

**DIE VERWENDUNG EREIGNISKORRELIERTER POTENTIALE
IN DER
PSYCHOPHYSIOLOGISCHEN AUSSAGEBEGUTACHTUNG
TÄUSCHUNGS-AUFDECKUNG MIT HILFE DER P300-KOMPONENTE**

Inauguraldissertation
zur
Erlangung der Doktorwürde

vorgelegt der

Philosophischen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität
zu Bonn

von
Prisca Jäger
aus
Troisdorf-Sieglar

Bonn 2002

Inhaltsverzeichnis

1. Zur Situation der Anwendung der Psychophysiologischen Aussagebegutachtung	1
2. Terminologische Vorbemerkungen	6
3. Verfahren der Psychophysiologischen Aussagebegutachtung	10
3.1 Die herkömmlichen Verfahren	10
3.1.1 Die Kontrollfragentechnik	10
3.1.1.1 Durchführung und Auswertung	10
3.1.1.2 Wirkungslogik	14
3.1.1.3 Forschungsstand zur Treffsicherheit	16
3.1.1.4 Varianten der Kontrollfragentechnik	16
3.1.1.5 Probleme der Kontrollfragentechnik	19
3.1.2 Die Tatwissentechnik	21
3.1.2.1 Durchführung und Auswertung	22
3.1.2.2 Wirkungslogik	23
3.1.2.3 Forschungsstand zur Treffsicherheit	25
3.1.2.4 Varianten der Tatwissentechnik	26
3.1.2.5 Probleme der Tatwissentechnik	27
3.2 Verwendung ereigniskorrelierter Potentiale	28
3.2.1 Die Verwendung der N400-Komponente	29
3.2.2 Die Verwendung der „Contingent Negative Variation“ (CNV)	33
3.2.3 Die Verwendung der P300-Komponente	36
3.2.3.1 Wirkungslogik und theoretischer Hintergrund	36
3.2.3.2 Untersuchungen zur P300	38
3.2.3.3 Bewertung der Verwendung der P300 zur Täuschungsaufdeckung	54
3.3 Allgemeine Bewertung der Verfahren	54

4. Untersuchungsparadigmen in der Psychophysiologischen Aussagebegutachtung	59
4.1 Feldstudien	59
4.2 Analogstudien	61
4.2.1 Karten- und Zahlentests	61
4.2.2 Täuschen bezüglich autobiographischer Informationen	62
4.2.3 Das Mock Crime	63
4.2.4 Hybridstudien	65
4.2.5 Bewertung der Forschungsszenarien	69
5. Ableitung der Fragestellungen	73
6. Methodik	77
6.1 Untersuchungsvariablen und Versuchsplan	77
6.1.1 Die Untersuchungsvariablen	77
6.1.1.1 Die unabhängigen Variablen	77
6.1.1.2 Die abhängigen Variablen	78
6.1.1.3 Zusätzliche Kontrollvariablen	79
6.1.2 Versuchsplan	79
6.2 Die Untersuchungsstichprobe	80
6.3 Materialien und apparativer Aufwand	81
6.3.1 Beobachtungsregeln	81
6.3.2 Filmmaterial	82
6.3.3 Feedbackbögen	83
6.3.4 Gudjonsson Suggestibility Scale (GSS)	85
6.3.5 Itemlisten	86
6.3.6 Rekognitionstest	88
6.3.7 Apparatur zur Stimulusdarbietung und Ableitung des EEGs	88
6.4 Die Untersuchungsdurchführung	89

6.5 Validierung der Methodik	95
6.5.1 Überprüfung der Annahme: Das Bonner Szenario erzeugt eine andere emotionale und motivationale Voraussetzung als ein Mock Crime	95
6.5.2 Validierung einzelner Vorannahmen des Bonner Szenarios	98
6.5.2.1 Misstrauen gegenüber dem vorgegebenen Untersuchungsanliegen	98
6.5.2.2 Feedbackbögen als Kriterium für die Beobachtungsleistung	99
6.5.2.3 Wirksamkeit des Vorhalts und resultierende Täuschungsmotivation	100
6.5.2.4 Probandenbelastung	101
6.6 Datenaufbereitung und statistische Auswertung	101
6.6.1 Datenaufbereitung	101
6.6.2 Auswertung zur Überprüfung der Annahme 1	102
6.6.3 Auswertung zur Überprüfung der Annahmen 3 und 4	103
6.6.4 Auswertung zur Überprüfung der Annahme 5	104
6.6.5 Auswertung zur Überprüfung des Einflusses der Kontrollvariablen	105
7. Ergebnisse	106
7.1 Annahme 1: Tatunspezifische Reize rufen ausgeprägte P300 hervor	106
7.2 Annahmen 3 und 4: Klassifizierbarkeit bezüglich Täterschaft	106
7.3 Annahme 5: Deliktthemenspezifische Unterschiede	126
7.4 Ergebnisse bezüglich der Kontrollvariablen	127

8. Diskussion	130
8.1 Beantwortung der Fragestellungen	130
8.2 Vergleich mit anderen Studien im Bereich der Täuschungsaufdeckung mit Hilfe der P300	132
8.3 Diskussion unter methodischen Gesichtspunkten	138
8.4 Theoretische Einbettung der Ergebnisse und Diskussion im Hinblick auf die Anwendung der Methode	140
9. Zusammenfassung	145
10. Literaturverzeichnis	148
Anhang	163

1. Zur Situation der Anwendung der Psychophysiologischen Aussagebegutachtung

„Man konnte fast den Eindruck gewinnen, die Auseinandersetzung mit der Psychophysiologischen Aussagebegutachtung findet in Deutschland nur noch in den Medien statt...“ (Vehrs, 2001). Zu dieser Feststellung sieht sich Vehrs gezwungen, obwohl es inzwischen vereinzelte Bemühungen von verschiedenen Seiten gibt, die Psychophysiologische Aussagebegutachtung (PAG) als Forschungsthema in Deutschland zu etablieren (Gödert, Rill & Vossel, 2001; Jäger, Schmidt & Scholz, 2000; Musch & Rossnagel, 1999; Vossel, Rill & Gödert, 2000). Die Attraktivität dieses Forschungsthemas wird durch die Möglichkeiten zur Anwendung der PAG beeinflusst.

Was die Anwendung der PAG im deutschen Strafverfahren betrifft, standen in der Geschichte der Rechtsprechung zur PAG lange Zeit ethische Bedenken im Vordergrund: Nach einer Entscheidung des Bundesgerichtshofes (BGH) im Jahre 1954 waren die mit Hilfe der PAG gewonnenen Befunde im Strafverfahren nicht verwertbar (1 StR 578/53). Die Begründung bezog sich auf strafprozessrechtliche Belange und erfolgte unabhängig von der wissenschaftlichen Beurteilung der Verfahren. Mit der Berücksichtigung der Ergebnisse, die mit Hilfe der PAG gewonnen wurden, habe das Gericht, so die Urteilsbegründung, gegen §136 der Strafprozessordnung verstoßen. Dieser Paragraph bezieht sich auf verbotene Vernehmungsmethoden. Zwar wird die PAG hier nicht namentlich genannt, doch der BGH befand, dass mit Anwendung dieser Methode eine unzulässige Beeinflussung der Willensbildung vorgenommen worden sei, auch wenn der Beschuldigte zuvor eingewilligt habe.

Im Jahre 1981 wurde die Frage nach der Verwendbarkeit der PAG erneut aufgeworfen. Während bei der Entscheidung 1954 ein belastendes Testergebnis bereits vorgelegen hatte, hatte nun ein rechtskräftig zu lebenslanger Freiheitsstrafe Verurteilter die Durchführung einer PAG beantragt. Der Vorprüfungsausschuss des 2. Senats des Bundesverfassungsgerichts sah 1981 in einer „derartigen Durchleuchtung“ der Person einen Verstoß gegen das durch das Grundgesetz geschützte Persönlichkeitsrecht (BVerfG, 1981, S. 446f.). Die Begründung des Beschlusses (2 BvR 166/81) wurde vielfach kritisiert (z.B. Schwabe, 1982; Amelung, 1982). Eines der Argumente war, dass es nicht verfassungswidrig sein könne, einen Unschuldsnachweis beibringen zu dürfen.

Eine neue Chance für die PAG tat sich in den 90er Jahren auf. Mit dem Anstieg der angezeigten Missbrauchsfälle, nahm auch die Anzahl derjenigen zu, die sich gefordert fühlten, ihre Unschuld nachzuweisen. Denn zwar gilt vor deutschen Gerichten das Prinzip „in dubio pro reo“, doch dieses Prinzip gilt bei einer Anschuldigung wegen Kindesmissbrauchs häufig nicht im familiären und sozialen Umfeld. Besonders wenn es an einem konkreten Tatvorwurf mangelt oder das betroffene Kind zu jung für eine umfassende Aussage ist (zu diesem Problem siehe Offe, 2001), ist es schwierig, die Sachlage zu beurteilen. Gerichte, die mit sorge- und umgangsrechtlichen Fragestellungen betraut sind, stehen auch in solchen Fällen vor dem Problem, eine Entscheidung treffen zu müssen. Aus dieser Situation heraus bot sich die PAG als Hilfsmittel an. Außerhalb des strafrechtlichen Bereichs ist das bei Gericht eingereichte Ergebnis einer PAG bei familienrechtlichen Entscheidungen berücksichtigt worden (z.B. Oberlandesgericht (OLG) Bamberg, 1995). Die strafgerichtliche Rechtsprechung hat sich der Beurteilung des Strafsenats dagegen überwiegend angeschlossen. Am 17. Dezember 1998 fällte der Bundesgerichtshof (in der Sache 1StR 156/98) folgendes Urteil:

1. Wirkt der Beschuldigte freiwillig an einer polygraphischen Untersuchung mit, so verstößt dies nicht gegen Verfassungsgrundsätze oder §136a StPO.
2. Die polygraphische Untersuchung mittels des Kontrollfragentests und – jedenfalls zum Zeitpunkt der Hauptverhandlung – des Tatwissentests führt zu einem völlig ungeeigneten Beweismittel i.S.d. §244 Abs. 3 Satz 2 4.StPO.
(BGH, 1999).

Es wird also in diesem Urteil nicht mehr der ethische Standpunkt von 1954 und von 1981 bestätigt. Insofern ist es als Fortschritt hin zur Anwendung der Methoden der PAG zu werten, dass in diesem Urteil an strafprozess- und verfassungsrechtlichen Bedenken nicht festgehalten wird, da prozessrechtliche Bedenken ein „K. O. – Kriterium“ sind, dem Forschungsergebnisse kaum etwas entgegensetzen können.

Die Bedenken richten sich gegen den wissenschaftlichen Stand bezüglich der PAG und wurden durch psychologische Gutachter vorgebracht. Kontrollfragen- sowie Tatwissenverfahren werden als „völlig ungeeignete Beweismittel“ (BGH, 1999, S. 205) eingestuft. Erläutert wurden in der Urteilsbegründung die beiden eben erwähnten, herkömmlich in der Praxis Verwendung findenden Methoden, bei denen standardmäßig

vier Körpersignale abgeleitet werden. Dazu gehören Herzfrequenz, Atmung, Blutdruck und Hautleitwert. Insgesamt wird das Kontrollfragenverfahren in der Urteilsbegründung als in „maßgebenden Fachkreisen nicht allgemein und zweifelsfrei als richtig und zuverlässig eingestufte Methode“ (BGH, 1999, S. 213) angesehen und ihm somit jeder Beweiswert abgesprochen. Der Tatwissentest ist von dieser Beurteilung zwar ausgenommen, wird aber aufgrund der Tatsache, dass dem zu Unrecht Beschuldigten das abgefragte Tatwissen unbekannt sein muss, ebenfalls als ein ungeeignetes Beweismittel zum Zeitpunkt der Hauptverhandlung eingestuft.

In der Urteilsbegründung werden zutreffend einige Schwachstellen und Mängel der empirischen Überprüfung der Verfahren der PAG genannt. Darüber hinaus sind jedoch einige substantielle Kritikpunkte nicht unwidersprochen geblieben (z.B. Offe & Offe, 2001). Auf die Schwachstellen der Kontroll- und Tatwissenverfahren wird in den Kapiteln 3 und 4 genauer eingegangen werden.

Der BGH hat zumindest für die nähere Zeit der strafrechtlichen Anwendung der PAG in der Praxis entgegengewirkt. Für den familienrechtlichen Bereich gilt dies bis dato nicht, wenn auch die strafrechtliche Praxis diesbezüglich Auswirkungen zeigt (Vehrs, 2001). Die wissenschaftliche Absicherung der Methode wurde vor ihren Einsatz in der Praxis gestellt, was zu begrüßen ist. Will man die Methoden und Möglichkeiten der PAG für die Praxis nutzbar machen, müssen wissenschaftliche Belege der Wirksamkeit herkömmlicher oder alternativer Verfahren erbracht werden. Um die Beibringung solcher Belege verwirklichen zu können, wird es notwendig sein, neue Ideen und Konzepte in die Forschung einzubringen, die die vorgebrachte Kritik berücksichtigen und neue Erkenntnisse ermöglichen.

Steller und Dahle (1999) sehen neben der Anwendung im Strafprozess und in Sorge- und Umgangsrechtverfahren weitere sinnvolle Anwendungsmöglichkeiten. So könnte die PAG im Ermittlungsverfahren bei bekannter Täterschaft eingesetzt werden, um zusätzliche Informationen zu erhalten, wie beispielsweise den Aufenthaltsort von entführten Personen, Namen gefährdeter Personen oder Informationen über Beuteverstecke. Es gibt Überlegungen, die dahin gehen, die Kriminalprognose oder Therapie mit einem Sexualstraftäter durch den Einsatz der PAG zu unterstützen. Ähnlich ließe sich die PAG zur Kontrolle der Einhaltung von Bewährungsaufgaben (z.B. der Auflage, sich einer bestimmten Person nicht zu nähern) einsetzen (Steller & Dahle, 1999).

Stadler (2001) weist auf solche Fälle hin, in denen Einlassungen des Beschuldigten sich darauf beziehen, dass er die in Frage stehende Handlung zwar vorgenommen habe, dies aber nicht in sexueller Absicht geschehen sei. In diesem Fall versagt die inhaltsorientierte Analyse der Zeugenaussage. Diese Konstellation eignet sich jedoch als Anwendungsfeld für den Kontrollfragentest.

Da jedes Anwendungsfeld eigene Probleme aufweist, ist zu fordern, dass die jeweilige Methode durch geeignete, auf dieses Feld zugeschnittene, Studien überprüft wird. Darüber hinaus sollte stets eine Kosten-Nutzen-Abwägung des Einsatzes der Methode stattfinden. So ist es zum Beispiel fraglich, ob die Vorteile des Einsatzes der PAG zur Kontrolle des Sexualstraftäters in der Therapie die Beschneidung der Menschenrechte und das Eindringen in die Privatsphäre rechtfertigen. Auch die Anwendung im Bereich der Personalauswahl und -überwachung sowie die Nutzung durch Versicherungen stellen fragliche Bereiche dar. In anderen Ländern ist die Anwendung in solchen Bereichen bereits die Regel. Horowitz (1999) beispielsweise schätzt die Anzahl der im Jahr 1997 in Israel durchgeführten PAG-Verfahren auf 18900. Den größten Anteil machen dabei Untersuchungen von Versicherungen aus.

Die vorliegende Untersuchung widmet sich dem Bereich der Sexualstraftaten. Hier steht häufig Aussage gegen Aussage. Die Glaubhaftigkeitsbegutachtung des Zeugen kann unbefriedigend ausfallen, wenn festgestellt wird, dass aufgrund von beeinflussenden Faktoren die Aussage verzerrt und damit nicht mehr verwertbar ist. Dies ist nicht nur deswegen ein unbefriedigendes Ergebnis, weil möglicherweise eine zu Recht beschuldigte Person nicht belangt werden kann, sondern auch deswegen, weil an einer zu Unrecht beschuldigten Person ein Restzweifel haften bleibt. Wie weiter oben bereits dargelegt, könnte die PAG einen geeigneten Entlastungsnachweis bei Beschuldigungen bezüglich Kindesmissbrauch oder anderer Sexualstraftaten darstellen. Voraussetzung zur Anwendung ist natürlich die wissenschaftliche Fundierung der einzelnen Methoden.

Eine Alternative zur herkömmlichen Registrierung peripherphysiologischer Parameter stellt die Nutzung ereigniskorrelierter Hirnpotentiale dar. Eine kleine Zahl von Studien (siehe Kapitel 3.2) hat sich dieser Thematik angenommen. Mit der vorliegenden Arbeit soll eine Methode der Anwendung ereigniskorrelierter Hirnpotentiale zur Täuschungsaufdeckung, nämlich die Registrierung der in einem Drei-

Kategorien-Oddball auftretenden P300-Komponente überprüft werden. Das Forschungsszenario, in dessen Rahmen diese Methode der PAG untersucht wird, soll in stärkerer Analogie zur Sexualstraftat stehen als bisher verwendete Szenarien.

2. Terminologische Vorbemerkungen

Diese Vorbemerkungen scheinen notwendig, da einige in den Medien und auch in der Fachliteratur verwendete Bezeichnungen („Lügendetektion“, „Polygraphentest“) für die bezeichneten Verfahren falsch oder zumindest irreführend sind. Zudem besteht in Expertenkreisen Uneinigkeit bezüglich der Bezeichnung der Psychophysiologischen Aussagebegutachtung (PAG).

Die Bezeichnung „Lügendetektion“, die sich im populärwissenschaftlichen Bereich hartnäckig hält, ist falsch, da es bei den angesprochenen Methoden nicht darum geht, Lügen als solche zu entlarven. Weder bezieht sich die Untersuchungslogik der bezeichneten Verfahren auf das Phänomen der „Lüge“, noch wird die Existenz einer speziellen Lügenreaktion angenommen, die mit Hilfe der Verfahren aufgedeckt werden soll. Die Annahme, dass es eine spezifische Reaktion gibt, die bei Täuschung auftritt, wird nicht vertreten. Allerdings gehen Rosenfelds (2002) Bemühungen in diese Richtung, wenn er annimmt, dass eine spezielle topographische Verteilung der P300-Amplitude bei Täuschung zu verzeichnen ist (siehe Kapitel 8). Da diese Annahme jedoch bei keiner der übrigen im Rahmen der PAG verwendeten Verfahren gemacht wird, muss eine dahin gehende Kritik „in's Leere gehen“ (Steller, 1987; Offe & Offe, 2001).

Das Anliegen der PAG besteht in der Aufdeckung von Täuschung oder Täuschungsintention, in dem Sinne, dass jemand, der das Zutreffen einer Beschuldigung abstreitet, der Täterschaft oder Nicht-Täterschaft überführt werden soll. Die PAG erfordert nicht einmal verbales Leugnen, sondern kann auch damit operieren, dass der Proband schweigt. Um den Unterschied in der Begrifflichkeit zu verdeutlichen, werden im Folgenden die Begriffe „Täuschung“ und „Lüge“ definiert.

„Täuschung“ ist der allgemeinere, „Lüge“ der spezifischere Begriff. Köhnken (1990) versteht unter dem Begriff der „Täuschung“ Verhaltensweisen, mit denen bei anderen Personen ein Eindruck erzeugt werden soll, von dem der Handelnde annimmt, dass er falsch ist. Diese Definition behandelt „Täuschung“ als ein Kommunikationsphänomen, das einen Kommunikator und einen Rezipienten voraussetzt. Hopper und Bell (1984) haben nach einer Clusteranalyse eine Typologie mit sechs unterschiedlichen Täuschungsformen vorgeschlagen: 1) Erfindung, Dichtung; 2) Scherzhafte Täuschung; 3) Lüge; 4) Straftaten (Täuschungsformen, die per Gesetz mit Strafe bedroht sind); 5)

Verschleierung; 6) „Nicht-Lügen“ (durch falsche Implikation in die Irre führen). Täuschung kann in vielfältiger Form, z.B. als scherzhafte Täuschung oder nonverbal erfolgen. Lüge ist dagegen eine verbale Täuschung (Köhnken, 1990). Eine Definition der Lüge, wie sie selbst zwischen Antipoden wie Augustinus (395/1979) und Nietzsche (1873/1980) kaum umstritten ist, besagt: Eine Lüge besteht im bewussten Behaupten eines nach eigener Überzeugung falschen Satzes, zumeist in täuschender, irreführender oder betrügerischer Absicht. Den Umstand, dass eine Lüge im Vergleich zu einer erlebnisfundierten Aussage eine komplexere kognitive Leistung erfordert, machte man sich bei der Konzeption inhaltlicher Glaubhaftigkeitskriterien (z.B. Undeutsch, 1967 oder Steller & Köhnken, 1989) zu Nutze. Diese Kriterien werden bei der Beurteilung von Zeugenaussagen angewandt.

Eine beschuldigte Person ist dagegen nicht gezwungen, eine Lüge, beispielsweise im Sinne eines erfundenen Alibis, zu produzieren. Sie hat das Recht, sich nicht zu äußern, oder kann die inkriminierte Tat schlicht abstreiten. Die Wirkungslogik der Verfahren der psychophysiologischen Aussagebegutachtung baut nicht auf dem Konzept der „Lüge“ auf, sondern auf intraindividuellen Reaktionsvergleichen bezüglich unterschiedlicher Fragen und Reizwörter.

Die Bezeichnung „Polygraphie“ rückt einen technischen Aspekt der Verfahren in den Vordergrund. Der Polygraph (Mehrkanalschreiber) ist Teil der apparativen Anordnung bei der Aufzeichnung mehrerer physiologischer Signale. Die Bezeichnung mag zu der irrigen Annahme verleiten, ein solches Verfahren sei eine Art „Wahrheitstest“. Ob eine Person „zu Recht beschuldigt“ wird oder nicht, würde nach dieser Annahme im Sinne einer Vorverurteilung von einer Maschine bestimmt. Es handelt sich bei der PAG auch nicht um ein bestimmtes Verfahren, wie besonders der Begriff „Polygraphentest“ suggeriert, sondern um eine ganze Klasse von Verfahren. Im englischsprachigen Raum unterscheidet die Verwendung dieses Begriffs die Praktiker (z.B. Abrams, 1989) von den eher forschungsorientierten Anwendern (z.B. Honts & Raskin, 1988).

Die Entscheidung für die Bezeichnung „Psychophysiologische Aussagebegutachtung“ fiel im Rahmen der vorliegenden Arbeit aus den folgenden Gründen: Die Psychophysiologie lässt sich als Teilgebiet der Biopsychologie von der Physiologischen Psychologie oder der Neuropsychologie vor allem durch das Untersuchungsobjekt und

die verwendeten Methoden abgrenzen. So werden physiologische Vorgänge vorwiegend am menschlichen Organismus mit nicht-invasiven Registrier- und Messmethoden untersucht. Die Physiologische Psychologie untersucht dagegen biologische Vorgänge und neuronale Strukturen mit direkter Reizung, Registrierung oder Zerstörung der Hirnaktivität, so dass sie in vielen Bereichen auf den Tierversuch angewiesen ist. Die Neuropsychologie konzentriert sich auf das menschliche Gehirn und arbeitet vorwiegend mit Patienten (Birbaumer & Schmidt, 1990). Man nimmt in der Psychophysiologie an, dass physiologische Veränderungen als Begleiterscheinungen psychologischer Prozesse auftreten, wodurch sie als objektiv erfassbare diagnostische Kriterien subjektiver Vorgänge verwendet werden können (Steller, 1987). Man verwendet also im hier angesprochenen Fall die Annahmen und Methoden der Psychophysiologie, um eine Beurteilung oder Begutachtung der Aussage vorzunehmen. Daher wird in Anlehnung an Steller (1987) die Bezeichnung „Psychophysiologische Aussagebegutachtung“ verwendet. Da sich auch in englischsprachigen Forschungsarbeiten die Bezeichnung *„psychophysiological detection of deception“* eingebürgert hat (Barland & Raskin, 1973; Bradley & Ainsworth, 1984; Honts, 1994; Waid & Orne, 1981), wird keine Notwendigkeit für die Übernahme des von Undeutsch (Undeutsch & Klein, 1999) vorgeschlagenen Begriffs „Physiopsychologische Verdachtsabklärung“ gesehen.

Der zweite Teil der Bezeichnung, „Aussagebegutachtung“ verweist auf die Parallelen zur inhaltsorientierten und verhaltensorientierten Aussagebegutachtung. In keinem Bereich der Aussagebegutachtung geht man davon aus, dass ein Merkmal oder eine eindeutige Merkmalskombination identifizierbar ist, die eine erlebnisbasierte von einer nicht erlebnisfundierte Aussage unterscheidet. Es handelt sich vielmehr um diagnostische Schlussfolgerungsprozesse. Bei der inhaltsorientierten Vorgehensweise wird aus dem Vorliegen geeigneter Merkmale (z. B. Steller & Köhnken, 1989) vor dem Hintergrund der Fähigkeiten der aussagenden Person, der geschilderten Tatumstände und unter Einbeziehung von Interviewmerkmalen und der Aussagegenese auf die Glaubhaftigkeit einer Aussage geschlossen. In ähnlicher Weise wie die inhaltlichen Merkmale können auch verhaltensorientierte oder psychophysiologische Merkmale in den diagnostischen Prozess einbezogen werden, um die Glaubhaftigkeitsbegutachtung auf Beschuldigte erweitern zu können. Voraussetzung ist natürlich, dass an diese

Merkmale die gleichen wissenschaftlichen Anforderungen gestellt werden wie an die inhaltsorientierten Merkmale.

Es ist nicht korrekt im Falle der PAG von einem „Test“ zu sprechen. Ein psychologischer Test ist ein „wissenschaftliches Routineverfahren zur Untersuchung eines oder mehrerer empirisch abgrenzbarer Persönlichkeitsmerkmale mit dem Ziel einer möglichst quantitativen Aussage über den relativen Grad der individuellen Merkmalsausprägung.“ (Lienert, 1994). Das Vorgehen der PAG ähnelt aber vielmehr einem einzelfallbezogenen Quasi-Experiment mit dem Ziel einer diagnostischen Entscheidung (Scholz, 1999). Die Registrierung der physiologischen Daten stellt nur einen Teil der PAG-Untersuchung dar und sie liefert keineswegs bereits eine quantitative Aussage über die Wahrscheinlichkeit einer tatsächlichen Täterschaft. Um solche Konnotationen zu vermeiden sollte die Bezeichnung „Test“ nicht im Zusammenhang mit der PAG verwendet werden. Aus diesem Grund werden im Folgenden die Begriffe „Kontrollfragentest“ oder „Tatwissentest“ vermieden.

3. Verfahren der Psychophysiologischen Aussagebegutachtung

Gegenstand der Betrachtung sind zunächst die herkömmlichen Verfahren der Psychophysiologischen Aussagebegutachtung (3.1). Anschließend wird die Möglichkeit, ereigniskorrelierte Potentiale zur Täuschungsaufdeckung einzusetzen, demonstriert (3.2). Es schließt sich eine Diskussion der Vor- und Nachteile der zuvor erläuterten Verfahren nach dem bisherigen Forschungsstand an (3.3).

3.1 Die herkömmlichen Verfahren

Im Folgenden werden Kontrollfragen- (3.1.1) und Tatwissentechnik (3.1.2) vorgestellt. Bei beiden Verfahren werden physiologische Reaktionen verglichen, die auf die Darbietung unterschiedlicher Reize hin auftreten. In der herkömmlichen Version der PAG werden elektrodermale Aktivität, Blutdruck, Atmung und Herzfrequenz als Biosignale abgeleitet.

3.1.1 Kontrollfragentechnik

In diesem Abschnitt werden Durchführung und Auswertung (3.1.1.1), Wirkungslogik (3.1.1.2), Forschungsstand (3.1.1.3), sowie Varianten (3.1.1.4) und Probleme (3.1.1.5) der Kontrollfragentechnik (KFT) skizziert.

3.1.1.1 Durchführung und Auswertung.

Am Institut für Forensische Psychophysiologie (IFP) in München bzw. Bayreuth erfolgt die praktische Durchführung und Auswertung einer PAG mit Hilfe der Kontrollfragentechnik beispielsweise in folgenden Schritten (Vehrs, 1999):

1. *Vorbereitung der Untersuchung*: Nachdem ein Gutachtenauftrag aufgrund eines Gerichtsbeschlusses erfolgt ist, wird der Beschuldigte angeschrieben und zu einem Untersuchungstermin, der etwa einen halben Tag in Anspruch nimmt, eingeladen. Die Fragestellungen fallen meist in den Bereich inkriminierter Sexualstraftaten. Nach ausführlichem Aktenstudium wird die Untersuchung geplant. Die Untersuchungsfragen werden vorformuliert. Die endgültigen Fragen werden während des Vortestinterviews festgelegt.

Die Grundform der Kontrollfragentechnik sieht drei Typen von Fragen vor: Relevante Fragen, Kontrollfragen und neutrale Fragen. Die relevanten Fragen beziehen

sich auf die inkriminierte Tat. Sie sollen sich auf den schwersten Vorwurf konzentrieren, kurz, klar und möglichst verhaltensnah formuliert sein. Die Beantwortung der Fragen soll möglichst mit „Ja“ oder „Nein“ erfolgen können. Bei Verdacht auf Kindesmissbrauch könnte eine solche Frage also folgendermaßen lauten: „Haben Sie am (Datum) die Scheide der (Name der geschädigten Person) mit ihrem Penis berührt?“. Die Kontrollfrage, die auch als „Vergleichsfrage“ bezeichnet wird, bezieht sich auf Themen, die in der relevanten Frage angesprochen werden. Auch diese Frage soll vom Probanden verneint werden. Eine entsprechende Kontrollfrage zur oben formulierten relevanten Frage könnte lauten: „Haben Sie vor Ihrer Hochzeit einmal ein Kind gesehen, das Sie gerne nackt gesehen hätten?“ Der Proband sollte diese Frage aus Gründen der sozialen Erwünschtheit und aus Bedenken bezüglich der Relevanz im aktuellen Fall verneinen. Gleichzeitig sollte bei der Verneinung Unsicherheit bezüglich der Korrektheit dieser Antwort bestehen. Neutrale Fragen beziehen sich nicht auf den Tatvorwurf und können z. B. den Namen zum Inhalt haben („Heißen Sie (Name des Probanden)?“).

Es gibt verschiedene „Formate“ der KFT, die sich hinsichtlich der Verwendung zusätzlicher Fragetypen, hinsichtlich Reihenfolge und Anzahl der verschiedenen Fragetypen unterscheiden. Zur Anwendung kommt am IFP meist ein Kontrollfragenverfahren, das an die „Zone Comparison Technique“ von Backster (1979, zitiert nach Matte, 1996, S.343) angelehnt ist und eines der verschiedenen Formate der KFT darstellt (siehe Kapitel 3.1.1.4).

Bei Ankunft wird der Proband über seine Rechte informiert. Eine schriftliche Einverständniserklärung bezüglich der Teilnahme wird eingeholt.

2. *Vortestinterview*: Das Vortestinterview stellt den wichtigsten Teil der PAG-Untersuchung dar, da hier die erforderliche Einstellung des Probanden bezüglich relevanter Fragen und Kontrollfragen erfolgt. Es dient zunächst dazu, biographische Daten zu erheben, die als Leitfaden für die Erarbeitung der Kontrollfragen dienen können. Berücksichtigt werden auch Variablen, die die physiologischen Parameter beeinträchtigen können. Dies können gesundheitliche Auffälligkeiten, Alkohol oder Drogenkonsum, Medikamenteneinnahme oder aktuelle emotionale Beeinträchtigungen sein.

Untersuchungstechnisch ist es von Relevanz, den Probanden von der Treffsicherheit der Methode zu überzeugen. Das ist zum einen wichtig, da dadurch der Täter eher als solcher klassifiziert werden kann. Zum anderen kann die Besorgnis bezüglich der Funktionsfähigkeit der Methode beim Nichttäter zu falsch positiven Ergebnissen führen. Matte (1996) weist darauf hin, dass das Interview nicht mit anklagender Haltung geführt werden sollte, da dies ebenfalls die Gefahr falsch positiver Ergebnisse birgt. Die Arbeitsweise der Messinstrumente wird vor oder während der Applikation der Elektroden erklärt. Nach Applikation der Messfühler wird zu Demonstrationszwecken ein sogenannter Stimulationstest durchgeführt, der mit Hilfe von Karten- oder Zahlentests (siehe Kapitel 4) die Funktionsfähigkeit der Methode vor Augen führen soll. Die Wirksamkeit des Stimulationstests wird durch verschiedene Studien in der Weise belegt (z. B. Bradley & Janisse, 1981), dass Probanden, bei denen der Test vor der PAG appliziert wurde, eher korrekt klassifiziert wurden. Ob der Proband von der Wirksamkeit des Verfahrens tatsächlich überzeugt ist, lässt sich nicht mit letzter Sicherheit feststellen.

Gemeinsam mit dem Probanden werden relevante Fragen und Kontrollfragen formuliert. Zunächst werden die relevanten Fragen generiert. Hierzu schildert der Proband, wessen er beschuldigt wird und wie er sich das Zustandekommen dieser Anschuldigung erklärt. Für die Generierung der Kontrollfragen wird der Proband bezüglich seines Sexuallebens exploriert. Im Rahmen dieser Exploration wird der Proband zum einen einige Eingeständnisse machen, die bei der Formulierung der Kontrollfragen ausgeklammert werden müssen. Es werden aber auch Sachverhalte besprochen, die der Proband nicht zugibt, zumal er annehmen muss, dass diese Informationen vom Untersucher zur Urteilsbildung herangezogen werden. Fragen nach solchen Sachverhalten, die sich zudem auf einen langen Zeitraum beziehen und recht vage formuliert werden, eignen sich als Kontrollfragen, wie z. B. „Haben Sie sich vor 1995 – außer dem, worüber wir gesprochen haben – schon einmal für ein nacktes Kind interessiert?“. Schließlich werden dem Probanden alle Fragen, auch neutrale und andere Fragen, vorgetragen. Im Rahmen der Entwicklung der Fragen erfolgt auch die Einstellung des Probanden auf die relevanten Fragen und Kontrollfragen, so dass diese die geforderte Bedeutsamkeit entwickeln. Das heißt, die Einstellung erfolgt so, dass der Täter seine Aufmerksamkeit auf die relevanten Fragen fokussiert, während sich die

Besorgnis des Nichttäters auf die Kontrollfragen konzentriert. Mit welchen Methoden dieses Ziel tatsächlich regelhaft erreicht werden kann, ist nicht geklärt. Dafür wurden bisher keine Standards ausgearbeitet.

3. Durchführung der Befragung unter Ableitung der physiologischen Parameter:

Die eigentliche Testphase ist der kürzeste Untersuchungsabschnitt. Die zuvor besprochenen Fragen werden in der vereinbarten Reihenfolge gestellt. Gewöhnlich erfolgen drei Durchgänge. In den Pausen zwischen den Durchgängen können die Kontrollfragen revidiert werden, falls dies erforderlich ist.

Die Sicherung der Daten erfolgt nur noch selten mit Hilfe eines Mehrkanalschreibers, sondern meist digitalisiert mit Personalcomputern, unterstützt durch entsprechende Software.

4. Die Nachtestbesprechung: Es folgt ein abschließendes Gespräch. Hier hat der Proband Gelegenheit, Beobachtungen während des apparativen Untersuchungsteils mitzuteilen, die eventuell relevante Alternativhypothesen zum Zustandekommen des Untersuchungsergebnisses liefern können.

5. Die Auswertung: Prinzipiell werden jeweils die Reaktionen auf Kontrollfragen mit denen auf relevante Fragen verglichen. Jede im Vergleich zur Kontrollfrage stärkere Reaktion auf eine relevante Frage vergrößert die Wahrscheinlichkeit, dass eine Person als „Täter“ klassifiziert wird. Eine verstärkte Reaktion auf eine Kontrollfrage vergrößert die Wahrscheinlichkeit einer Beurteilung als „Nichttäter“.

Zur Abschätzung der Reaktionsstärke werden verschiedene Merkmale und Parameter herangezogen. Bezüglich der Atmung werden beispielsweise Atemfrequenz und -tiefe als abhängige Variablen verwendet. Auch eine Apnoe beim Ein- oder Ausatmen oder eine Veränderung der Baseline fließen in die Beurteilung mit ein (Matte, 1996). Die Auswertung und Interpretation der Daten kann per Hand oder computergestützt erfolgen. Man kann außerdem eine global-intuitive Vorgehensweise von einem numerischen Ansatz unterscheiden. Im ersten Fall entscheidet der Auswerter nach Durchsicht der Daten eher intuitiv über Klassifikation des Probanden. Beim numerischen Ansatz wird ein sogenanntes „semi-objektives“ numerisches Schätzverfahren angewendet, das auf Backster (1963) zurückgeht. Hierbei ist festgelegt, welche Reaktionen miteinander zu vergleichen sind. Zur Quantifizierung der Reaktionsunterschiede sind einige Regeln vorgegeben. Je nachdem, wie die Vergleiche ausfallen,

werden Zahlenwerte vergeben, die zu einem Summenscore aufaddiert werden. Diese Vorgehensweise ist semi-objektiv, weil die Skalierung der Reaktionsunterschiede regelgeleitet aber letztendlich subjektiv erfolgt. Einen Überblick über die verwendeten Parameter, Auswertungsalgorithmen und Grenzwerte gibt Matte (1996) für verschiedene Formate der KFT.

3.1.1.2 Wirkungslogik.

Da eine unmittelbare Erfassung des Bedeutungsgehaltes von Stimuli auf der Grundlage von physiologischen Begleiterscheinungen nicht möglich ist, vergleicht man bei der Kontrollfragentechnik peripherphysiologische Reaktionen auf relevante Reize mit denen auf Kontrollreize. Bei der herkömmlichen Version werden Blutdruck-, Atmungs-, Herzfrequenz- und Hautleitwertänderungen als peripherphysiologische Reaktionen registriert, obwohl prinzipiell auch andere Biosignale sinnvoll einbezogen werden könnten (siehe z. B. Curio & Scholz, 1991).

Die Kontrollfragentechnik stellt eine Erweiterung der Relevant-Irrelevant-Technik (Reid, 1947) dar. Im Rahmen der Relevant-Irrelevant-Technik werden die Reaktionen auf relevante Fragen mit denen auf neutrale Fragen, die jeweils verneint werden sollen, verglichen. Da auch dem Nichttäter die Bedeutsamkeit der relevanten Frage nach der inkriminierten Tat offensichtlich ist, ist diese Methode sehr anfällig für falsch-positive Entscheidungen. Die Einführung der Kontrollfrage soll diesem Effekt entgegenwirken. Die Kontrollfrage hat ebenfalls belastenden Charakter, bezieht sich jedoch nicht auf das inkriminierte Geschehen. Dem Probanden wird vermittelt, dass seine Reaktionen auf die Kontrollfrage in der Weise Auswirkungen auf die Beurteilung haben werden, dass sie z. B. Auskunft über die Persönlichkeit des Beschuldigten geben. Diese Information kann vorgeblich dazu dienen, einzuschätzen, ob es wahrscheinlich ist, dass die betreffende Person die in Frage stehende Tat begangen hat. Auf diese Weise kann man Aussagen zu der Bedeutsamkeit des relevanten Reizes in Relation zu einem Vergleichsstimulus machen. Entsprechend kommt den Eigenschaften des Vergleichsstimulus entscheidende Bedeutung zu. Er muss folgende Kriterien (Steller & Dahle, 1999) erfüllen:

1. *Vergleichbarkeitskriterium*: Darunter ist die Forderung zu verstehen, dass der Kontrollstimulus keine eigenständigen, unkontrollierten Signalqualitäten aufweisen

darf. Eine Bedeutungsverschiebung zwischen beiden Fragen soll nur durch ihre Eigenschaft als relevante Frage beziehungsweise Kontrollfrage verursacht werden. Im Sinne der Vergleichbarkeit ist es beispielsweise wichtig, dass die Kontrollfrage ebenfalls pädophile Neigungen thematisiert, wenn die relevante Frage dies tut, so dass die relevante Frage nicht allein durch Ansprechen dieser Thematik bedeutsam ist. Matte (1996) hat zu Standardisierungszwecken einen Katalog von Kontrollfragen zu unterschiedlichen Delikttypen aufgestellt.

2. *Niveaukriterium*: Im Rahmen der Kontrollfragentechnik versucht man, die Kontrollfrage so zu formulieren, dass drei Kriterien erfüllt sind: Sowohl der Täter als auch der Nichttäter soll erstens diese Frage wahrheitswidrig verneinen, zweitens sollen beide dies mit Täuschungsabsicht tun und drittens sollen beide dabei befürchten, dass eine Aufdeckung dieser Täuschung dazu beiträgt, sie als Täter zu klassifizieren. Man versucht dies zu erreichen, indem man das Vergleichbarkeitskriterium ausnutzt, so dass es dem Probanden plausibel erscheint, dass die Beantwortung der Frage mit der inkriminierten Tat in Zusammenhang steht.

Voraussetzung für eine korrekte Klassifikation ist, dass der Kontrollreiz für den Nichttäter eine höhere Bedeutsamkeit aufweist als der relevante Reiz. Die Bedeutsamkeit der Kontrollfrage darf jedoch für den Täter die Bedeutsamkeit der relevanten Frage nicht übertreffen. Dies stellt die eigentliche Herausforderung der Methode dar. Wenn dies nicht gelingt, sind die Ergebnisse nicht verwertbar. Das Vortestinterview hat daher in erster Linie die Zielsetzung Kontrollfragen von geeignetem Niveau individuell für den Probanden zu bestimmen.

3. *Unabhängigkeitskriterium*: Die Formulierung der Kontrollfragen sollte so sein, dass ihre Bedeutsamkeit nicht von der Tatbegehung abhängt. Sie sollten einen stabilen Vergleichsmaßstab darstellen.

Anhand dieser Kriterien kann überprüft werden, ob eine hinreichend eindeutige Interpretation der Reaktionen auf die Reize möglich ist. Je stärker die Kriterien im Einzelfall verletzt wurden, um so weniger kann die Interpretation eindeutig ausfallen, da konfundierende Variablen die Reaktionen mitbedingen.

3.1.1.3 Forschungsstand zur Treffsicherheit.

Raskin und Honts (2002) fassen den Forschungsstand bezüglich der Treffsicherheit der PAG mit Hilfe der Kontrollfragentechnik folgendermaßen zusammen: Die mittlere Trefferquote in Analogstudien beträgt 91%. Im Mittel werden 90% der Scheintäter korrekt klassifiziert und 92% der Nichttäter. Die Rate der falsch-positiven Ergebnisse liegt bei 8%, die der falsch-negativen bei 10%. In Feldstudien lag die Trefferquote im Mittel bei 90,5%. Wurden Beschuldigte nach einem Außenkriterium (z. B. Geständnis, Verurteilung) klassifiziert, betrug die korrekte Klassifikation bezüglich der Täter 98%, bezüglich der Nichttäter betrug sie 75%. Die Rate der falsch-positiven Klassifikationen betrug also 25%, die der falsch-negativen 2%. Die Trefferquoten wurden unter Vernachlässigung der unentscheidbaren Fälle angegeben.

Steller und Dahle (1999) stellen zusammenfassend fest, dass die empirisch gefundenen Trefferquoten Zufallsklassifikationen deutlich überlegen sind. Theoretische Annahmen und empirische Belege bezüglich der Kontrollfragentechnik verweisen auf ein Überwiegen der falsch-positiven Klassifikationen und damit auf eine hohe Sensitivität. Die KFT eignet sich daher besonders als Entlastungsverfahren.

3.1.1.4 Varianten der Kontrollfragentechnik.

Die Kontrollfragentechnik wird in unterschiedlichen Formaten durchgeführt. Unterschiede bestehen in der Anzahl und Reihenfolge der verwendeten Fragen, sowie hinsichtlich der Hinzuziehung von verschiedenen Fragetypen und dem Vorgehen bei der Auswertung. Neben den bereits vorgestellten Kontroll-, neutralen und relevanten Fragen werden weitere Typen von Fragen hinzugezogen. Die wichtigsten Typen werden im Folgenden kurz vorgestellt:

Sogenannte „*Symptomatische Fragen*“ betreffen beispielsweise Sorgen des Probanden, die vom Untersuchungsleiter nicht intendiert sind und Einfluss auf die Ergebnisse haben. Eine starke Reaktion auf diese Frage ist in dem Sinne symptomatisch oder alarmierend, dass sie anzeigt, dass den Probanden möglicherweise nicht intendierte Sorgen beschäftigen, was zu einer verminderten Reaktion auf Kontroll- und relevante Fragen führen kann („super-dampening-effect“, Matte, 1996). Eine symptomatische Frage kann beispielsweise folgendermaßen lauten: „Sind Sie überzeugt davon, dass ich Ihnen nur Fragen stellen werde, die wir zuvor besprochen haben?“. Zeigen sich starke

Reaktionen auf diese Frage, wird dies mit dem Probanden besprochen und versucht, die entsprechende Besorgnis auszuräumen. Anschließend erfolgt ein erneuter Durchlauf. Einer Untersuchung von Capps, Knill und Evans (1993) zufolge, reduziert die Verwendung dieser Frage die Anzahl der unentscheidbaren Fälle gegenüber der Durchführung der PAG ohne diese Frage. Dementsprechend ist sie Bestandteil zahlreicher Formate der KFT.

Eine *vorbereitende, relevante Frage* bereitet die eigentlich relevante, tatbezogene Frage vor und soll eine Überreaktion auf die erste relevante Frage verhindern. Sie kann z. B. folgendermaßen lauten: „Bezüglich (inkriminierte Tat) haben Sie vor, die Wahrheit zu sagen?“. Die Frage wird entsprechend ihrer Funktion zu Beginn und vor den relevanten Fragen gestellt. Ist die Reaktion auf die vorbereitende, relevante Frage höher als auf die eigentlich relevanten Fragen, ist zu überprüfen, ob der Proband Gegenmaßnahmen angewendet hat. Horvarth (1994) fand keinen Vorteil der Verwendung der vorbereitenden, relevanten Frage. Diese Studie wird allerdings von Matte (1996) in methodischer Hinsicht kritisiert. Auch diese Frage ist Bestandteil vieler Formate.

Die „*Fear of Error-Kontrollfrage*“ und „*Hope of Error-relevante Frage*“ werden verwendet, weil man sich von ihnen einen zusätzlichen Hinweiswert erwartet. Aus dem Vergleich der Reaktionen auf die beiden Fragen ergibt sich ein zusätzlicher Hinweiswert auf die Täterschaft oder Nicht-Täterschaft unabhängig von der Thematisierung des inkriminierten Delikts. Die „*Fear of Error-Kontrollfrage*“ lautet „Befürchten Sie, dass bei der Testdurchführung ein Fehler gemacht wurde?“. Die „*Hope of Error-relevante Frage*“ lautet „Hoffen Sie, dass bei der Testdurchführung ein Fehler gemacht wurde?“. Diese Fragen werden nach den eigentlichen Kontroll- und relevanten Fragen gestellt. Bei der Auswertung werden die Reaktionen auf diese beiden Fragen miteinander verglichen. Die Wertung erfolgt analog der Vergleiche zwischen Kontrollfragen und relevanten Fragen. Studien, die die Effektivität der *Fear of Error-Kontrollfrage* und der *Hope of Error-relevante Frage* zeigen, stehen noch aus. Beide Fragen sind z. B. Bestandteil der Backster Zone Comparison Technik (Backster, 1979, zitiert nach Matte, 1996, S. 343).

Kontrollfragen lassen sich hinsichtlich der Verwendung von Zeitschranken unterscheiden. Das heißt, man kann eine Kontrollfrage auf einen bestimmten Zeitraum

begrenzen, z. B. „Haben sie nach Ihrem 18. Geburtstag...?“. Außerdem werden Kontrollfragen, die das inkriminierte Geschehen einbeziehen, von solchen unterschieden, die es ausschließen. Durch die Verwendung von Zeitschranken und Formulierungen, die das inkriminierte Delikt ausschließen, sollen Zweideutigkeiten vermieden werden.

Von unterschiedlichen Anwendergruppen wurden verschiedene Formate erstellt. Matte (1996) erwähnt z. B. die Folgenden: „Quadri-Track Zone Comparison“ Technik, „Backster Tri-Zone Comparison“-Technik, „Utah Zone Comparison“-Technik. Sie unterscheiden sich hinsichtlich der Anzahl der Paare von vergleichbaren Kontrollfragen und relevanten Fragen. Unterschiedliche Formen von Kontroll- und zusätzlichen Fragen werden verwendet. Keines der Formate kann empirisch begründet als eindeutig überlegen bezeichnet werden.

Die gerichtete Lügenkontrollfragentechnik oder „directed lie question technique“ (DLQT/ obwohl auch hier nicht gelogen, sondern nur wahrheitswidrig abgestritten wird) gleicht im Prinzip der KFT. Auch hier werden die Reaktionen auf relevante Fragen den auf Vergleichsfragen gegenübergestellt. Der Unterschied in der Durchführung besteht darin, dass der Proband explizit aufgefordert wird, die Vergleichsfrage wahrheitswidrig zu beantworten. Horowitz, Kircher, Honts und Raskin (1997) geben folgende Beispiele für gerichtete Lügenkontrollfragen: „Haben Sie jemals gelogen?“, „Haben Sie jemals eine Vorschrift oder eine Regel gebrochen?“ oder „Haben Sie jemals einen Fehler gemacht?“. Das heißt, es sollte sich um persönlich bedeutsame, verhaltensbezogene Vergleichsfragen handeln, die kaum jemand wahrheitswidrig verneinen kann. Im Prinzip sind triviale Fragen auch denkbar, haben sich aber in der oben genannten Untersuchung von Horowitz et al. (1997) als weniger geeignet erwiesen. Um den Signalcharakter der Frage zu erhöhen wird dem Probanden mitgeteilt, dass mit dieser Frage der Vergleichsstandard erstellt wird, um festzustellen, ob der Proband hinreichend stark bei einer Täuschung reagiere. Die Wirkungslogik der DLQT impliziert, dass der Nichttäter ein uninterpretierbares Ergebnis vermeiden möchte und so die Vergleichsfrage einen höheren Signalwert aufweist als die relevante Frage. Die bedeutsamere Frage für den Täter bleibt die relevante Frage, da er sich bewusst ist, dass er auch hier täuscht. Vorteile sind hier die Standardisierbarkeit der Fragen, die nicht mehr individuell formuliert werden müssen und die Möglichkeit, den Probanden direkt

zur Täuschung aufzurufen, ohne ihn auf subtilere Weise zum Leugnen verleiten zu müssen. Allerdings wird durch die inhaltliche Entfremdung das Vergleichbarkeitskriterium verletzt. Auch wird das Unabhängigkeitskriterium insoweit verletzt, als die Vergleichsfrage für Täter und Nicht-Täter eine unterschiedliche Signalqualität erhält, da der Täter ein uninterpretierbares Ergebnis nicht unbedingt vermeiden möchte. Dies führt zu einer erschwerten Interpretation der Ergebnisse. Außerdem wird eine „Einladung zur Manipulation“ durch die „Aufklärung“ bezüglich der Vergleichsfragen befürchtet (Steller & Dahle, 1998).

3.1.1.5 Probleme der Kontrollfragentechnik.

Folgende Annahmen sind insofern problematisch, als sie im Einzelfall als Prämissen vorausgesetzt werden, ohne überprüft zu werden: 1. Der Proband ist von der Bedeutsamkeit der Kontrollfrage für das Ergebnis der PAG überzeugt und täuscht bzgl. dieser Frage. 2. Die Bedeutsamkeit der Kontrollfrage übertrifft tatsächlich die Bedeutsamkeit der relevanten Frage für den Nichttäter. 3. Für den Täter ist die Bedeutung der relevanten Frage dominant.

Validierungsstudien zeigen zum Teil zwar sehr gute Klassifikationsergebnisse, allerdings wurden bisher nicht die oben erläuterten Annahmen bezüglich der Wirksamkeit überprüft, sondern es wurde lediglich indirekt bei erfolgreicher Klassifikation auf die Korrektheit der Prämissen geschlossen. Scholz (1999) hat als eine Möglichkeit Bedeutsamkeitsskalierungen durch die Probanden vorgeschlagen, um die Prämissen selbst bestätigen zu können.

Ansatzweise wurde dies von Horowitz, Kircher, Honts und Raskin (1997) bereits geleistet. Nachdem fünf Durchgänge eines PAG-Verfahrens absolviert wurden, beantworteten die Probanden in dieser Untersuchung einen Fragebogen. Dieser forderte, jede Frage hinsichtlich ihrer Bedeutsamkeit bezüglich des Testergebnisses zu bewerten und anzugeben, wie stark nach subjektivem Empfinden auf diese Reize reagiert worden sei. Als unabhängige Variablen wurden der Täterstatus (Täter vs. Nichttäter) und der Fragentyp (relevante Frage, Kontrollfrage, neutrale Frage) variiert. Als abhängige Variable ging das Bedeutsamkeitsrating bezüglich des Testergebnisses in die Analyse ein. Die Ratingskala reichte dabei von „1 = sehr wichtig“ bis „7 = absolut unwichtig“ für das Bedeutsamkeitsrating und von „1 = starke Reaktion“ bis „7 = kaum Reaktion“

für die Einschätzung der eigenen physiologischen Reaktion. Eine 2 x 3 faktorielle ANOVA ergab keine signifikante Interaktion zwischen Täterstatus und Fragentyp, wenn das Bedeutsamkeitsrating als abhängige Variable verwendet wurde. Es zeigte sich ein Haupteffekt für den Fragentyp. Relevante Fragen wurden von allen Probanden als bedeutsamer eingeschätzt als Kontrollfragen und neutrale Fragen. Wurde als abhängige Variable die Einschätzung der eigenen physiologischen Reaktionen auf die gestellten Fragen verwendet, ergab die ANOVA den erwarteten Interaktionseffekt. Das heißt, Täter schätzten ihre Reaktionen auf relevante Fragen höher ein als auf Kontrollfragen. Dagegen schätzten Nichttäter ihre Reaktionen auf Kontrollfragen höher ein als auf relevante Fragen. Informationen darüber, inwieweit diese Einschätzungen mit den tatsächlich registrierten physiologischen Daten übereinstimmen, werden nicht angegeben.

Vehrs (1999) schlägt eine nicht verbale Erlebnisskalierung zur Bedeutsamkeitseinschätzung von Fragen vor. Weitere Untersuchungen zu dieser Thematik mit differenzierteren Skalierungsmethoden werden benötigt.

Die Möglichkeit von Manipulationsversuchen durch den Probanden stellt ein weiteres Problem der KFT dar. Motorische „Countermeasures“ (Zungenbeißen, Muskelkontraktionen, Atmungsveränderungen usw.) sind vom Probanden schwer so anzuwenden, dass eine falsch-negative Klassifikation resultiert (Honts, Raskin, Kircher & Hodes, 1988). Dies erfordert Einsicht in die Wirkungslogik und setzt eine gewisse Übung voraus, da sie gezielt eingesetzt werden muss, um die Reaktionen auf Kontrollfragen zu verstärken und die Dosierung der motorischen Aktivität schwierig ist. Bei geeignetem Training ist die erfolgreiche Anwendung allerdings möglich (Rovner, Raskin & Kircher, 1979). Eine Möglichkeit motorische Gegenmaßnahmen aufzudecken, besteht in elektromyographischen Aufzeichnungen (Honts & Hodes, 1983). Diese Aufzeichnungen können Anspannungen verschiedener Muskelpartien anzeigen, die zu Manipulationszwecken erfolgen. Sedierende Drogen könnten prinzipiell die Chance erhöhen, ein nichtklassifizierbares Ergebnis zu erhalten. Durch Drogenscans oder Verhaltensbeobachtung könnte solchen Gegenmaßnahmen der Probanden entgegen gewirkt werden. Auch Biofeedbacktraining oder mentale Gegenmaßnahmen (Erhöhung der Aktivität oder Denken an anregende Situationen bei Kontrollfragen) sind prinzipiell geeignet die Ergebnisse der KFT zu manipulieren. Die Anwendung der sogenannten

„Stanislavsky“-Technik¹, mit deren Hilfe Schauspieler versuchen, sich durch persönliche Erinnerungen in einen erforderlichen emotionalen Zustand zu versetzen, führte allerdings im Rahmen einer Laborstudie nicht zu vermehrt falsch-negativen Klassifikationen (Dawson, 1980). Die Probanden versuchten mit Hilfe dieser Technik allerdings nicht, die Reaktionen auf Kontrollfragen gezielt zu erhöhen, sondern waren instruiert, sich als eine zu Unrecht beschuldigte Person zu fühlen.

Kritisiert wird darüber hinaus die Generalisierbarkeit der vorhandenen Simulationsstudien, die die Treffsicherheit der KFT überprüfen sollen (BGH, 1998). Hinsichtlich der entsprechenden Feldstudien wird die Validität der Außenkriterien, die den Täterschaftsstatus bestimmen, in Zweifel gezogen (siehe Kapitel 4).

Zusammenfassend lässt sich feststellen, dass die Voraussetzungen der KFT und ihre Wirkungslogik problematisch sind. In der Praxis versucht man, die Wirksamkeit der Kontrollfragen und relevanten Fragen bestmöglich herzustellen, ohne genau angeben zu können, mit welchen Methoden dies gelingt und ob es tatsächlich gelungen ist. Es ist fraglich, ob die erforderlichen Kriterien im Einzelfall tatsächlich erfüllt werden. Die bisherige Forschung liefert zwar Belege für gute Klassifizierbarkeit mit Hilfe der KFT, liefert aber keine Erkenntnisse über die Erfüllung der theoretischen Grundannahmen. Auf Grund der bisherigen Forschung zu den sogenannten „Countermeasures“ lässt sich nicht ausschließen, dass Probanden solche erfolgreich anwenden können. Verschiedene Formate der KFT widmen sich einiger verbesserungswürdiger Aspekte. So kann beispielsweise durch Verwendung der DLQT eine bessere Standardisierbarkeit erreicht werden. Häufig entstehen, wie im Fall der DLQT, durch Neuerungen allerdings neue Probleme hinsichtlich der formulierten Kriterien, die Kontrollfragen erfüllen sollen. Die genannten Probleme der KFT sind somit bisher nicht zufriedenstellend gelöst.

3.1.2 Die Tatwissenteknik

In diesem Abschnitt werden Durchführung und Auswertung (3.1.2.1), Wirkungslogik (3.1.2.2), Forschungsstand (3.1.2.3), Varianten (3.1.2.4) und Probleme (3.1.2.5) der Tatwissenteknik (TWT) skizziert.

¹ Benannt nach dem russischen Theaterdirektor, Schauspieler und Lehrer Konstantin Stanislavsky.

3.1.2.1 Durchführung und Auswertung.

Die TWT kommt in Deutschland nur selten zur Anwendung (Vehrs, 1999). Matte (1996) führt die Tatwissenteknik folgendermaßen durch:

1. *Vorbereitung der Untersuchung:* Für die Anwendung der TWT wird vorausgesetzt, dass Informationen zum Tatgeschehen vorliegen, die nur dem Täter und den Ermittlungsbehörden bekannt sind. Von der Anzahl dieser Informationen ist die Zahl der relevanten Fragen abhängig. Die Fragen und Antwortalternativen werden vorformuliert. Eine relevante Frage könnte bei einer Beschuldigung wegen sexuellen Kindesmissbrauchs lauten: „Welche Farbe hatte die Unterwäsche der (Geschädigten)?“, wenn sich aus dem Tatvorwurf ergibt, dass der Beschuldigte nur bei Zutreffen dieses Vorwurfs über dieses Wissen verfügen kann. Die Antwortalternativen könnten dann „schwarz“, „grau“, „grün“, „blau“, „braun“, „rot“ sein. „Blau“ könnte die relevante Antwortalternative darstellen, da die Unterwäsche der Geschädigten tatsächlich diese Farbe hatte.

Neben relevanten Fragen werden auch sogenannte Pufferfragen formuliert, die keine tatbezogenen Antwortalternativen einbeziehen. Ein Multiple-Choice-Format wird verwendet (Wo stand der Schreibtisch, aus dem der Umschlag entwendet wurde. Stand er a) links an der Wand, b) rechts an der Wand, c) in der Mitte des Zimmers...). Im Falle einer relevanten Frage könnte Antwortalternative c) mit den tatsächlichen Gegebenheiten am Tatort übereinstimmen, im Falle einer Pufferfrage würde keine der Alternativen zutreffen. Eine andere Möglichkeit besteht in der Verwendung von „Ja-Nein“-Fragen (Wissen Sie, dass die Farbe des gestohlenen Autos blau war? Wissen Sie, dass sie grün war?...) (Office of Technology Assessment (OTA), 1987). Die Zahl der Pufferfragen, die Zahl der Alternativantworten und die Platzierung der relevanten Fragen bzw. der relevanten Antwortalternativen kann variiert werden. Weder die relevante Frage noch die relevante Antwortalternative stehen allerdings an erster Stelle, da die erste Frage, beziehungsweise die erste Antwortalternative unabhängig von ihrem Inhalt beinahe immer eine vergleichsweise starke Reaktion hervorruft.

2. *Vortestinterview:* Dem jeweiligen Probanden wird die Vorgehensweise erläutert. Es wird erklärt, dass die Aufzeichnungen der physiologischen Daten sich bei nicht vorliegender Täterschaft während der gesamten Untersuchung nicht verändern. Beim Täter dagegen steige die Erregung solange, bis die relevanten Fragen beantwortet

seien. Eventuell könne der Proband, der täuschen wolle, diese Erregungssteigerung selbst wahrnehmen. Durch diese Erklärung soll die Relevanz der entsprechenden Frage und Antwort erhöht werden.

Anschließend werden die einzelnen Fragen mit dem Probanden besprochen. Ziel ist es herauszufinden, inwieweit dem Beschuldigten Informationen über relevante Reize bekannt sind. Die Reihenfolge der Fragen während der Ableitung wird allerdings nicht bekannt gegeben. Da die Wirkungslogik erläutert wurde, kann dem jeweiligen Probanden begründet werden, warum es wichtig ist, aus anderen Quellen erworbenes Tatwissen vorab als solches zu erkennen zu geben.

3. *Durchführung der Befragung unter Ableitung der physiologischen Parameter:* Nach Applikation der Elektroden erfolgen mindestens zwei Durchgänge. Abgeleitet werden meist Blutdruck, Herzfrequenz, Atmung und Hautleitfähigkeit. Auch hier ist die Hinzuziehung anderer physiologischer Variablen denkbar. Die Fragen und Antwortalternativen werden vorgelesen. Verbale Antworten sind keine notwendige Voraussetzung für die erfolgreiche Klassifizierung (Gustafson & Orne, 1965). Alternative Präsentationsmodi, wie visuelle Darbietung der Fragen und Antworten, aber auch die Darbietung von Tatortfotos oder Stimmproben von am Tatgeschehen beteiligten Personen sind denkbar.

4. *Die Nachtestbesprechung:* Nach der Durchführung besteht die Gelegenheit, die erhaltenen Kurven, die den Verlauf der abgeleiteten physiologischen Variablen wiedergeben, gemeinsam mit dem Probanden zu besprechen. Auf diese Weise kann überprüft werden, inwieweit die Ergebnisse durch andere Ursachen als das Vorhandensein von Tatwissen erklärt werden können.

5. *Auswertung:* Für die Auswertung werden nur die relevanten Fragen berücksichtigt. Die Reaktionen auf relevante Reize, also auf relevante Antwortalternativen, werden mit denen auf neutrale Reize, also nichtbedeutsame Antwortalternativen, verglichen. Für die verwendeten Auswertungsalgorithmen und Grenzwerte siehe Lykken (1959).

3.1.2.2 Wirkungslogik.

Die Wirkungslogik beruht darauf, dass sich der Täter vom Nichttäter durch tatspezifische Informationen unterscheidet, die nur er haben kann. Bei dieser Technik

werden daher tatrelevante Reize in den Kontext neutraler aber semantisch und syntaktisch möglichst gleichartiger Reize eingebettet. Für den Nichttäter ergibt sich kein Unterschied zwischen relevanten und neutralen Reizen. Für den Täter hat der relevante Stimulus² eine andere Bedeutung als die übrigen Stimuli.

Die peripherphysiologischen Reaktionen auf relevante und nicht relevante Antwortalternativen werden miteinander verglichen. Eine systematisch stärkere Reaktion auf relevante Reize im Vergleich zu den neutralen Items identifiziert den Täter. Die Wahrscheinlichkeit des zufälligen Auftretens einer erhöhten Reaktion auf relevante Reize lässt sich bei gegebener Anzahl von Alternativmöglichkeiten und Zahl der Durchgänge angeben. Um die Wirksamkeit des Verfahrens gewährleisten zu können müssen einige Voraussetzungen gegeben sein:

1. Das Vergleichbarkeitskriterium muss auch bei der TWT gegeben sein. Die Alternativwahlmöglichkeiten müssen echte, plausible Alternativen darstellen. Eine zufällige Reaktionsverteilung bei einer Person ohne Tatwissen ist nur unter der Voraussetzung anzunehmen, dass die relevanten Items nicht durch irgendeine andere Eigenschaft als den Tatbezug von den übrigen Items zu unterscheiden sind. Es muss gewährleistet sein, dass die übrigen Items ihrerseits ebenso ungewöhnlich oder gewöhnlich sind wie das relevante Item und keines durch eine Besonderheit die Aufmerksamkeit auf sich zieht.

2. Die Ermittlungen zum Sachverhalt müssen soweit gediehen sein, dass dem Untersucher genügend geeignete, tatbezogene Informationen bekannt sind, um ein geeignetes Itemset erstellen zu können.

3. Die Detailinformationen müssen in der Weise geeignet sein, dass sie dem Täter sicher bekannt sind. Sie müssen von solcher Relevanz sein, dass der Täter sie mit einiger Sicherheit im Zusammenhang mit seiner Tat wahrgenommen hat. Hinsichtlich der Formulierung wird gefordert, dass der Täter diese Wahrnehmungen auch in der Wortwahl repräsentiert findet.

4. Schließlich dürfen dem Nichttäter die verwendeten Detailinformationen nicht bekannt sein. Für die praktische Anwendung der TWT stellt die Generierung von entsprechenden Items nicht zuletzt deswegen ein solches Problem dar, weil „unschuldig“ erworbene tatrelevante Informationen das Verfahren zu einem falsch

² Die Begriffe „Reiz“, „Stimulus“ und „Item“ werden im Folgenden weitgehend synonym verwendet.

positiven Ergebnis führen können. Einige Studien scheinen die Robustheit des TWT gegenüber solchem Wissen zu belegen: Giesen und Rollison (1980) sowie Stern, Breen, Watanabe und Perry (1981) konnten zeigen, dass Nichttäter sich auch dann noch von Tätern unterschieden, wenn ihnen tatspezifisches Wissen dargeboten worden war. Dieses Wissen war den Nichttätern allerdings in einem anderen Kontext als dem des inkriminierten Geschehens dargeboten worden. Bradley, MacLaren & Carle (1996), Bradley und Rettinger (1992) sowie Bradley und Warfield (1984) fanden, dass ein Teil der Nichttäter mit Hilfe des „Guilty Action Tests“ (eine Form der TWT, siehe Kapitel 3.1.2.4) korrekt als solche klassifiziert wurden, obwohl ihnen im Zusammenhang mit der in Frage stehenden Tat spezifisches Wissen mitgeteilt worden war.

3.1.2.3 Forschungsstand zur Treffsicherheit.

Elaad (1998) wertete 15 Analogstudien zur Treffsicherheit der TWT aus. Die Trefferquoten betragen im Durchschnitt 80,6% bezüglich der Scheintäter und 95,9% bezüglich der Nichttäter. In 11 von diesen 15 Studien gab es keine falsch-positiven Klassifikationen. Bisher existieren nur wenige Feldstudien zur Treffsicherheit der TWT. Elaad (1990) und Elaad, Ginton und Jungman (1992) überprüften sie im Rahmen von zwei Feldstudien. Im Durchschnitt lagen die falsch-positiven Klassifikationen bei 2% in der erstgenannten beziehungsweise bei 5% in der letztgenannten Studie. Diese Angaben sind vergleichbar mit denen, die in Analogstudien gefunden wurden. Die Rate der falsch-negativen Klassifikationen liegt mit 42% beziehungsweise 20% höher als in den Analogstudien. Die hohen Fehlerraten in diesem Bereich könnten damit erklärt werden, dass die durchschnittliche Anzahl der verwendeten relevanten Fragen in den referierten Studien bei 2 beziehungsweise bei 1,8 liegen. Außerdem wurde bei Elaad (1990) als einzige physiologische Variable der Hautleitwert herangezogen. Theoretische Annahmen und empirische Befunde verweisen auf ein Überwiegen der falsch-negativen Klassifikationen und damit auf eine höhere Spezifität der Methode. Die TWT empfiehlt sich somit als Belastungsnachweis.

3.1.2.4 Varianten der Tatwissentechnik.

Nach Matte (1996) unterscheidet sich Lykken's „Guilty Knowledge Test“ (GKT) – also die klassische TWT – vom Peak of Tension Test (POT) lediglich darin, dass das relevante Item bei ersterem eine zufällige Position einnimmt, bei letzterem immer eine zentrale. Unter „Concealed Knowledge Test“ versteht er die Anwendung dieser Tests als Demonstration, wie beispielsweise den Zahlentest, der vor einer KFT durchgeführt wird. Der POT kann entweder als „Known Solution Peak of Tension Test“ durchgeführt werden, wenn ein Proband der Täterschaft überführt werden soll oder als „Probing Peak of Tension Test“, wenn die Anwendung bei starkem oder sicherem Täterschaftsverdacht auf die Gewinnung weiterer Informationen, wie z. B. Verstecke von Raubgut, Entführungsoffern, Namen von Hintermännern etc. abzielt.

Der POT unterscheidet sich vom GKT ferner durch die Art der Itemformulierung (OTA, 1987). Beim POT würde es z. B. heißen: „Bezüglich der Farbe des gestohlenen Autos, wissen Sie, dass es grün war? Wissen Sie, dass es schwarz war? usw.“ Ein Item des GKT wäre folgendermaßen formuliert: „Wo war die Beute versteckt? A) in dem Papierkorb, B) in der Schublade etc.“

Einen Vorteil hinsichtlich der Klassifizierbarkeit von Nicht-Tätern mit Tatwissen soll der sogenannte „Guilty Actions Test“ (GAT) haben. Er soll also die Zahl der falsch-positiven Entscheidungen gegenüber dem GKT vermindern (Bradley, MacLaren & Carle, 1996). Die Formulierung der relevanten Fragen wird hierbei gegenüber dem GKT dahin gehend abgewandelt, dass die in Frage stehenden Handlungen auf den Probanden bezogen werden. Auf diese Weise stellt die Verneinung der richtigen Alternative keine Täuschung dar, wenn es sich bei dem Probanden um einen Nichttäter mit Tatwissen handelt. Folglich würde eine Frage, die im GKT „In welchem Raum wurde der Diebstahl begangen?“ (am Tatwissen orientiert) lautet, im GAT folgendermaßen formuliert: „In welchem Raum haben Sie den Diebstahl begangen?“ (an Täterschaft und Tatwissen orientiert). Die praktische Vorgehensweise ist analog zur Vorgehensweise beim GKT. Bradley, MacLaren und Carle (1996) konnten im Rahmen einer Analogstudie zeigen, dass mit Hilfe der GAT 50% der Nichttäter mit Tatwissen korrekt als Nichttäter klassifiziert wurden. Demgegenüber wurden in der gleichen Studie mit Hilfe der GKT lediglich 10% der Nichttäter mit

Tatwissen korrekt klassifiziert. Die Rate der falsch-positiven Fehler ist allerdings auch mit Hilfe der GAT noch sehr hoch.

3.1.2.5 Probleme der Tatwissenteknik.

Die oben genannten Voraussetzungen an das Verfahren begründen zugleich seine Probleme. Die Auswahl und Formulierung geeigneter Reize stellt die größte Schwierigkeit dar. Die verwendeten Reize müssen Informationen darstellen, die vom Täter mit großer Sicherheit während der Tat wahrgenommen und bis zur Durchführung der PAG behalten wurden. Gleichzeitig dürfen diese Informationen, gemäß der theoretischen Prinzipien des Verfahrens, dem Nichttäter nicht bekannt sein. Dies dürfte zumindest zum Zeitpunkt der Hauptverhandlung in einem Strafverfahren so gut wie unmöglich sein.

Die Möglichkeit zur Verfälschung der Ergebnisse durch „Countermeasures“ besteht auch hier, z. B. durch die Einnahme eines Beruhigungsmittels (Waid, Orne, Cook & Orne, 1981). Durch Biofeedback- bzw. Autohypnosetraining soll ebenfalls eine Reaktionsunterdrückung bei relevanten Reizen erfolgen (Corcoran, Lewis & Garver, 1978). Mentale Maßnahmen, wie rückwärts zählen, können die Aufmerksamkeit reduzieren und so die Reaktion auf relevante Reize unterdrücken. Mentale und motorische Gegenmaßnahmen können aber auch eingesetzt werden, um die Reaktionen auf neutrale Stimuli zu erhöhen. Diese Maßnahme erwies sich in einer Arbeit von Lykken (1960) als wenig erfolgreich, da er die Abweichungen von einer Zufallsverteilung als Täuschungsmerkmal identifizieren konnte. Honts, Devitt, Winbush und Kircher (1996) konnten dagegen zeigen, dass auch die TWT nicht immun gegenüber mentalen und motorischen Gegenmaßnahmen ist. Ein Training war auch hier erforderlich, um geeignete Reaktionen auf neutrale Stimuli zu erzeugen.

Zusammenfassend lässt sich über die TWT sagen, dass sie hinsichtlich der Wirkungslogik von weniger restriktiven Voraussetzungen ausgeht. Voraussetzung für die Anwendbarkeit ist dabei, dass sich geeignete Items formulieren lassen. Die Schwierigkeit besteht eindeutig in der Generierung von Items, die ein Täter sicher als relevante Items wahrnimmt und über die ein Nichttäter keine Informationen hat. Einige Studien belegen, dass es möglich ist, mit Hilfe der TWT, beziehungsweise dem GAT, Probanden mit sogenanntem „unschuldigem Tatwissen“ (siehe Kapitel 3.1.2.2 und

3.1.2.4) korrekt einzuordnen. Diese Studien sind aber entweder ökologisch nicht valide – wenn das Tatwissen in einem anderen Kontext gelernt wurde – oder ergeben zu hohe Fehlerraten. Es besteht auch bei der TWT die Möglichkeit, dass der Proband Gegenmaßnahmen ergreift, die die Testergebnisse verzerren können. Diese Kritikpunkte reduzieren die Anwendbarkeit der TWT im Rahmen des intendierten Einsatzes der PAG bei inkriminierten Sexualstraftaten.

3.2 Verwendung ereigniskorrelierter Potentiale

Einer der wichtigsten methodischen Zugänge zur Erforschung der Zusammenhänge zwischen physiologischen Reaktionen und Kognitionen ist die Aufzeichnung elektrischer Aktivität des Gehirns. Das Elektroenzephalogramm (EEG) stellt elektrische Spannungsschwankungen des Gehirns dar, wobei auf der X-Achse die Zeit in Millisekunden-Intervallen (ms-Intervallen) und auf der Y-Achse die Potentiale in Mikrovolt (μV) abgetragen werden. Die EEG-Signale stellen die summierte Aktivität aus einer Vielzahl elektrischer Prozesse dar, die in vielen kortikalen aber auch subkortikalen Strukturen ablaufen.

Das Gehirn wirkt nicht nur als Generator für die Spontanaktivität³, sondern es reagiert auch mit typischen Potentialverläufen auf bestimmte innere und äußere Ereignisse. „Unter ereigniskorrelierten Hirnpotentialen (EKP) verstehen wir alle elektrokortikalen Potentiale, die vor, während und nach einem sensorischen, motorischen oder psychischen Ereignis im Elektroenzephalogramm (EEG) messbar sind“ (Birbaumer & Schmidt, 1999). Bei einem ereigniskorrelierten oder evozierten Potential handelt es sich um eine komplexe Wellenform mit mehreren Extrema. Da die Amplituden der EKP klein sind und durch andere Wellen des Spontan-EEGs überlagert werden, lassen Sie sich mit bloßem Auge im Spontan-EEG nicht erkennen. Um ein ereigniskorreliertes Potential aus dem EEG-Signal zu extrahieren, verwendet man ein Mittelungsverfahren. Dabei kommt ein wichtiger Unterschied zwischen Spontan-EEG und EKP zum tragen: Die EKP stehen in direktem Zusammenhang zu einem Reiz und treten in Bezug auf diesen in ähnlicher Form auf. Die Spontanaktivität tritt dagegen unabhängig vom gegebenen Reiz auf. Die Durchschnittskurve des Spontan-EEGs setzt

³ Unter der Spontanaktivität, deren Aufzeichnung entsprechend Spontan-EEG genannt wird, versteht man die ununterbrochen an der Schädeloberfläche registrierbaren Spannungsschwankungen.

sich aus zufällig variierenden positiven und negativen Werten zusammen. Mit steigender Zahl der Mittelungen geht diese Durchschnittskurve gegen Null. Die Amplitudenwerte, die aus dem evozierten Potential resultieren, bleiben dagegen unverändert und heben sich mit steigender Anzahl der Mittelungen immer deutlicher von dem Rauschen des Spontan-EEGs ab.

Mit anderen Worten: Das evozierte Potential wird an t_i aufeinanderfolgenden Zeitpunkten abgetastet. Die Anzahl dieser Zeitpunkte ist gleich T und geschieht für N ereignisbezogene Reaktionen. Betrachtet man ein bestimmtes t_x , so ist die Amplitude des ereignisbezogenen Potentials zum Zeitpunkt t_x für jede der N Reaktionen gleich. Die Amplitude des Spontan-EEGs variiert dagegen zufällig, so dass der Mittelwert der Amplituden zum Zeitpunkt t_x bei steigendem N gegen Null geht. Führt man diese Mittelung für alle Amplituden der T Zeitpunkte durch, resultiert das ereigniskorrelierte Potential.

Man unterscheidet zwischen *exogenen* und *endogenen* EKPs. Sensorische oder exogene Komponenten treten mit einer Latenz bis zu 100 ms nach einem Reiz auf. Die Amplituden dieser Komponenten ändern sich in erster Linie in Abhängigkeit von den physikalischen Charakteristika des Reizes. Alle Komponenten, die ab einer Latenz von 50 ms auftreten, zeigen Variationen in Abhängigkeit von psychischen Veränderungen und sind nicht mehr von den physikalischen Reiz- und Reaktionsbedingungen allein abhängig. Da man den Ursprung ihrer Variabilität innerhalb des Organismus vermutet, heißen sie *endogene* Komponenten.

Zunächst werden zwei Möglichkeiten der Verwendung endogener ereigniskorrelierter Hirnpotentiale, nämlich der N400 (3.2.1) und der „Contingent Negative Variation“ (CNV) (3.2.2), zur Täuschungsaufdeckung beschrieben. Anschließend wird eine dritte Möglichkeit, die Verwendung der P300-Komponente im Rahmen der PAG, vorgestellt (3.2.3).

3.2.1 Die Verwendung der N400-Komponente

Die N400 ist eine negative Welle, die mit einer Latenz von ca. 400 ms nach einem Reiz auftritt. Diese Komponente steht im Zusammenhang mit der semantischen Verarbeitung von Sprache. Sie tritt z. B. dann auf, wenn ein Stimuli präsentiert wird, der unerwartet

oder semantisch inkongruent in Bezug auf den Kontext des vorhergehenden Satzes ist. Ein erwarteter oder kongruenter Reiz erzeugt dagegen keine ausgeprägte N400.

Im klassischen Paradigma besteht die Aufgabe darin, Sätze, deren Wörter nacheinander auf einem Bildschirm erscheinen, still zu lesen. Die Sätze sollen mit der Intention gelesen werden, später inhaltliche Fragen beantworten zu können. Ein Teil der Sätze endet mit einem syntaktisch korrekten, aber semantisch inkongruenten Wort (z. B. „A robin is a truck“). Diese unerwarteten Satzergänzungen rufen eine ausgeprägte N400 hervor (Kutas & Hillyard, 1980).

Die N400 tritt auch dann auf, wenn Sätze in dem Sinn falsch beendet werden, dass sie zuvor gelernten Informationen widersprechen. Hat ein Proband beispielsweise zuvor gelernt, dass Diana Ärztin ist, ruft der Satz „Diana ist Rechtsanwältin“ eine stärker ausgeprägte N400-Komponente hervor als der Satz „Diana ist Ärztin“. Der N400-Effekt tritt nicht nur bei der sequentiellen Darbietung von Sätzen auf, sondern auch bei der Darbietung eines Kontextaussage-Testwort-Formats: „Ein Vogel: Amsel.“ (Neville, Kutas, Chesney & Schmidt, 1986). Dabei werden sequentiell der Kontext und das Testwort dargeboten. Diese Darbietungsform wurde auch in der nachfolgend vorgestellten Untersuchung verwendet.

Den Umstand, dass die N400 auftritt, wenn Sätze in dem Sinn falsch enden, dass sie zuvor erworbenem Wissen widersprechen, machten Boaz, Perry, Raney, Fischler und Shuman (1991) für die PAG nutzbar. Die Annahme war, dass ein Proband, der Tatwissen besitzt, da er die entsprechende Tat begangen hat, auf nicht zusammenpassende Kontext-Testwörter-Paare, deren Inkongruenz sich durch den Tatzusammenhang ergibt, mit einer im Vergleich zu kongruenten Paaren ausgeprägten N400 reagieren wird. Für einen Probanden, der die Tat nicht begangen hat und der auch keine Informationen darüber besitzt, werden alle Paare eher neutralen Charakter besitzen. Daraus ergibt sich, dass bei diesen Probanden auch keine N400-Unterschiede auftreten sollten.

Die Stichprobe bestand 1.) aus einer Gruppe von 59 Probanden, die durch eine Zeitungsannonce rekrutiert worden waren, und 2.) aus einer Gruppe von 16 Psychologiestudenten. Eine weitere 3.) Gruppe bestand aus Probanden der Gruppe 1, die nach drei Monaten ein zweites Mal getestet worden waren. Aus Gruppe 1 wurden 29, aus Gruppe 2 wurden 12 und aus Gruppe 3 wurden 14 Probanden der

Experimentalgruppe zugeteilt. Die übrigen Probanden wurden der Kontrollgruppe zugeteilt und simulierten kein Verbrechen. Als abhängige Variable fungierten die Amplitudenwerte der N400-Komponenten.

Probanden der Experimentalgruppe sahen einen Film über einen Einbruch, der aus der Perspektive des Einbrechers gefilmt war. Der Film begann mit dem Eindringen in ein fremdes Haus und zeigte daraufhin, wie verschiedene Gegenstände gestohlen und Akte von Vandalismus ausgeführt wurden. Die Probanden der Kontrollgruppe sahen einen Film, der verschiedene Alltagsszenen aus New York zeigte. Anschließend wurden alle Probanden beschuldigt, den Einbruch begangen zu haben. Sie sollten an einer PAG teilnehmen, wobei nur die Probanden der Experimentalgruppe über tatspezifisches Wissen verfügten. Es wurde ein EEG von zwei Ableitorten (Cz, Pz)⁴ aufgezeichnet. Zehn Paare von Items wurden präsentiert. Jedes Item bestand aus der Kontextaussage und dem Testwort. Jede Kontextaussage wurde sowohl mit einer korrekten (Distraktor) als auch mit einer falschen Ergänzungsinformation (tatspezifisches Item) versehen. Auf diese Weise bestand jedes Itempaar aus einem Distraktor und einem tatspezifischen Item. Der Kontext bezog sich auf örtliche Informationen (z. B. „in the kitchen“) oder auf Angaben zu gestohlenen oder zerstörten Details (z. B. „sports equipment“). Ein Item konnte also z. B. folgendermaßen aussehen: „Sportausrüstung: Gestohlen“.

Es wurden sowohl Grand Averages⁵ als auch eine Auswertung für jeden Einzelfall vorgenommen: Die Grand Averages wurden getrennt für die Experimental- und die Kontrollgruppe berechnet. Für die Einzelfallanalyse wurden die EEG-Daten pro Proband für tatspezifische Items und Distraktoren zu EKP-Kurven gemittelt. Für jeden Proband wurden in einem bestimmten Zeitfenster die Amplitudenwerte gemittelt. Auf diese Weise konnten für jeden Probanden Differenzen zwischen dem Amplitudenmittelwert bezüglich tatspezifischer Items und Distraktoren berechnet werden. Dies wurde für beide Ableitorte vorgenommen. Anhand der Gruppe 1 waren zuvor

⁴ Die mit Hilfe der am Kopf befestigten Elektroden registrierten hirnelektrischen Potentiale sind abhängig vom Ableitort. Um zu reproduzierbaren und vergleichbaren Ergebnissen zu kommen, wurde die Elektrodenplatzierung standardisiert. Das Standardsystem ist das sogenannte 10-20 System (Jasper, 1958). Als Bezugspunkte gelten Nasion (Übergang von Nasenrücken zu Stirn) und Inion (als Einbuchtung am Hinterhauptsknochen fühlbar), sowie die beiden präaurikulären Punkte (Vertiefungen unterhalb des Jochbeins, auf Höhe der Ohren). Indem man die Strecken zwischen den jeweiligen Bezugspunkten nach dem Schema 10%, 20%, 20%, 20%, 10% aufteilt, wobei sich die Prozentwerte auf den jeweiligen Gesamtabstand beziehen, erhält man die Achsen dieses Koordinatensystems und die Standardpositionen. Die Standardposition Cz befindet sich an der Stelle, an der sich die beiden Achsen schneiden, Pz befindet sich auf der Sagittallinie im Parietalbereich.

⁵ Grand Averages sind über alle Probanden einer Bedingung gemittelte EKP-Kurven.

Grenzwerte berechnet worden, die optimal zwischen Kontroll- und Experimentalgruppe trennten. Es waren sowohl Grenzwerte für die Differenzen der Amplitudenmittelwerte als auch die zugehörigen Zeitfenster in einem Bereich zwischen 200 und 560 ms nach Erscheinen des Testworts auf dem Bildschirm ermittelt worden, so dass maximal viele Probanden korrekt klassifiziert werden konnten. Anhand dieser Grenzwerte wurden alle Probanden in Probanden mit Tatwissen und in Probanden ohne solches unterteilt. Mit Hilfe eines χ^2 -Tests wurde ermittelt, ob die Trefferraten größer als zufällig waren.

In der Experimentalgruppe fiel für beide Ableitorte die Amplitude der N400 bezüglich tatspezifischer Items größer aus als bezüglich der Distraktoren (siehe Abbildung 1). In der Kontrollgruppe unterschieden sich die Amplituden der N400 für tatspezifische Items und Distraktoren nicht. Anhand der ermittelten Grenzwerte konnten in Gruppe 1 78%, in Gruppe 2 87,5% und in Gruppe 3 64% der Probanden korrekt zugeordnet werden. Diese Zahlen galten jeweils sowohl für die bei Cz als auch für die bei Pz abgeleiteten Daten. Mit Hilfe des χ^2 -Tests konnte belegt werden, dass die Trefferquoten für alle Gruppen größer als zufällig waren.

Die hier verwendete Methode zur Klassifikation im Einzelfall lässt sich nicht auf die praktische Anwendung übertragen. Eine Referenzstichprobe zur Gewinnung geeigneter Grenzwerte steht hier nicht zur Verfügung. Schwierigkeiten bei der Gewinnung allgemeiner Grenzwerte bestehen darin, dass diese sich mit neuem Reizmaterial und anderen Untersuchungsumständen wahrscheinlich ändern werden. Obwohl es sich bei der Anwendung der gewonnenen Grenzwerte auf Retest-Werte der Eichstichprobe handelt, wurden eher niedrige Trefferquoten erzielt. Weitere Untersuchungen und die Entwicklung anderer Auswertungsalgorithmen sind für die Beurteilung dieses Verfahrens notwendig.

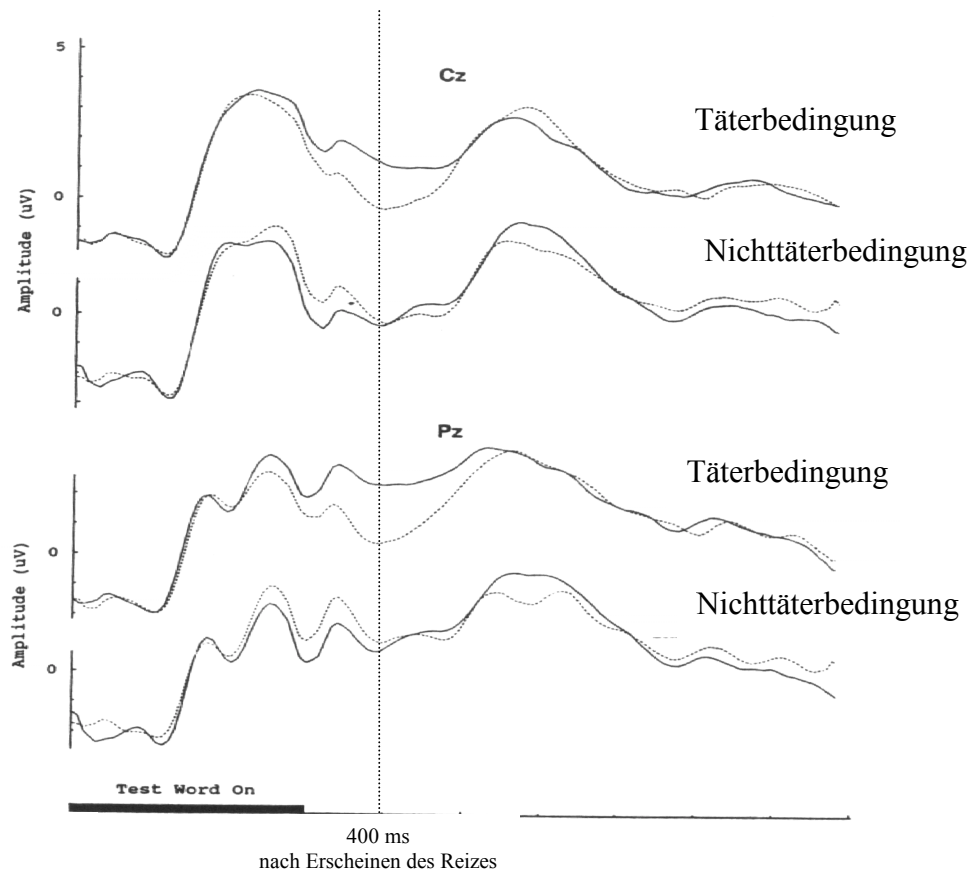


Abbildung 1

Grand Averages für Distraktoren (durchgezogene Linie) und tatspezifische Items (gepunktete Linie) in der Untersuchung von Boaz, Perry, Raney, Fischler & Shuman (1991). Auf der X-Achse ist die Zeit in ms abgetragen. Ca. 400 ms nach Erscheinen des Testworts lässt sich bei Probanden der Täterbedingung im Mittel eine ausgeprägtere N400 auf tatspezifische Reize als auf Distraktoren feststellen. Dies gilt für die Nichttäterbedingung nicht.

3.2.2 Die Verwendung der „Contingent Negative Variation“ (CNV)

Die „Contingent Negative Variation“ (CNV) ist eine langsame negative Potentialverschiebung, die während der Vorbereitungsphase einer Reaktion auftritt (Walter, Cooper, Aldridge, McCallum & Winter, 1964). Diese negative Verschiebung ist über zentralen und frontalen Regionen des Gehirns am ausgeprägtesten (Brunia & Haagh, 1986). Man interpretiert die CNV als Indikator für eine präparatorische Aktivierung der Hirnrinde. Sie besteht, wenn das Erwartungsintervall lang genug ist, aus einer frühen, sogenannten O- oder Orientierungs-Welle, und einer späten, sogenannten E- oder Erwartungs-Welle (Loveless & Sanford, 1974).

Das typische Paradigma zur Erzeugung der CNV sieht ein Ankündigungssignal (S1) vor, das den Beginn einer Vorwarnperiode markiert. S1 kündigt einen weiteren Reiz (S2) an, bei dessen Auftreten der Proband eine motorische Reaktion auszuführen hat. Forscher, die die funktionelle Bedeutung der CNV untersuchen, haben dieses Paradigma auf verschiedene Arten variiert. Diese Variation erfolgte hinsichtlich der Aufgabe, die Probanden bei Erscheinen von S2 zu erfüllen hatten, hinsichtlich der Diskriminierbarkeit von S1, hinsichtlich der Länge der Vorwarnperiode, hinsichtlich der Auftretenswahrscheinlichkeiten beider Stimuli, hinsichtlich der Anwesenheit von Distraktoren etc. (Donchin, Ritter & McCallum, 1978).

Distraktoren, die in dem Intervall zwischen den beiden Reizen S1 und S2 dargeboten werden, führen zu einer niedrigeren CNV-Amplitude und einer längeren Reaktionszeit nach Wahrnehmung von S2. Diese Effekte werden der Abwendung der Aufmerksamkeit zugeschrieben und deshalb als Ablenkungseffekt bezeichnet (Travis & Tecce, 1998).

Hira und Matsuda (1998) wendeten diese Idee zur Täuschungsaufdeckung an. Die Annahme war, dass die Verheimlichung eines zuvor ausgewählten Frauenbildes zu einem Ablenkungseffekt mit einer verminderten CNV-Reaktion führen würde.

Die Stichprobe bestand aus zwölf japanischen Studenten. Jeder Proband nahm sowohl an der Kontroll- als auch an der Experimentalbedingung teil. Als abhängige Variablen dienten mittlere Amplituden, die für das Intervall zwischen dem zuletzt dargebotenen Gesicht und dem S2-Reiz ermittelt wurden.

Die S2-Reize stellten Landschaftsbilder dar, auf deren Darbietung hin eine motorische Reaktion in Form eines Tastendrucks erfolgen sollte. Vor dem S2-Reiz wurden jeweils vier Frauenbilder (S1-Reize) dargeboten. In der Bedingung, in der Probanden Nichttäter darstellten, die nichts verheimlichen wollten, bearbeiteten die Probanden lediglich diese Aufgabe. Anschließend wählten sie ein Frauenbild aus und sollten diese Auswahl während der Ableitung verheimlichen. Dies stellte die Täterbedingung dar. Abgeleitet wurde von Fz⁶, Cz und Pz.

Es wurden Grand Averages für beide Bedingungen über alle zwölf Probanden gebildet. Die mittlere Intervallamplitude vor den S2-Reizen wurde für alle drei

⁶ Ableitort im frontalen Bereich der Sagittallinie.

Ableitorte berechnet. Eine MANOVA für eine unabhängige (Experimental- versus Kontrollgruppe) und drei abhängige Variablen (Ableitorte) wurde berechnet.

Es zeigte sich, dass die CNV in der Bedingung, in der die Probanden eines der Bilder verheimlichen sollten, unterdrückt wurde (siehe Abbildung 2). Die mittleren Intervallamplituden unterschieden sich signifikant zwischen Verheimlichungs- und Kontrollbedingung.

Prinzipiell scheint diese Vorgehensweise also geeignet, um Verheimlichungstendenzen aufzudecken. In der Praxis könnten tatrelevante Informationen die Ablenkungsreize darstellen. Eine Einzelfallauswertung und Klassifikation der Probanden wird nicht beschrieben. Somit fehlen auch Angaben über Grenzwerte und Methoden der Einzelfallbewertung. Insgesamt werden mehr Studien zu diesem Forschungsgebiet benötigt, um diesen Ansatz adäquat auch bezogen auf die Spezifität des Ablenkungseffekts hinsichtlich der Verheimlichungstendenzen beurteilen zu können. Inwiefern die Kenntnis der zu verheimlichenden Information ohne Verheimlichungsintention zu einer Verminderung der CNV führt, ist nicht geklärt.

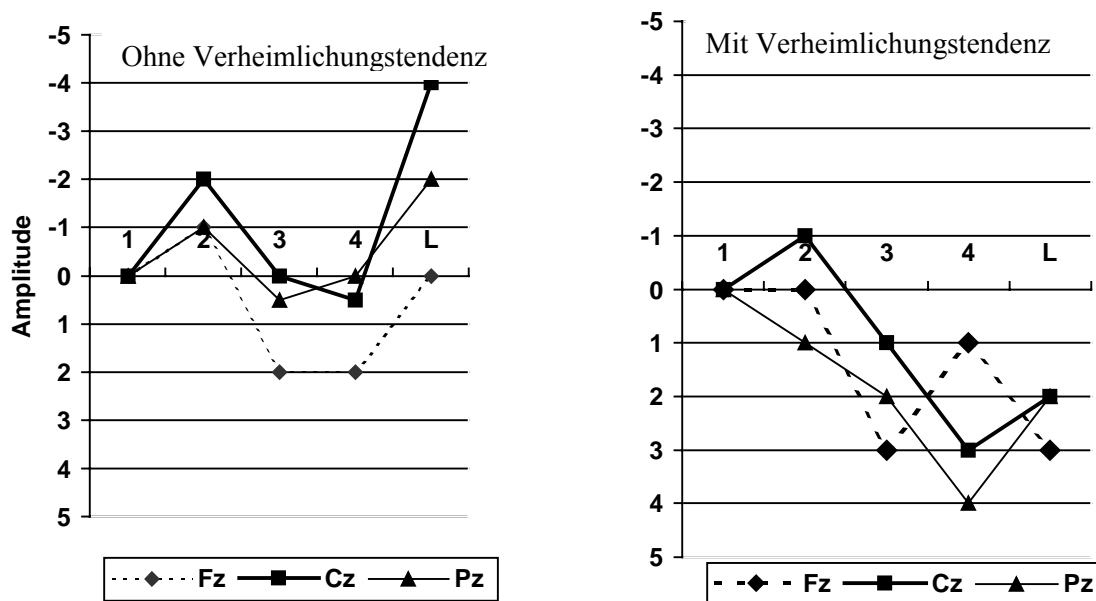


Abbildung 2

Mittlere Intervallamplituden (in μV) für Probanden mit und ohne Verheimlichungstendenz gegenüber einem der vier Frauenfotos für das Intervall vor der Präsentation des jeweiligen Fotos. Die Zahlen 1 bis 4 beziehen sich auf die Frauenbilder, L bedeutet Landschaftsbild. Die Amplituden sind vor der Darbietung von L in der Bedingung mit Verheimlichungstendenz sehr viel weniger negativ.

3.2.3 Die Verwendung der P300-Komponente

Zunächst werden Wirkungslogik und theoretischer Hintergrund der Verwendung der P300-Komponente zur Täuschungsaufdeckung erläutert (3.2.3.1). Anschließend werden relevante Untersuchungen vorgestellt (3.2.3.2). Schließlich wird dieser Untersuchungsansatz unter den Aspekten der Leistungsfähigkeit, der Qualität der Studien und der Anwendbarkeit auf die Praxis bewertet (3.2.3.3).

3.2.3.1 Wirkungslogik und theoretischer Hintergrund.

Vermutlich wurde keinem anderen ereigniskorrelierten Potential so viel Forschungsinteresse entgegengebracht wie der P300-Komponente (Coles & Rugg, 1995). Es handelt sich um eine positive EKP-Welle mit einer Latenz von 300 bis 1000 ms, die als P300, P3 oder P3b bezeichnet wird. Die Latenz variiert mit der kognitiven Verarbeitungszeit, die unter anderem mit der Stimuluskomplexität zusammenhängt. Die Amplitude liegt meist zwischen 5 und 20 μV . Ein Identifizierungsmerkmal ist außerdem die topographische Verteilung der Komponente. Die P300 tritt parietal (Pz) mit der größten Amplitude auf. Bei Fz ist die Komponente am geringsten ausgeprägt und die maximale Amplitude bei Cz liegt meist zwischen diesen beiden Werten (Rosenfeld, 2002).

Die P300 wurde zuerst von Sutton, Braren, Zubin und John (1965) beschrieben. Das Standardparadigma zur Erzeugung dieser Komponente stellt eine einfache auditorische oder visuelle Diskriminationsaufgabe dar. Eine Sequenz unterschiedlicher Reize wird präsentiert, die sich in zwei Klassen einordnen lassen. Die Reize der beiden Kategorien werden mit unterschiedlicher Auftretenswahrscheinlichkeit dargeboten, was als „Oddball-Paradigma“ bezeichnet wird. Die Aufmerksamkeit des Probanden gilt dabei den selteneren Reizen, da diese per Instruktion mit Knopfdruck beantwortet oder gezählt werden sollen. Die P300 zeigt eine größere Amplitude bezüglich der aufgabenrelevanten, seltenen Reize als bezüglich der neutralen Reize. Die Amplitude der P300 ist abhängig von der Zuteilung von Gedächtnis- und Aufmerksamkeitsressourcen, die steigt, wenn ein Reiz selten und bedeutsam ist (Squires, Donchin, Herning & McCarthy, 1977).

Wichtig für den Einsatz der Methode zur Täuschungsaufdeckung ist das direkte Verhältnis der P300-Amplitude zur Bedeutsamkeit der Reize. Ähnlich der Tatwissen-

technik lassen sich seltene Reize als tatspezifisches Wissen und die häufigen Reize als neutrale Items konstruieren. Nur derjenige, der die Tat tatsächlich begangen hat, sollte den tatspezifischen Reiz als solchen erkennen können. Ein Nichttäter sollte einen tatspezifischen Reiz genauso verarbeiten wie einen neutralen Reiz, da ihm die entsprechenden Tatinformationen fehlen. In diesem Sinne stellt die P300 einen Index für die Rekognition von Tatinformationen dar. Die stimulusspezifische Bedeutsamkeit entwickelt sich lediglich für den Täter. Entsprechend sollte sich auf tatspezifische Reize hin eine P300 mit größerer Amplitude als auf die neutralen Reize ergeben. Bei einem Nichttäter ohne Tatwissen sollte dieser Unterschied nicht auftreten.

Das Problem dieser Vorgehensweise besteht darin, dass es keine Gewähr dafür gibt, dass der Proband die Items auch tatsächlich beachtet. Ein Proband mit Täuschungsabsicht könnte eine Klassifikation als Nichttäter erreichen, indem er seine Aufmerksamkeit von den dargebotenen Reizen ablenkt. Die Klassifikation als „Nichttäter“ wäre in dieser Hinsicht zweideutig, da sie sowohl bedeuten könnte, dass der Proband seine Aufmerksamkeit von den Reizen ablenkt, als auch, dass er nicht in der Lage war, die relevanten Reize trotz Aufmerksamkeit als solche zu erkennen. Außerdem besteht auf diese Weise lediglich die Möglichkeit, Reaktionen auf den tatspezifischen Reiz mit denen auf neutrale Reize zu vergleichen. Die Möglichkeit, diese Reaktion einer Reaktion auf relevante Vergleichsreize gegenüberzustellen, besteht nicht.

Gegenüber dieser Zwei-Kategorien-Oddball-Version lässt sich in Annäherung an das Kontrollfragenverfahren eine Drei-Kategorien-Technik konstruieren (siehe Tabelle 1). Neben neutralen und tatspezifischen Items, die sich für einen Nichttäter nicht voneinander unterscheiden, wird ein tatunspezifischer Reiz als Vergleichsreiz eingeführt, der seine Relevanz durch eine gesonderte Aufgabenstellung erhält und gewährleistet, dass die Reize auch tatsächlich beachtet werden. Sowohl Täter als auch Nichttäter erlernen eine Liste von Items, auf die sie mit Knopfdruck reagieren sollen. Die bezüglich der tatunspezifischen Reize auftretende P300 sollte, so die Annahme, mit der bei einem Täter auftretenden P300 bezüglich tatspezifischer Reize vergleichbar sein. Dagegen ist die P300 bezüglich tatspezifischer Reize, die beim Nichttäter auftritt, eher vergleichbar mit der bezüglich der neutralen Items.

*Tabelle 1**Unterschiedliche Itemtypen im Drei-Kategorien-Oddball-Paradigma.*

Itemtyp	
Tatunspezifisch (= U)	Relevant für die Deliktgruppe
Neutraler Reiz (= N)	Irrelevant für die Deliktgruppe und für die vorgeworfene Tat
Tatspezifisch (= S)	Relevant bezüglich der vorgeworfenen Tat

3.2.3.2 Untersuchungen zur P300.

Zunächst wird eine Untersuchung beschrieben, die die Leistungsfähigkeit des Zwei-Kategorien-Oddball-Paradigmas zur Täuschungsaufdeckung zum Gegenstand hat: Rosenfeld, Nasman, Whalen, Cantwell und Mazzeri (1987) nahmen an, dass die Verheimlichung der Auswahl eines Gegenstandes dazu führen würde, dass eine P300 mit ausgeprägter Amplitude auf die Nennung dieses Gegenstandes hin registriert werden könne.

Die Stichprobe bestand aus 16 Studenten. Zehn Probanden wurden der Experimentalgruppe, sechs der Kontrollgruppe zugeordnet. Die P300-Amplitudenwerte in einem Zeitfenster von 400-700 ms nach Erscheinen des Reizes wurden zu einem Wert gemittelt, der als abhängige Variable diente.

Zunächst suchte jeder Proband einen von neun wertvollen Gegenständen (Perlenkette, Fotoapparat, Beutel mit Münzen etc.) aus, den er gerne behalten würde. Diesen Wunsch sollte jeder Proband mit hundert Worten begründen. Die Instruktion beinhaltete, dass die Probanden sich vorstellen sollten, den gewählten Gegenstand gestohlen zu haben, während sie einer PAG unterzogen wurden. Die Probanden sahen auf einem Bildschirm sukzessiv die Bezeichnungen von acht Gegenständen, die zuvor nicht thematisiert worden waren. Diese Bezeichnungen stellten die neutralen Items dar. Den Probanden der Experimentalgruppe wurde außerdem die Bezeichnung des zuvor ausgewählten Gegenstandes dargeboten. Diese Bezeichnung stellte das tatspezifische Item dar. Für Probanden der Kontrollgruppe stellte der neunte Reiz einen neutralen Reiz dar. Der von ihnen ausgewählte Gegenstand befand sich nicht unter den dargebotenen Items. Um zu gewährleisten, dass die Probanden die dargebotenen Reize beachteten, erhielten sie die Instruktion, bei Auftauchen des tatspezifischen Items auf dem

Bildschirm jeweils lautlos „Nein“ zu sagen. Ein EEG wurde von einer nicht näher bezeichneten Elektrodenposition auf der Sagittallinie abgeleitet.

Die Auswertung erfolgte zunächst durch visuelle Inspektion der P300-Amplitudenwerte. Dabei wurden jeweils die P300-Komponenten bezüglich der neun Reize pro Person miteinander verglichen. Außerdem wurde ein t-Test für abhängige Stichproben berechnet, bei dem der mittlere Amplitudenwert bezüglich des relevanten mit dem eines zufällig ausgewählten, neutralen Reizes verglichen wurde.

In der Experimentalgruppe traten bei 9 (90%) Probanden hinsichtlich der P300-Komponente Unterschiede zwischen dem tatspezifischen Item und den neutralen Items auf. Bei einem Probanden der Experimentalgruppe konnte dieser Unterschied nicht festgestellt werden. Die 6 (100%) Probanden der Kontrollgruppe zeigten diese Unterschiede nicht. Der Unterschied zwischen neutralen und relevanten Reizen erwies sich in der Experimentalgruppe als signifikant.

An dieser Vorgehensweise ist zu kritisieren, dass die Probanden aufgefordert wurden, jeweils das tatspezifische Item mit „Nein“ zu beantworten. Dies sollte die Aufmerksamkeit der Probanden auf diese Reize lenken. Es lässt sich nicht voraussetzen, dass ein real Beschuldigter sich an diese Anweisung halten würde, wenn er etwas zu verbergen hätte. Somit wurde das bereits im vorigen Abschnitt erwähnte Aufmerksamkeitsproblem der Zwei-Kategorien-Oddball-Version durch das Procedere dieser Untersuchung im Hinblick auf die Anwendungspraxis nicht zufriedenstellend gelöst.

Eine Drei-Kategorien-Oddball-Version stellten Rosenfeld, Cantwell, Nasman, Wojdac, Ivanov und Mazzeri (1988) vor. Die Autoren nahmen an, dass sowohl ein Reiz, der per Instruktion zu zählen war (tatspezifisches Item), als auch die Nennung eines zuvor „entwendeten“ Gegenstandes (tatspezifisches Item) eine ausgeprägte P300 hervorrufen würde. Bei neutralen Stimuli sollte dies nicht der Fall sein.

Insgesamt nahmen 15 Personen an der Untersuchung teil. Dabei handelte es sich um 14 Studenten und einen professionellen „Polygrapher“. Aufgrund von Datenverlust und unkooperativem Verhalten eines Probanden gingen letztlich Daten von 12 Probanden in die Auswertung ein. Zehn Probanden wurden der Experimental- und fünf Probanden der Kontrollgruppe zugeteilt. Zur Gewinnung der abhängigen Variablen wurde in einem Zeitfenster von 350 bis 800 ms nach dem Reiz der Bereich von 100 ms

ermittelt, der bei Mittelung aller Amplitudenwerte den größten Amplitudenmittelwert ergab.

Die Durchführung entsprach der Vorgehensweise der zuvor beschriebenen Untersuchung. Die Probanden der Kontrollgruppe sahen während der Ableitung wiederum nur neutrale Reize. Den Probanden der Experimentalgruppe wurde außerdem die Bezeichnung des zuvor ausgewählten Gegenstandes dargeboten. Der Unterschied des Vorgehens bestand darin, dass einer der neutralen Reize per Instruktion zum tatsunspezifischen Reiz erklärt wurde. Das Auftreten dieses Reizes sollte von allen Probanden während der Ableitung gezählt werden. Ein EEG wurde von Fz, Cz und Pz abgeleitet.

Die Ergebnisse werden von den Autoren lediglich für die bei Pz abgeleiteten Daten berichtet. Der maximale Amplitudenmittelwert ging als abhängige Variable in eine gruppenstatistische Auswertung ein. Ein t-Test für abhängige Stichproben wurde berechnet. Verglichen wurden dabei die Amplitudenmittelwerte bezüglich eines neutralen mit denen bezüglich des tatspezifischen Items.

In der Experimentalgruppe ergab der t-Test für abhängige Stichproben einen signifikanten Unterschied zwischen den Amplitudenmaßen bezüglich der neutralen und der tatspezifischen Items. In der Kontrollgruppe trat dieser Unterschied nicht auf.

Es wurde keine individuelle Auswertung vorgenommen. Die gruppenstatistischen Ergebnisse sind nur begrenzt aussagekräftig, da die Stichprobe und damit die Wahrscheinlichkeit, signifikante Effekte zu erhalten, klein ist.

Eine methodisch ausgefeiltere Untersuchung legten Rosenfeld, Angell, Johnson und Qian (1991) vor. Die Annahme war, dass in der Experimentalgruppe bezogen auf die Amplituden $B > C$ und in der Kontrollgruppe $B = C$ gelten sollte. Für die Experimentalgruppe stellte B die tatsächlich begangene Tat dar, für Probanden der Kontrollgruppe stellte es ein neutrales Item dar. Vergehen C hatte keiner der Probanden begangen, es wurde als Kontrollitem behandelt, mit dem die Reaktionen auf B verglichen wurden.

Die Stichprobe bestand aus 28 männlichen Psychologiestudenten. Dreizehn Probanden wurden der Experimentalgruppe, 15 Probanden der Kontrollgruppe zugeordnet. Als abhängige Variable dienten die Amplitudenwerte der P300-Komponenten.

Alle Probanden erhielten eine Liste mit 13 Beschreibungen einer antisozialen Handlung oder eines Vergehens (z. B. „in einem Test betrogen“, „ein Fahrrad gestohlen“, „das Geld eines Freundes gestohlen“). Auf dieser Liste sollten sie jeweils angeben, ob sie das entsprechende Vergehen begangen hatten oder nicht. Informationen über die begangenen bzw. nicht begangenen Delikte erhielten die Untersuchungsleiter durch eine den Probanden verborgene Überwachungskamera. Die Probanden wurden in dem Glauben gelassen, dass nur sie den ausgefüllten Bogen zu Gesicht bekämen, zumal sie ihn anschließend mitnehmen oder wegwerfen durften. Diese Information nutzten die Untersuchungsleiter, um ein Itemset mit insgesamt neun Items für die Durchführung einer PAG zu konstruieren. Nachdem die Probanden jeweils die Liste mit den Vergehen bearbeitet hatten, folgte ein Rehearsal-Durchgang. Die Aufgabe der Probanden bestand darin, einen linken Antwortknopf zu betätigen, wenn eine nichtbegangene antisoziale Handlung auf dem Bildschirm präsentiert wurde und den rechten Knopf, wenn eine begangene beschrieben wurde. Die Probanden der Experimentalgruppe sahen die Items, die auch anschließend während der tatsächlichen EEG-Ableitung präsentiert wurden. In der Kontrollgruppe wurde für die Aufgabe während der EEG-Ableitung das begangene Vergehen durch ein nicht begangenes ersetzt. Anschließend sollten die Probanden sich vorstellen, im Rahmen eines Einstellungsgespräches einer PAG bezüglich ihrer Reputation unterzogen zu werden. Das EEG wurde wiederum von Fz, Cz und Pz abgeleitet. Für Probanden der Experimentalgruppe wurde eine der angekreuzten antisozialen Handlungen als tatspezifisches Ereignis verwendet. Bezeichnungen nicht begangener antisozialer Handlungen dienten als neutrale Items. Als tatspezifisches Item wurde die Formulierung „took deception test“ verwendet. Indem die Probanden dieses Item jeweils bejahten, bestätigten sie korrekt ihre Teilnahme an der PAG. Alle übrigen Items wurden per Tastendruck verneint. Für die Kontrollgruppe stellte das tatspezifische Item ein nicht begangenes Vergehen dar.

Ausgewertet wurde sowohl gruppenstatistisch als auch einzelfallorientiert. Die hier dargestellten Ergebnisse beziehen sich auf die bei Pz abgeleiteten Werte. Für die gruppenstatistische Auswertung wurden die maximalen Amplituden der P300-Komponente und der nachfolgenden negativen Welle bestimmt. Die Differenz zwischen beiden Punkten diente als abhängige Variable (Peak to Peak-Amplitude). Eine 2 (Experimental-/Kontrollgruppe) x 2 (tatspezifisches Item/Kontrollitem) Varianzanalyse

wurde durchgeführt. Für die einzelfallorientierte Auswertung verwendeten die Autoren einen vierstufigen Auswertungsalgorithmus, der ein Bootstrapverfahren⁷, das jeweils Peak to Peak-Amplitudenwerte von gemittelten Kurven zwischen tatspezifischen und Kontrollitems vergleicht, mit einbezieht.

Die Varianzanalyse ergab den erwarteten signifikanten Interaktionseffekt zwischen Gruppenzugehörigkeit und Itemtyp (siehe Abbildung 3). In der Experimentalgruppe war die Differenz der Mittelwerte bezüglich der Peak to Peak-Amplitude zwischen tatspezifischem Item und Kontrollitem größer als in der Kontrollgruppe. Mit Hilfe des Bootstrapverfahrens wurden 12 von 13 (92%) Probanden der Experimentalgruppe und 13 von 15 Probanden (86,6%) der Kontrollgruppe korrekt klassifiziert. Das heißt, dass eine falsch negative und zwei falsch positive Klassifizierungen auftraten.

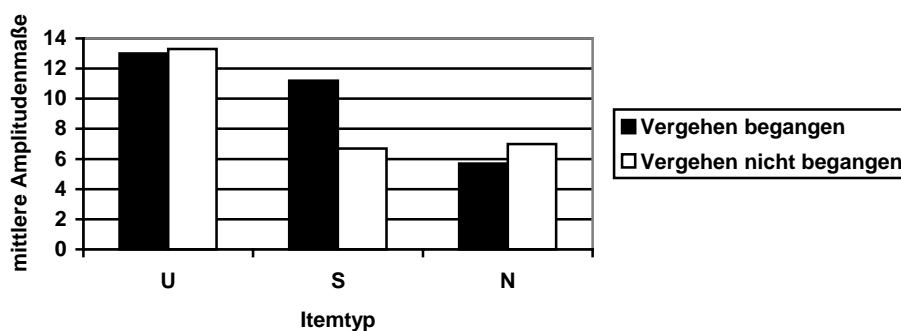


Abbildung 3

Mittlere maximale P300-Amplituden (in μV) der Probanden, die das in Frage stehende Vergehen begangen hatten, und der Probanden, die das in Frage stehende Vergehen nicht begangen hatten (Rosenfeld et al., 1991). U = tatspezifisch, S = tatspezifisch, N = neutral

Problematisch hinsichtlich der internen Validität ist, dass nicht mit Sicherheit bekannt ist, ob die Probanden tatsächlich wahrheitsgemäß angaben, ob sie ein Vergehen begangen hatten oder nicht. Hinsichtlich externer Validität besteht folgendes Problem: Bei der beschriebenen experimentellen Vorgehensweise könnte es sein, dass die Probanden nicht aufgrund der Begehung der Tat und resultierender Verheimlichungstendenzen anhand der P300-Komponenten entlarvt wurden. Die P300-Reaktionen könnten sich durch das Listenausfüllen und die Rehearsal-Prozedur dadurch ergeben

⁷ Voraussetzungsarmes, statistisches Verfahren, das häufig zum Vergleich von EKP-Kurven eingesetzt wird (siehe Kapitel 6.7.3 und Anhang E).

haben, dass das begangene Vergehen bei den vorangehenden Aufgaben wie ein tatunspezifisches Item behandelt wurde. Die Bedeutsamkeit der Items, die ein begangenes Delikt bezeichnen, würde somit nicht aus ihrer Tatspezifität resultieren, sondern daraus, dass sie im Gegensatz zu neutralen Items zuvor aufgabenrelevant waren. Wenn dies der Fall wäre, würden sich die mit Hilfe der vorgestellten Studie gewonnenen Erkenntnisse nicht auf die Verwendung in der Praxis übertragen lassen.

Johnson und Rosenfeld (1992) schalteten in einer ähnlich angelegten Studie diese Konfundierung aus, indem sie zunächst die PAG durchführten und anschließend die Informationen über begangene Delikte einholten. Es galt die gleiche Annahme wie bei Rosenfeld, Angell, Johnson und Qian (1991).

Die Stichprobe bestand aus 42 Studenten. Daten von 31 Probanden gingen in die Auswertung ein. Ausschlusskriterien waren: Keine identifizierbare P300; mehr als 50% der registrierten EEG-Epochen mit Artefakten, wie z. B. Blinks oder Augenbewegungen; die Instruktionen nicht befolgt zu haben und drei oder mehr der als Items verwendeten Vergehen begangen zu haben. Die Notwendigkeit des letzten Ausschlusskriteriums ergibt sich aus den erforderlichen Auftretenswahrscheinlichkeiten im Oddball. Hat ein Proband drei antisoziale Handlungen begangen, resultieren 3 tatspezifische plus 1 tatunspezifischer Reiz. Es ergibt sich eine 4/9 Auftretenswahrscheinlichkeit für bedeutsame Reize. Diese Wahrscheinlichkeit ist zu hoch, als dass man noch von der seltenen Darbietung bedeutsamer Reize in einem Oddball sprechen könnte. Dieses Problem konnte sich bei dieser Studie ergeben, da erst nach der EEG-Ableitung festgestellt wurde, welche Vergehen die Probanden begangen hatten und welche nicht. Die Zuteilung zur Kontroll- oder Experimentalbedingung ergab sich daraus, ob die Probanden nach der PAG angaben, dass sie die in Frage stehende antisoziale Handlung schon einmal ausgeführt hatten oder nicht. Als abhängige Variable wurde das in der früheren Untersuchung der Arbeitsgruppe beschriebene Peak to Peak-Amplitudenmaß verwendet.

Die Durchführung erfolgte bis auf das Listenausfüllen und das Rehearsal analog zu der von Rosenfeld, Angell, Johnson und Qian (1991) beschriebenen Vorgehensweise. Die Rehearsalprozedur vor der EEG-Ableitung war unspezifisch in der Weise, dass das tatspezifische Item nicht anders behandelt wurde als die neutralen Items. Die Probanden sollten jeweils zwei Sätze für jedes Item formulieren und laut vorlesen.

Diese Sätze hatten die Form „Ich habe (Handlung) in den letzten fünf Jahren begangen“ und „Ich habe (Handlung) in den letzten fünf Jahren nicht begangen.“ Die Probanden glaubten, dass ihre Stimme und ihr EEG während dieser Prozedur zwecks Täuschungsaufdeckung analysiert wurden. Außerdem wurden den Probanden verschiedene Fragen zu den Vergehen gestellt, die jedoch nicht das Begehen an sich thematisierten. Auf diese Weise sollte gewährleistet sein, dass die Probanden sich mit den verschiedenen antisozialen Handlungen auseinandersetzen. Für die EEG-Ableitung wurde eine antisoziale Handlung als tatspezifisches Item gewählt, von der die Autoren durch Voruntersuchungen wussten, dass ca. 50% der Studenten sie bejahen würden. Es handelte sich um das Vergehen, einen gefälschten Ausweis verwendet zu haben. Im Anschluss an die EEG-Ableitung füllten die Probanden wiederum unter Beobachtung einen Fragebogen bezüglich der thematisierten Vergehen aus, bei dem sie angaben, ob sie die Vergehen begangen hatten oder nicht.

Zur Einzelfallauswertung wurde der gleiche Algorithmus wie bei Rosenfeld et al. (1991) verwendet.

Von 14 Probanden der Kontrollgruppe wurden 11 Probanden korrekt als Probanden klassifiziert (79%), die das fragliche Vergehen nicht begangen hatten. Die drei Probanden, die falsch positiv klassifiziert wurden, hatten eine plausible Erklärung für dieses Ergebnis, wie z. B. „Ich habe selbst nie eine Unterschrift gefälscht, aber ich musste daran denken, dass meine Schwester dies immer tut.“ Von den 17 Tätern der Experimentalgruppe wurden 13 korrekt klassifiziert (67%).

Die Klassifikation anhand der P300-Amplituden war also auch ohne das möglicherweise konfundierende Rehearsal bei einem Großteil der Probanden korrekt. Die in den beiden Studien verwendete Vorgehensweise bei der Durchführung der PAG birgt jedoch ein weiteres Problem: Die beschriebene PAG-Prozedur der Arbeitsgruppe um Rosenfeld verwendete als tatspezifische Informationen die Bezeichnung eines Vergehens. Ein solches Vorgehen eignet sich lediglich für ein Screeningverfahren, bei dem keine bestimmte Straftat fokussiert wird. In diesem Fall möchte der Proband bezüglich aller genannten Vergehen unschuldig erscheinen. Im Strafverfahren besteht eine Anklage bezüglich eines bestimmten Delikts. Allein aus diesem Grund weist die Nennung der inkriminierten Handlung in diesem Fall eine andere Qualität auf als die Nennung vergleichbarer Delikte. Dass ein aus verschiedenen Deliktbezeichnungen

bestehendes Itemset unter diesen Umständen unverändert angewendet werden kann, ist nicht wahrscheinlich.

Farwell und Donchin haben in einer Untersuchung von 1991 ebenfalls vorgestellt, wie man ein Drei-Kategorien-Oddball zur Täuschungsaufdeckung nutzbar machen kann. Die Untersuchung unterschied sich von den zuvor beschriebenen vor allem dadurch, dass die Reaktionen auf tatunspezifische Items als Vergleichsreize in die Auswertungsprozedur miteinbezogen. Die Annahmen lauteten entsprechend, dass die Kurvenverläufe bezüglich tatspezifischer und tatunspezifischer Items sich bei einem Probanden mit Tatwissen mehr gleichen sollten als die Kurvenverläufe bezüglich tatspezifischer und neutraler Items. Umgekehrt sollten sich bei einem Probanden ohne Tatwissen die Kurvenverläufe bezüglich tatspezifischer und tatunspezifischer nicht so ähnlich sein wie die Verläufe bezüglich tatspezifischer und neutraler Items.

An der Untersuchung nahmen 20 Studenten teil. Jeder Proband wurde einmal einer PAG bezüglich eines Tatszenarios unterzogen, zu dem er Tatwissen besaß und einmal bezüglich eines Szenarios, zu dem er kein Tatwissen besaß. Als abhängige Variable wurden die Amplitudenmaße der P300 verwendet.

Die Probanden begingen ein sogenanntes „Mock Crime“⁸ aus dem Themenbereich „Spionage“. Diese Spionagehandlung umfasste das Auswendiglernen bestimmter relevanter Informationen und die Durchführung der Handlung, die im Überreichen eines Umschlags an eine bestimmte Person bestand. Jeweils zehn Probanden lernten mit Hilfe eines Computerprogramms die Instruktionen für eines von zwei Spionageszenarien, ohne sich darüber bewusst zu sein, dass ein weiteres Szenario existierte. Einen Tag später sollten alle Probanden ihre simulierten Spionagehandlungen während einer EEG-Ableitung verheimlichen. Bezüglich des nicht durchgeführten Spionageszenarios, das ebenfalls Gegenstand eines EEG-Durchgangs war, hatten die Probanden keine tatspezifischen Informationen erhalten. Die auswendig gelernten Instruktionen enthielten sechs Zwei-Wort-Items, die tatspezifische Items darstellten. Zu jedem tatspezifischen Item waren fünf neutrale Items generiert worden, die die gleiche Silbenzahl aufwiesen und der selben semantischen Kategorie angehörten. Die tatunspezifischen Items waren zufällig aus den neutralen ausgewählt worden. Tabelle 2 gibt einen Teil der verwendeten Items wieder. Die tatunspezifischen Items wurden vor

⁸ Scheinverbrechen. Siehe Kapitel 4.3

der EEG-Ableitung auswendig gelernt und mussten während der Ableitung per Instruktion mit Betätigen einer rechten Antworttaste beantwortet werden. Alle übrigen Items sollten mit der linken Taste beantwortet werden. Abgeleitet wurde von drei Ableitorten, nämlich Fz, Cz und Pz.

Die Einzelfallauswertung erfolgte anhand eines korrelativen Bootstrapverfahrens (siehe Kapitel 6.7.3).

Tabelle 2

Ein Ausschnitt der in der Untersuchung von Farwell und Donchin (1991) verwendeten Itemlisten.

Szenario 1			Szenario 2		
tatspezifisch	tatunspezifisch	neutral	tatspezifisch	tatunspezifisch	neutral
Blue Coat	Green Hat	Brown Shoes Red Scarf Gray Pants Black Gloves	White Shirt	Green Tie	Beige Suit Red Vest Tan Belt Black Socks
Phil Jenks	Tim Howe	Ray Snell Neil Rand Gene Falk Ralph Croft	Dale Spence	Wayne Brant	Glenn Platt Walt Rusk Tod Ames Earl Dade
...

Erläuterung zu Tabelle 2: Szenario 1 und 2 beziehen sich auf zwei unterschiedliche Versionen des Spionageszenarios, bei denen unterschiedliche Informationen eine Rolle spielten. „Blue Coat“ und „Phil Jenks“ könnten beispielsweise tatspezifische Informationen darstellen, da im Rahmen des Spionageszenarios 1, einem Phil Jenks in einem blauen Mantel der Umschlag überreicht werden musste (Die Autoren geben keine Informationen darüber). Tatunspezifische und neutrale Items beziehen sich nicht auf das Szenario und wurden hinsichtlich gleicher Silbenzahl und gleicher semantischer Kategorie passend zu den tatspezifischen Items generiert.

Die EKP-Verläufe der Täterbedingung und der Nichttäterbedingung sind in Abbildung 4 und 5 wiedergegeben.

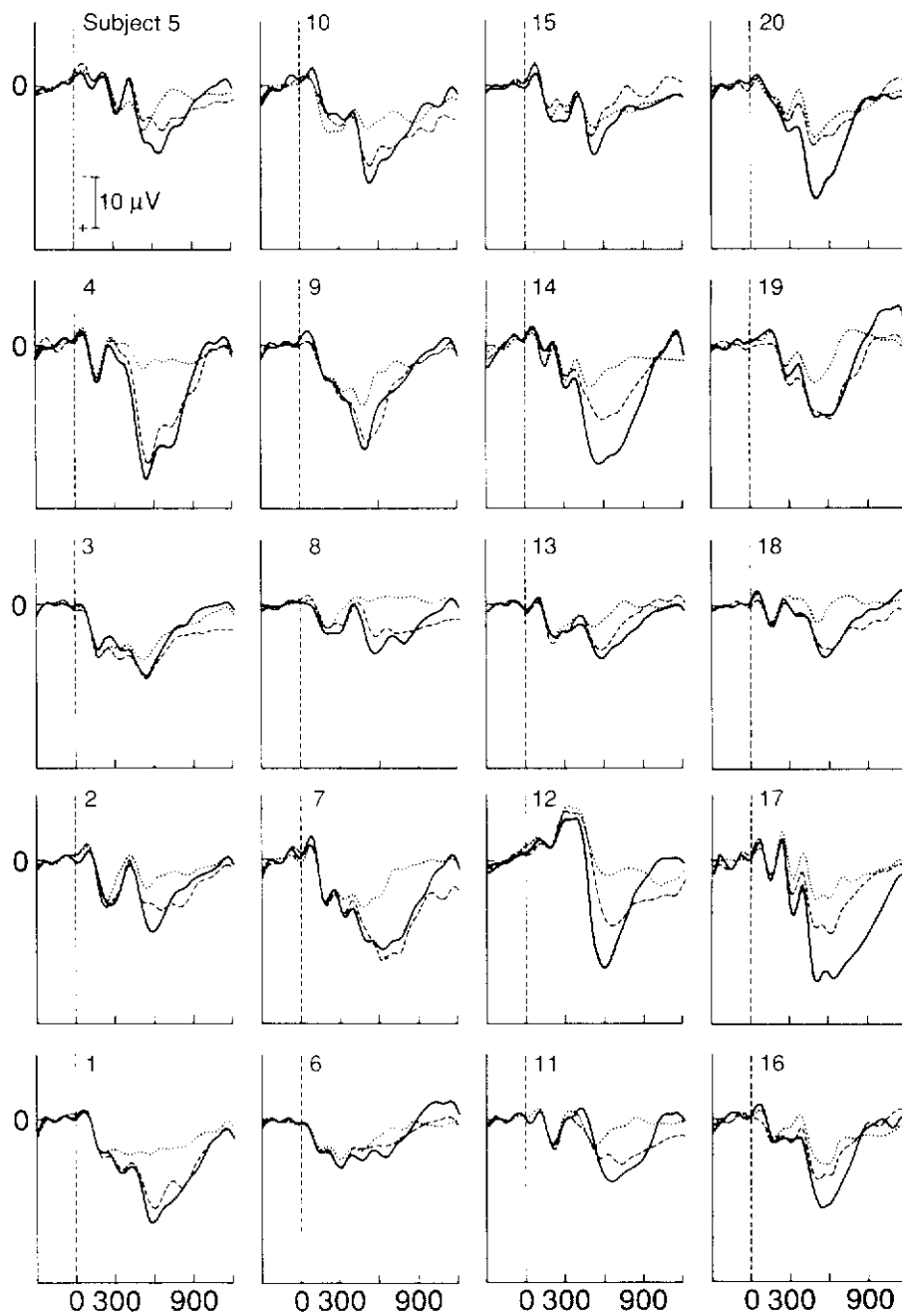


Abbildung 4

EKP-Kurven für jeden der zwanzig Probanden in der Täter-Bedingung bei Farwell und Donchin (1991). Die Verläufe bezüglich tatspezifischer Items ähneln denen bezüglich tatunspezifischer Items.

(---- tatspezifische, neutrale, — tatunspezifische Reize).

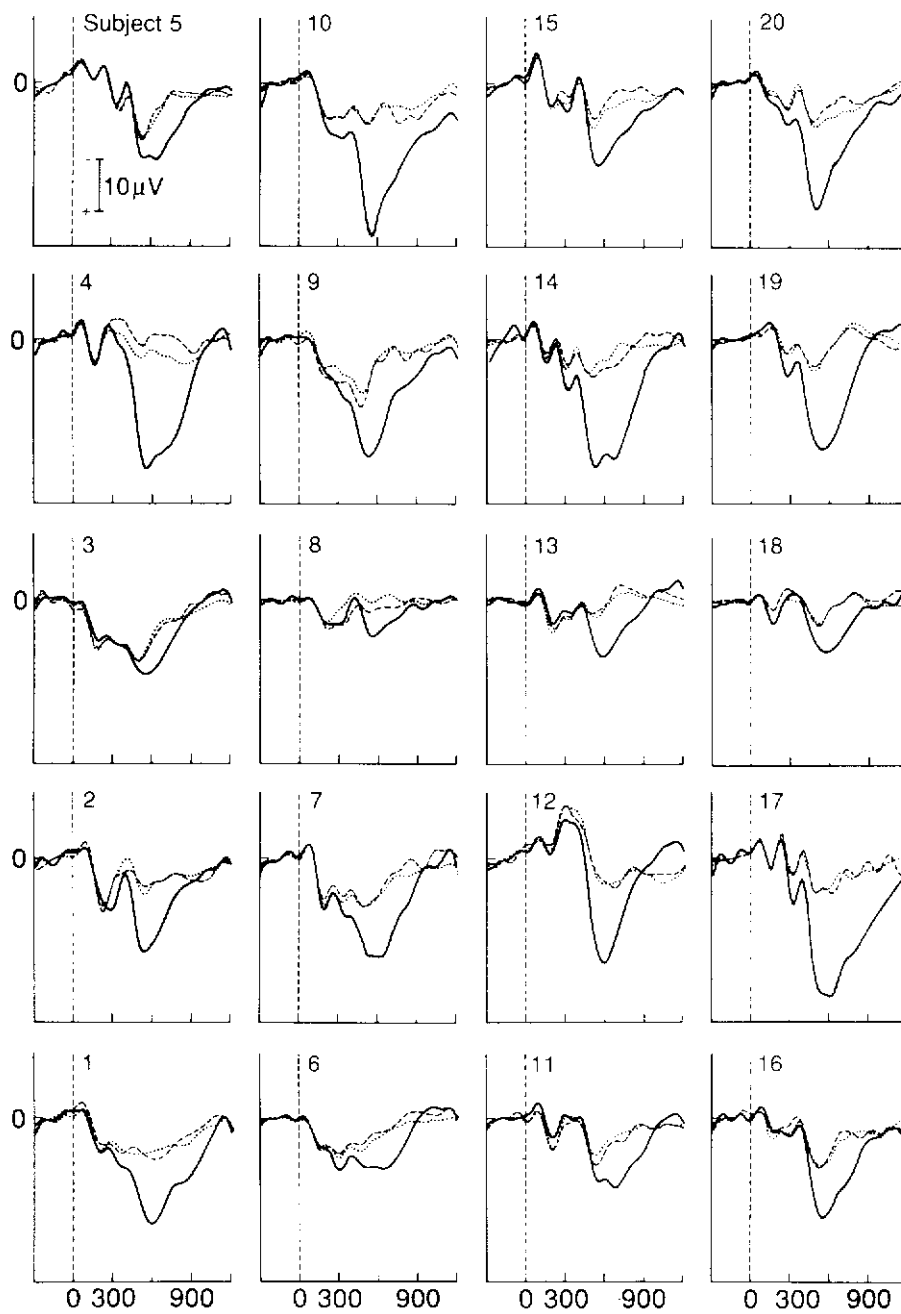


Abbildung 5

EKP-Kurven für jeden der zwanzig Probanden in der Nichttäter-Bedingung bei Farwell und Donchin (1991). Die Verläufe bezüglich tatspezifischer Items ähneln denen bezüglich neutraler Items.

(---- tatspezifische, neutrale, — tatunspezifische Reize).

Mit Hilfe der P300-Komponenten konnte sicher zwischen den Bedingungen unterschieden werden. Die Treffsicherheit lag bei 100%, bei einer Rate von 12,5% unentscheidbarer Fälle.

Die Autoren konnten also mit großer Sicherheit diejenigen Probanden identifizieren, die mit tatspezifischer Information versehen worden waren. Die externe Validität wird dadurch verringert, dass das Tatwissen in Form von Instruktionen auswendig gelernt wird. Dabei wird dieses Wissen intentional erworben. In der Praxis werden die Informationen, die später zur Generierung von tatspezifischen Items verwendet werden, inzidentell bei Begehung der Tat erworben. Diesen Lernprozessen unterliegen teilweise unterschiedliche neuronale Prozesse (Baldwin & Kutas, 1997). Es ist nicht bekannt wie sich diese Unterschiede auf die Klassifizierbarkeit anhand der P300 auswirken.

Allen, Iacono und Danielson (1992) beziehungsweise Allen und Iacono (1997) sehen den Einsatzbereich des Drei-Kategorien-Oddball-Paradigmas zur Täuschungsaufdeckung im Bereich der Neurologie. Hier könnte es nutzbringend angewendet werden, um festzustellen, ob Patienten Gedächtnisstörungen simulieren oder aggravieren. Wenn das Vorhandensein von Gedächtnisinhalten von einem Patienten verneint wird, könnte die P300 dazu dienen, dies zu bestätigen oder dies zu widerlegen. Die Informationen, deren Erinnerbarkeit überprüft werden soll, werden dabei als tatspezifische Informationen behandelt. Werden außerdem tatunspezifische und neutrale Items während einer EEG-Ableitung dargeboten, gelten die gleichen Annahmen wie in der Studie von Farwell und Donchin (1991).

Die Stichprobe bestand im Rahmen dieser Simulationsstudie aus sechzig Studenten. Alle Probanden wurden mit den tatspezifischen Informationen versehen. Für die Kontrollbedingung wurden die gleichen Daten noch einmal ausgewertet, wobei die Reaktionen auf sechs neutrale Items als Reaktionen auf tatspezifische Items behandelt wurden. Als weitere unabhängige Variable wurde die Motivationsbedingung variiert. Es gab drei unterschiedliche Instruktionsversionen. Eine bestand lediglich in der Aufforderung, die erste Liste zu verheimlichen. Die zweite Instruktionsbedingung stellte die explizite Aufforderung zur Täuschung dar. In der dritten Bedingung wurden den Probanden eine Belohnung für einen erfolgreichen Täuschungsversuch versprochenen. Jeweils 20 Probanden wurden einer dieser Motivationsbedingungen

zugeteilt. Als abhängige Variable wurden die Amplitudenwerte der P300 verwendet. Die Probanden lernten eine Liste mit sechs Worten (tatspezifische Items), die der gleichen semantischen Kategorie (z. B. Tiere, Werkzeuge, Metalle) angehörten. Als Lernkriterium galt die zweimalige korrekte Wiedergabe der Liste in normaler und in umgekehrter Reihenfolge. Es folgte eine dreißigminütige Pause, nach der eine weitere Liste mit sechs Worten (tatspezifische Items) einer anderen semantischen Kategorie bis zum Erreichen des Kriteriums gelernt wurde. Während einer unmittelbar anschließenden EEG-Ableitung waren die Probanden gehalten, das Erlernen der ersten Wortliste zu verheimlichen. Items aus fünf neuen Kategorien dienten als neutrale Items. Abgeleitet wurde von Cz und Pz und zwei weiteren nicht näher bezeichneten Ableitorten.

Die Darstellung der Ergebnisse beschränkt sich auf die Ableitung bei Pz. Es wurden drei unterschiedliche Auswertungsstrategien für die Klassifizierung im Einzelfall verwendet: Eine Strategie, die das Bayes-Theorem anwendet, ein Bootstrapverfahren, bei dem Amplitudenmaxima verglichen wurden (siehe Kapitel 6.7.3) und ein Bootstrapverfahren, bei dem wie bei Farwell und Donchin (1991) Kreuzkorrelationen verglichen wurden. Für die gruppenstatistische Auswertung wurde eine 3 (Itemtyp/Within) x 3 (Motivationsbedingung/Between) Varianzanalyse mit Messwiederholung unter Verwendung der Amplitudenmaße der P300 als abhängiger Variable berechnet. Außerdem wurden Signalentdeckungskurven bezüglich der Klassifikationsgenauigkeit ermittelt und ein Flächenmaß („Area under Curve“, AUC-Wert) als Index für die Klassifikationsgüte berechnet (siehe Kapitel 6.7.3).

Die Varianzanalyse ergab einen Haupteffekt für den Itemtyp und außerdem einen Haupteffekt für die Motivationsbedingung, aber keine signifikanten Interaktionen zwischen Itemtyp und Motivationsbedingung. Die korrekte Klassifikationsrate der Probanden in der Kontrollbedingung lag für das korrelative Verfahren bei 72%. Für die Experimentalbedingung betrug die Trefferquote 87%. Signalentdeckungs-Kurven (ROC-Kurven) erbrachten „Area Under Curve“-Werte (AUC-Werte) über .90, was für eine sehr gute Klassifikationsgenauigkeit spricht. Ein Vergleich der ROC-Kurven nach Motivationsbedingung ergab, dass mit steigender Motivation zu täuschen auch die Wahrscheinlichkeit einer korrekten Klassifikation stieg. Ein Vergleich der Auswertungsmethoden ergab eine signifikante Unterlegenheit der Verwendung von

Amplitudenvergleichen gegenüber dem korrelativen Verfahren und dem Bayes-Verfahren.

Die Untersuchung liefert wertvolle Hinweise auf die Bedeutung der Motivation für die P300. Die Übertragbarkeit der Klassifikationsergebnisse auf den Anwendungskontext der PAG wird durch das artifizielle Untersuchungsparadigma (Listenerlernen) eingeschränkt.

Zhou, Yang, Liao und Zou (1999) untersuchten im Rahmen eines Mock-Einbruchs-Paradigmas die Anwendbarkeit der P300-Komponente zur Täuschungsaufdeckung.

Ihre Stichprobe bestand aus 20 männlichen Studienanfängern. Zehn Personen aus der Experimentalgruppe simulierten in dieser Studie einen Einbruch, 10 Probanden der Kontrollgruppe begingen diesen Einbruch nicht. Als abhängige Variablen dienten die maximalen P300-Amplituden (in einem Fenster von 300 bis 800 ms nach dem Reiz) beziehungsweise die Flächen unter der EKP-Kurve.

Probanden der Experimentalgruppe brachen, mit Handschuhen und Werkzeug ausgestattet, die Tür eines Zimmers auf, knackten instruktionsgemäß einen Tresor und entwendeten eine Goldkette. Um zu gewährleisten, dass die Probanden den Hergang gut erinnerten, mussten sie diesen anschließend einem Untersuchungsleiter aufsagen. Ein bis zwei Tage später fand die EEG-Ableitung statt. Die Probanden der Kontrollgruppe hatten am ersten Tag frei, ihre Teilnahme begann erst mit der EEG-Ableitung am zweiten Tag. Vor der EEG-Ableitung wurden allen Probanden 8 Fotos vom Einbruchsszenario gezeigt, die als tatspezifische Reize dienten. Während der Ableitung wurden außerdem 32 Bilder als neutrale Reize dargeboten. Als tatspezifische Reize dienten weitere 8 Fotos aus dem Einbruchsszenario, die die Probanden zuvor noch nicht gesehen hatten. Die Instruktion lautete, jedes Mal dann einen Knopf zu betätigen, wenn ein vom begangenen Einbruch bekanntes Foto gezeigt wurde. Das EEG wurde von Fz, Cz und Pz abgeleitet.

Paarweise durchgeführte t-Tests wurden getrennt für die Experimental- und Kontrollgruppe zwischen tatspezifischen Reizen und neutralen Reizen berechnet.

In der Experimentalgruppe ergab sich ein signifikanter Unterschied zwischen neutralen und tatspezifischen Items. Für die Kontrollgruppe zeigte sich dieser Unterschied erwartungsgemäß nicht. Der Unterschied zwischen tatspezifischen und

neutralen Items war in der Experimentalgruppe sowohl bei Verwendung der maximalen Amplitude, als auch bei Verwendung des Integrals als abhängige Variable signifikant.

Die Aussagekraft dieser Untersuchung wird durch den geringen Stichprobenumfang erheblich gemindert. Die Verwendung von Fotos vom Tatort schafft insofern größere Realitätsnähe, als hier die Notwendigkeit des Erlernens von Itemlisten zum Einprägen der tatspezifischen Items entfällt. Die Verwendung von Bildmaterial für die Verwendung im Realfall dürfte allerdings im Hinblick auf Sexualstraftaten noch schwieriger zu realisieren sein als die Verwendung von Wortmaterial. Eine Auswertungsprozedur, die eine Klassifikation im Einzelfall ermöglichen würde, fehlt in dieser Arbeit.

Zhou, Yang, Liao und Zou stellten 2000 eine weitere Untersuchung vor, die der ersten in beinahe allen Punkten gleicht. Diesmal bestand die Stichprobe aus 30 Probanden und es gab eine zusätzliche Experimentalgruppe, die den Diebstahl nicht durchführte, sondern lediglich den Tatort besuchte. Es gab also eine zusätzliche Gruppe, die Nichttäter darstellte, aber Informationen über den Tatort besaß. Die t-Tests ergaben signifikant größere Amplituden für die tatspezifischen Reize in der Experimentalgruppe, die das Mock Crime beging, als in der Kontroll- und in der Besuchergruppe. Das bedeutet, dass „unschuldig“ erworbenes Tatwissen nicht zu einer falsch positiven Klassifikation führte. Die Stichprobe ist jedoch zu klein und die verwendeten Auswertungsmethoden sind unzureichend, um fundierte Schlussfolgerungen aus diesen Ergebnissen ziehen zu können.

Charakteristika je einer Untersuchung der vier Studiengruppen werden in Tabelle 3 im Überblick dargestellt.

Tabelle 3

Verschiedene Aspekte von Studien, in denen das Drei-Kategorien-Oddball-Paradigma zur Täuschungsaufdeckung verwendet wurde.

Aspekte der/des	Farwell & Donchin (1991)	Rosenfeld et al. (1991)	Allen & Iacono (1997)	Zhou, Yang, Liao & Zou (1999)
Stichprobe: Alter (in Jahren)	Studenten 19-27	Studenten, männl. 18-22	Studenten ?	Studenten, männl. 19-20
Stichprobengröße	20 (40)	32	60 (120)	20
Durchführung: Täuschungsszenario	Mock Crime/ Spionage	Antisoziale Handlungen	Lernen von Wortlisten	Mock Crime/ Einbruch
Wie wird tatspezifische Info erworben?	Auswendig- lernen und Tatbegehen	Tatbegehen	Auswendiglernen	Begehen u. Rekognizieren
Zeitabstand zwischen Tatszenario und PAG	1 Tag	Unmittelbar nach Wiederholungs- prozedur	Ca. 1 Stunde	1 oder 2 Tage
Reizmaterial	36 Zwei-Wort- Items	9 Kurzsätze	42 Wörter aus 7 semantischen Kategorien	48 Bilder
Tatunspezifische Reize	Ohne Tatzusammen- hang	“Took deception- test”	Eine semantische Kategorie z. B. Kleidung	Bilder vom Einbruchsort
Tatspezifische Reize	Spionage	Antisoziale Handlung	Eine semantische Kategorie z. B. Tiere	Bilder vom Einbruchsort
Anzahl der Darbietungen	144 x 3 pro Bedingung	9 x 12 = 108	42 x 5 = 210	48 x 5 = 240
Darbietungszeit (ms) Interitemintervall Insgesamt	300 1550 ca. 45 Min	?	306 1,999 ca. 45 Min.	Je 500 800 6 Min.
Designs:	UV: Itemtyp Schuldbe- dingung AV: Amplituden- maße	UV: Itemtyp Schuldbedingung AV: Amplitudenmaße	UV: Itemtyp Schuldbedingung Motivation AV: Amplitudenmaße	UV: Itemtyp Schuldbedingung AV: Amplitudenmaße Integrale
Auswertung:	Einzelfall: Bootstrap/ Korrelativ	Einzelfall: Amplitudenvgl. Gruppen- vergleiche: ANOVA	Einzelfall: Bootstrap/ korrelativ + Amplitudenvgl + Bayes Gruppenvgl.: ANOVA	Gruppen- vergleiche: t-Test
Ergebnisse:	100% korrekt klassifiziert 12,5% unentscheidbar	Signifikante Interaktions- effekte Hitrate 89,3%	AUC zwischen .94 und .98 Signifikante Interaktions- effekte	EG höhere max. Amplituden u. Integrale bei tatspezifischen Reizen als bei irrelevanten
Prozent korrekt Täter Nichttäter	90% 85%	92% 86,6%	87% (korrel. 72% Verfahren)	Keine Klassifikation

3.2.3.3 Bewertung der Verwendung der P300 zur Täuschungsaufdeckung.

Zusammenfassend lässt sich sagen, dass die Studien, die bisher zur Täuschungsaufdeckung mit Hilfe der P300-Komponenten vorliegen, sehr hohe Trefferquoten erbrachten. Insgesamt belegen die Ergebnisse, dass die Kurvenverläufe bezüglich tatspezifischer und tatunspezifischer Items sich bei einem Probanden mit Tatwissen mehr gleichen als die Kurvenverläufe bezüglich tatspezifischer und neutraler Items. Umgekehrt sind sich bei einem Probanden ohne Tatwissen die Kurvenverläufe bezüglich tatspezifischer und tatunspezifischer nicht so ähnlich wie die Verläufe bezüglich tatspezifischer und neutraler Items. Ebenso findet man bestätigt, dass bei Probanden mit Tatwissen, die P300-Amplituden auf tatspezifische Items größer ausfallen als auf neutrale Items. Dieser Unterschied tritt bei Probanden ohne Tatwissen nicht auf.

Da die P300 bisher nicht in der Praxis der PAG eingesetzt wurde, handelt es sich bei den Studien um Analogstudien. Die ökologische Validität der vorgestellten Studien ist bezüglich verschiedener Aspekte zu kritisieren. Besonders problematisch ist ein Übertrag der Ergebnisse auf den Bereich inkriminierter Sexualstraftaten. Ein Teil der Studien fokussiert Screeningverfahren, ein anderer diagnostische Tests zur Überprüfung von Simulation im klinisch neurologischen Bereich als Anwendungsbereich. Weitere fokussierte Anwendungsbereiche sind Spionage oder Einbruch. Ein Szenario, dass die Situation im Falle einer inkriminierten Sexualstraftat simuliert, sollte erstellt werden, um die Treffsicherheit der Methode unter diesen Bedingungen zu überprüfen.

Bezüglich der Anwendbarkeit könnte ein Vorteil der Verwendung der P300 zur Täuschungsaufdeckung darin bestehen, dass sie gegenüber „unschuldig erworbenem Tatwissen“ robust ist. Einen ersten Hinweis darauf liefert die Untersuchung von Zhou, Yang, Liao und Zou (2000). Eine methodisch bessere Untersuchung sollte diese Befunde jedoch zunächst erhärten, bevor daraus weitere Schlüsse bezüglich der Anwendbarkeit gezogen werden.

3.3 Allgemeine Bewertung der Verfahren

Die Urteilsbegründung des BGH (siehe Kapitel 1) lässt bezüglich der Methoden der PAG und hier besonders bezüglich des Kontrollfragenverfahrens, den Eindruck entstehen, dass es sich um wissenschaftlich nicht belegte und nicht belegbare Verfahren

handelt, die daher von der Anwendung ausgeschlossen sind. Wie oben dargelegt wurde, ergeben sich sowohl aus der Wirkungslogik (z. B. Formulierung geeigneter Kontrollfragen), aus der Forschungslage (z. B. nicht vergleichbare Deliktthematik) und auch aus der Anwendungspraxis (z. B. Gefahr der willentlichen Verfälschung durch den Probanden) einige derzeit noch ungelöste Probleme, so dass diesbezügliche Kritik berechtigt ist. Andere Kritikpunkte, die vom BGH aufgegriffen wurden, betreffen Behauptungen, die im Rahmen der PAG nicht aufgestellt werden und somit nicht greifen. So kann den Verfahren der PAG nicht zum Vorwurf gemacht werden, dass sie nicht geeignet sind, eine „specific lie response“ zu registrieren, weil sie eine solche nicht postuliert. Auch ist nicht richtig, dass eine „wesentliche Prämisse des Kontrollfragenverfahrens“ darin besteht, dass bezüglich des Erregungsgrades zwischen Täter und Nichttäter ein Unterschied der Reaktionen auf relevante Fragen besteht (BGH, 1999, S.214). Die Annahme ist vielmehr, dass das Verhältnis der Reaktionen auf Kontroll- und auf relevante Fragen sich bei Tätern und Nichttätern unterscheidet. Auch sollte daraus, dass die vorliegenden Feld- und Analogstudien als unzureichend betrachtet werden, nicht geschlossen werden, dass die Methoden der PAG der Forschung prinzipiell nicht zugänglich sind. Mit den Problemen der Feld- und Laborexperimente und der Übertragung ihrer Ergebnisse auf die Praxis beschäftigt sich Kapitel 4.

Obwohl frühere prozessrechtliche Bedenken und Bedenken gegenüber eventuellen Menschenrechtsverletzungen in diesem Urteil des BGHs nicht geteilt werden, spielten diese dennoch eine lange Zeit eine wesentliche Rolle. Die Öffentlichkeit begegnet der PAG noch immer mit großer Skepsis. Vom wissenschaftlichen Standpunkt her ist festzustellen, dass die PAG im Vergleich beispielsweise zur inhaltsorientierten Glaubhaftigkeitsbegutachtung sehr viel seltener Forschungsgegenstand ist. Die wenigen Forschungsgruppen, die sich mit der Thematik beschäftigen, sind sich bezüglich der korrekten Bezeichnung der hier als PAG bezeichneten Methode nicht einig. Insgesamt kann festgestellt werden, dass das Potential bezüglich der Methoden der PAG bisher in der Forschung nicht ausgeschöpft wurde. Bei Betrachtung der Gesamtmenge der Untersuchungen fällt auf, dass bei der Vielzahl möglicher Parameter, die aus vegetativen Begleitprozessen abgeleitet werden können, nur ein geringes Spektrum für die PAG angewendet wird (Scholz, 1999).

Probleme wie die methodisch zufriedenstellende Registrierung der Bedeutsamkeit von Kontroll- und relevanten Fragen werden bisher noch kaum angegangen (siehe Kapitel 3.1.1.5).

Selbst, wenn man die meisten Probleme der herkömmlichen Verfahren lösen könnte, hat die Verwendung peripherphysiologischer Parameter einen weiteren prinzipiellen Nachteil: Das differentielle Auftreten peripherphysiologischer Reaktionen auf relevante beziehungsweise neutrale oder Kontrollfragen bei TWT und KFT wird vielfach mit der Orientierungsreaktion sensu Sokolov (1967) bzw. Gray (1982) erklärt. Demnach wird eine Orientierungsreaktion ausgelöst, wenn eine Diskrepanz zwischen gespeicherten Informationen und wahrgenommenen Informationen besteht. Das heißt, die Orientierungsreaktion setzt ein Vergleichssystem voraus, das bestehende mit neu erworbenen Informationen vergleicht (Öhman, Hamm & Hugdahl, 2000). So stellt die Orientierungsreaktion auch ein zentrales Konzept im „Systemmodell der Entstehung autonomer Reaktionen bei der PAG“ nach Steller (1987) dar. Im Rahmen dieses Modells spielt ihr Auftreten eine Rolle als unspezifischer „Neuigkeitsdetektor“. Wichtiger für die registrierbaren Reaktionsunterschiede ist ihre Funktion als „Bedeutsamkeitsdetektor“. Die Stärke der individuellen Signifikanz eines Reizes soll dabei positiv mit der Höhe der korrespondierenden physiologischen Erregung korrelieren. Definitiver Bestandteil für das Vorliegen einer Orientierungsreaktion ist die Habituation dieser Reaktion. Dies ist jedoch kein Bestandteil einer Täuschungsreaktion. Nach Scholz (1999) lassen sich die differentiellen Reaktionen, die bei der PAG auf unterschiedliche Reize hin auftreten, nicht mit Hilfe der Orientierungsreaktion erklären, da Habituation bezüglich relevanter Reize bei der fraglichen schuldigen Person nicht erwartet werden kann.

Furedy, Davis und Gurevich (1988) weisen darauf hin, dass im Felde der PAG differentielle emotionale Reaktionen, Besorgtheit, Erregtheitszustände oder, wie im Falle der P300, Unterschiede in der Auftretenshäufigkeit von relevanten und neutralen Reizen herangezogen werden ohne auf das Konzept der Täuschung zurückgreifen zu müssen, um die unterschiedlichen physiologischen Reaktionen auf unterschiedliche Reize bei TWT und KFT zu erklären.

Von Scholz (1999) wird darauf hingewiesen, dass mit der Registrierung peripherphysiologischer Prozesse primär subkortikale Aktivierung erfasst wird.

Täuschung ist allerdings ein intentionaler kognitiver Prozess. Da es sich um ein intentionales kognitives Phänomen handelt und dieses nur bei kontrollierter Aufmerksamkeit beobachtet werden kann, wenn ein limitiertes Kapazitäts-Kontroll-System aktiv ist, bieten sich als abhängige Variablen langsame Hirnpotentiale und ereigniskorrelierte Potentiale an. Bei der Verwendung der CNV (contingent negative variation) wird ausgenutzt, dass eine Negativierung bioelektrischer Felder in Vorbereitung auf eine Entscheidung zu registrieren ist. Die Variabilität der Komponente wird mit Erwartungseffekten, mentalem Priming, Assoziationseffekten und der Aufmerksamkeit in Zusammenhang gebracht (Fabiani, Gratton & Coles, 2000). Endogene Komponenten ereigniskorrelierter Potentiale erfordern Aufmerksamkeit und eine aktive Informationsverarbeitung (Fabiani, Gratton & Coles, 2000). Sie sind damit prinzipiell geeignete Variablen, um Begleitprozesse von Täuschung abzubilden.

Ein Vorschlag wäre es daher, die Methoden, die bei Registrierung peripher-physiologischer Variablen angewendet werden, durch Methoden zu ergänzen, die die Registrierung ereigniskorrelierter Potentiale erfordern. Auf diese Weise könnten die herkömmlichen Methoden der PAG wertvolle Ergänzung erfahren.

Die Verwendbarkeit der P300-Komponente im Drei-Kategorien-Oddball zur Täuschungsaufdeckung ist im Vergleich zur Verwendung der N400 oder der CNV besser belegt (siehe Kapitel 3.2). Sie wurde diesbezüglich öfter und in verschiedenen Kontexten untersucht. Dass die P300 einen direkten Zusammenhang zur Bedeutsamkeit der Reize aufweist (siehe Kapitel 3.2.3.1), macht sie besonders interessant für die Ermittlung von Täuschungsverhalten. Trotz der Tatsache, dass die Belege vielversprechend sind, weisen die vorliegenden Untersuchungen im Hinblick auf die Anwendung im Fall einer tatsächlichen Beschuldigung in einem Sexualstraftatsfall Mängel hinsichtlich externer Validität auf (siehe Kapitel 3.2.3.2 und 3.2.3.3).

Die Täuschungsaufdeckung mit Hilfe der P300 umfasst sowohl Merkmale der Tatwissen- als auch der Kontrollfragentechnik. Die Verwendung von Kontrollstimuli in Form der tatunspezifischen Items gleicht der Vorgehensweise bei der Kontrollfragentechnik. Bei beiden Verfahren besteht die Möglichkeit, die Reaktion auf den relevanten, tatspezifischen Reiz sowohl mit neutralen als auch mit dem tatunspezifischen Reiz beziehungsweise dem Kontrollreiz zu vergleichen. Die Itemformulierung gleicht eher der Tatwissenteknik. Auch unterscheiden sich Täter von

Nichttätern dahin gehend, dass sie entweder entsprechendes Tatwissen besitzen oder nicht. Damit besteht für die Verwendung der P300-Komponente das gleiche Problem, das auch die TWT hat. Die Voraussetzung, dass relevantes Wissen nur dem Täter zugänglich ist, lässt sich im Realfall zum Zeitpunkt der Hauptverhandlung, wenn eine Person einer Sexualstraftat beschuldigt wird, nicht erfüllen. Diese Vorgehensweise erscheint also für den Anwendungskontext nicht angemessen. Es sei denn, es würde sich zeigen, dass das Oddball-Verfahren gegenüber tatspezifischem Wissen, das von einem Nichttäter erworben wird, robust ist.

Steller und Dahle (1999) stellen die Behauptung auf, es dürfte in Fällen von sexuellem Missbrauch schwer sein, gleichwertige sexualbezogene Items zu generieren, die in ihrem Belastungsgehalt den relevanten oder tatspezifischen Items gleichwertig sind. Allerdings ist nicht einzusehen, warum dies schlicht unmöglich sein soll. Wäre ein tatspezifisches Item in einem Oddball-Paradigma „unter Kleid fassen“, ließen sich entsprechend, „in Hose fassen“, „unter Rock fassen“, „in Ausschnitt fassen“, als neutrale Items und „unter Hemd fassen“ als tatunspezifisches Item formulieren, wenn diese Items in dem Fall keine inkriminierten Handlungen umfassten. Die Probleme entstehen vielmehr dadurch, dass es schwierig ist, Informationen zu finden, die dem Nichttäter zum Zeitpunkt der Hauptverhandlung nicht bekannt sind. Die Itemformulierung der Tatwissenteknik kann aber prinzipiell übernommen werden. Das heißt, es werden tatspezifische Details als relevante Reize verwendet anstatt einer direkten Beschuldigung.

Falls die P300 auch dann noch sicher zwischen den Beschuldigungsbedingungen differenziert, wenn sich Probanden in den beiden Bedingungen nicht durch das Wissen um die tatspezifischen Reize unterscheiden, dann bietet sie einen Vorteil gegenüber dem TWT. Es gilt folgende Annahme: Allein durch die Begehung der Tat und die damit einhergehende Notwendigkeit zu täuschen lassen sich Täter von Nichttätern anhand der P300-Komponenten, die auf neutrale, tatspezifische und tatunspezifische Reize auftreten, differenzieren. Diese Annahme soll in der vorliegenden Untersuchung überprüft werden.

4. Untersuchungsparadigmen in der Psychophysiologischen Aussagebegutachtung

Verfahren der PAG wurden in Feld- oder Analogstudien erprobt. Im Folgenden werden verschiedene Untersuchungsstrategien vorgestellt, die bisher herangezogen wurden, um Methoden der PAG zu validieren. Zunächst werden die Feldstudien (4.1), anschließend die Simulationsstudien (4.2) thematisiert.

4.1 Feldstudien

Als „Feldstudien“ werden im Folgenden in Anlehnung an die amerikanische Terminologie (OTA, 1983) solche Studien bezeichnet, deren Daten anhand realer Fälle gewonnen werden. Das heißt, die psychophysiologischen Aussagebegutachtungen dienen primär der Beurteilung tatsächlich beschuldigter oder verdächtigter Personen. Die so gewonnenen Daten und resultierenden Klassifikationen der Probanden werden mit einem Kriterium verglichen, das unabhängig von den PAG-Daten eine Aussage zur Glaubhaftigkeit der Aussage des begutachteten Probanden macht. Als Kriterien wurden bisher der Prozessausgang (Barland & Raskin, 1976, zitiert nach Steller, 1987, S. 36), Geständnisse (z. B. Horvarth, 1977; Horvarth & Reid, 1971) oder Expertenurteile (z. B. Bersh, 1969) bezüglich des in Frage stehenden Falles oder eine Kombination aus diesen (Honts, 1996) als Kriterium verwendet.

Die Ermittlung eines validen Außenkriteriums stellt das Hauptproblem der Feldstudien dar. Prozessentscheidungen sind nicht immer richtig. Sie sind abhängig von der Beweislage. Außerdem kann das Prinzip „in dubio pro reo“ dieses Kriterium verzerren (Steller, 1987). Ebenso können auch Experten-Urteile falsch sein, wobei allerdings der Fehler, einen Schuldigen für unschuldig zu erklären, im Vergleich zu Gerichtsentscheidungen weniger häufig auftreten wird. Falsche Geständnisse sind nicht ungewöhnlich. Kassin und Kiechel (1996) unterscheiden drei Arten von falschen Geständnissen: Freiwillig falsche Geständnisse, die ohne äußeren Druck erfolgen; erzwungen-compliance Geständnisse, die erfolgen, um einem aversiven Verhör zu entgehen, um einen versprochenen Vorteil zu erlangen oder eine angedrohte Konsequenz zu vermeiden; erzwungen-internalisierte Geständnisse, bei denen der Proband zu der Überzeugung gelangt, dass er tatsächlich zu Recht beschuldigt wird.

Ein weiteres methodisches Problem stellt die Konfundierung der Kriterien mit den anhand der PAG gewonnenen Klassifizierungen dar. Die Wahrscheinlichkeit für ein Geständnis ist größer, wenn die PAG eine Klassifizierung als „Täter“ (positiver Befund) nahe legt. Dies besonders, da in den USA die Nachtestbesprechungen bei einem positiven Befund häufig mit der Intention geführt werden, ein Geständnis zu erwirken (Iacono, 1991). Liegt ein negativer Befund vor, wird dies nicht intendiert, vielmehr konzentrieren sich die Bemühungen nun auf einen anderen Verdächtigen. Auf diese Weise bestätigen Geständnisse meist die PAG-Ergebnisse. Der Bewertung desjenigen, der die PAG durchgeführt hat, und die der Experten oder des Gerichts beruhen teilweise auf den gleichen Informationen, z. B. aufgrund des Aktenstudiums. Dieser Umstand führt ebenfalls dazu, dass die Treffsicherheit der PAG überschätzt wird. Aus diesem Grund werden in einigen Studien die Ergebnisse der PAG von Beurteilern reanalysiert, die sich lediglich auf die PAG-Daten stützen können (z. B. Horvarth, 1977). Patrick und Iacono (1991) fanden dementsprechend auch höhere Trefferquoten für die Originalauswertungen als für die Reanalysen.

Das Problem selektiv verzerrter Stichproben entsteht dadurch, dass nur Fälle in die Untersuchungen eingehen, bei denen sich ein Außenkriterium bilden lässt. Es müssen also Geständnisse vorliegen oder die Beweislage muss einigermaßen eindeutig sein, so dass die Experten zu einer Entscheidung gelangen.

Diese methodischen Mängel haben den BGH veranlasst, in seinem Urteil vom 17.12.1998 festzustellen, dass Feldstudien im Bereich der PAG wertlos seien, weil das Außenkriterium fehlerbehaftet sei. Wie Offe und Offe (2001) bemerken, irrt der BGH mit dieser Annahme. Denn selbst, wenn das Kriterium nicht absolut valide ist, wird es nicht gleich häufig für Täter und Nicht-Täter zutreffen. Auch sind längst nicht alle Möglichkeiten zur Verbesserung solcher Studien ausgeschöpft. Es können beispielsweise auch mehrere Außenkriterien verwendet werden oder die Kriterien an sich verbessert werden, indem z. B. auf die Unabhängigkeit des Kriteriums vom Ergebnis der PAG geachtet wird oder Feldstudien prospektiv anstatt mit post hoc analysierten Daten durchgeführt werden. Feldstudien zur PAG sind sicherlich verbesserungswürdig, aber sie sind nicht grundsätzlich als Forschungsmethode im Bereich der PAG zu verwerfen.

Ähnliche Probleme, wie unzulängliche Außenkriterien oder Konfundierung durch gleiche Informationen von Gutachter und Expertengremium, gibt es auch in anderen Forschungsbereichen, so z. B. bezüglich der Forschung zum Thema der inhaltsorientierten Glaubhaftigkeitsanalyse. Hier wird beispielsweise von Steller und Dahle (1999) vorgeschlagen, beide Forschungszugänge – Feldstudien und Laborexperimente – ergänzend zu nutzen.

Solange das in Frage stehende Verfahren in Deutschland in der Praxis allerdings kaum angewendet wird, wird es schwierig sein, hier eine geeignete Feldstudie durchzuführen. Bevor eine Methode in der Praxis Verwendung findet, sollte sie sich zudem in Laborexperimenten bewährt haben.

4.2 Analogstudien

„Eine Analogstudie ist eine Untersuchung, bei der die zu untersuchende Variable und/oder die zu untersuchende Stichprobe von dem Bereich abweicht, über den etwas ausgesagt werden soll“ (Häcker & Stapf, 1994). Entsprechend werden bei Analogstudien, die zur Validierung einer Methode der PAG durchgeführt werden, Personen untersucht, die nicht deswegen einer PAG unterzogen werden, weil sie tatsächlich einer Straftat beschuldigt werden. Im Folgenden werden die verschiedenen Untersuchungsprozeduren, die bisher im Rahmen der Laborforschung angewendet wurden, um die Durchführung einer PAG im Felde zu simulieren, vorgestellt. Es handelt sich hierbei um so genannte Karten- oder Zahlentests (4.2.1), das Täuschen bezüglich autobiographischer Informationen (4.2.2), das „Mock Crime“ (4.2.3) und die so genannten „Hybridstudien“ (4.2.4). Im letzten Kapitel (4.2.5) wird eine Wertung der vorgestellten Paradigmen hinsichtlich ihrer Praktikabilität und ökologischen Validität in Bezug auf die Anwendung der PAG im Rahmen der Sexualstraftaten vorgenommen.

4.2.1 Karten- und Zahlentests

Im Karten- oder Zahlentestparadigma werden die Probanden angewiesen, eine zuvor gezogene Nummer oder Karte in einer nachfolgenden PAG-Sitzung zu verheimlichen. Hierbei werden meist alle Probanden zur Täuschung motiviert. Dabei simuliert das Ziehen der Karte die Straftat, die während der PAG verheimlicht werden soll.

Das Karten- oder Zahlentestparadigma wurde beispielsweise in Untersuchungen

von Cutrow, Parks, Lucas und Thomas (1972), Gustafson und Orne (1963, 1964) oder von Horvath (1979) verwendet. Das Procedere unterschied sich bei Gustafson & Orne (1964) dahin gehend, dass Probanden mit einbezogen wurden, die blanke Karten gezogen hatten und somit alle abgefragten Karten wahrheitsgemäß verneinen konnten.

Der Vorteil des Karten- oder Zahlentestparadigmas ist darin zu sehen, dass es schnell und leicht durchführbar ist. Allgemein wird jedoch angenommen, dass dieses Paradigma aufgrund seiner Lebensferne nicht geeignet ist, um mit seiner Hilfe den diagnostischen Wert einer Methode der PAG für die Praxis einzuschätzen (OTA, 1987). Der Wert der Studien, die dieses Paradigma verwenden, wird eher in der Möglichkeit experimenteller Variation von Faktoren gesehen, von denen ein Einfluss auf die Validität der PAG angenommen wird. Ergebnisse dieser Studien tragen wesentlich zur Theoriebildung bei (Steller, 1987).

4.2.2 Täuschen bezüglich autobiographischer Informationen

Es wird auch eine Variante der zuvor genannten Vorgehensweise verwendet, bei der nicht Karten oder Zahlen verheimlicht werden sollen, sondern autobiographische Informationen, wie z. B. der eigene Name oder das Geburtsdatum.

Dieses Paradigma wurde z. B. von Elaad (1994), Elaad und Ben-Shakhar (1990) und Lykken (1960) für Untersuchungen bezüglich der Tatwissentechnik verwendet. Eine Abwandlung gegenüber der Abfrage von biographischen Daten stellt die Untersuchung von Bradley und Cullen (1993) dar. Die Probanden wurden instruiert, ein unangenehmes, peinliches Ereignis, das sie zuvor detailliert beschrieben hatten, während eines Kontrollfragenverfahrens zu verheimlichen. Sie wurden ebenfalls einem Kontrollfragenverfahren bezüglich eines nicht erlebten peinlichen Ereignisses unterzogen. Die Probanden verheimlichten das Ereignis aber nicht aus eigenem Antrieb. Sie waren vielmehr entsprechend instruiert und sollten eine monetäre Belohnung für den Fall erhalten, dass ihre Täuschung unerkant blieb. Rosenfeld, Angell, Johnson und Quian (1991) verwendeten, wie bereits beschrieben, Vergehen wie die Verwendung eines gefälschten Ausweises oder Rauschmittelkonsum als Stimuli.

In diesem Paradigma müssen also keine nebensächlichen Details, wie z. B. Zahlen, auswendig gelernt werden. Die Informationen sind aufgrund der Biographie oder der Erlebnisse des Probanden salient. Allerdings müssen die Probanden auch in

diesem Paradigma zur Täuschung während der PAG instruiert werden. Wenn die Probanden im Rahmen des Experiments bereit sind, das peinliche oder moralisch verwerfliche Ereignis zu berichten, muss durch zusätzliche Motivation erreicht werden, dass die Probanden dieses Ereignis während der PAG wieder verheimlichen.

4.2.3 Das Mock Crime

Ein großer Teil der Analogstudien auf dem Gebiet der PAG basiert auf dem so genannten „Mock Crime-Paradigma“. Üblicherweise besteht das Mock Crime-Szenario aus einem Tat- und einem Täuschungsszenario. Die Handlungen werden vor Begehen des Mock Crimes durch die Instruktion genau vorgegeben. Probanden, die Täter simulieren sollen, führen das Scheinverbrechen aus. Probanden, die einen Nichttäter simulieren sollen, tun dies nicht. Als Grund für dieses Vorgehen wird den Probanden gleich zu Beginn der Forschungshintergrund erläutert, nämlich, dass es sich um die Erprobung einer Methode der PAG handelt. Themen dieser gespielten kriminellen Handlungen waren in bisher durchgeführten Untersuchungen „Einbruch“, „Diebstahl“, „Spionage“ oder „Mord“. In den meisten Untersuchungen müssen die Handlungen auch instruktionsgemäß durchgeführt werden. Auf die Durchführung der „Tat“ folgt die spielerische Festnahme und Beschuldigung des jeweiligen Probanden. Der Proband ist instruiert, die Tat in jedem Fall abzustreiten. Während der folgenden PAG soll er instruktionsgemäß versuchen, den Untersuchungsleiter zu überzeugen, dass er die vorgeworfenen Handlungen nicht ausgeführt hat. Für den Fall, dass er die in Frage stehende Tat begangen hat, muss der Proband also täuschen und anderenfalls die Begehung der Tat wahrheitsgemäß abstreiten. Um seine Motivation zu erhöhen wird dem Probanden in einem Großteil der Studien für einen erfolgreichen Versuch eine Belohnung versprochen.

Typische Mock Crime-Studien stellen zum Beispiel die Untersuchungen von Barland und Raskin (1975), Bradley und Black (1998), Honts, Raskin und Kircher (1987), Horowitz, Kircher, Honts und Raskin (1997) oder Raskin und Hare (1978) dar, um nur einige zu nennen. Um die Situation für den Probanden stressreicher zu gestalten, mussten in einigen Studien zur Simulation eines Mordes Schüsse aus einer echten Waffe auf eine Puppe abgefeuert werden (Timm, 1987). Bei O`Toole, Yuille, Patrick und Iacono (1994) wurde die Einbruchssituation so gestaltet, dass verschiedene

Blechdosen bei dem Versuch, den Gegenstand aus einer Manteltasche zu nehmen, scheppernd umfielen. Honts, Raskin und Kircher (1994) drohten ihren Probanden damit, dass sie durchaus festgenommen werden könnten, da die übrigen beteiligten Personen nicht informiert seien.

In anderen Studien begnügte man sich mit der schriftlichen Fassung des Tatszenarios. Die Probanden lasen über ihre Tat und stellten sie sich vor, führten die Handlungen jedoch nicht aus (z. B. Bradley & Warfield, 1984). Filmsequenzen mit strafrechtlich relevantem Inhalt stellen ebenfalls eine Variante des Tatszenarios im Mock Crime dar. Boaz, Perry, Raney, Fischler und Shuman (1991), Honts, Devitt, Winbush und Kircher (1996), sowie Iacono, Boisvenu und Fleming (1984) ließen die Probanden z. B. einen Film über einen Einbruch sehen. Um das Involvement zu erhöhen, war das Video jeweils in eine Geschichte eingebunden und wurde aus einer speziellen räumlichen Perspektive gefilmt. Die Probanden sollten sich instruktionsgemäß vorstellen, sie wären Beteiligte der gezeigten Situation und würden die Diebstähle bzw. Zerstörungen selbst vornehmen.

In einigen Studien lernten Probanden im Rahmen eines Spionagekontext Codewörter. Waid, Wilson und Orne (1981) und Waid, Orne und Wilson (1979) ließen die Probanden Wortlisten auswendig lernen, die sie später verheimlichen sollten. Bei diesen Wörtern handelte es sich um Codewörter, die in einem Spionagekontext nur den Tätern bekannt wären. Nach Aussage der Autoren bietet dieses Vorgehen eine effiziente Möglichkeit, Tatwissen und Täterschaft gleichzeitig experimentell zu erzeugen (Orne, Thackrey & Paskewitz, 1972).

Eine weitere Prozedur, die man als Mock Crime auffassen könnte, stellt eine ältere Arbeit von Berrien (1942) vor. In dieser Untersuchung wurden die Probanden instruiert, Ratten zu töten und diesen „Mord“ später zu verheimlichen.

Ein kritischer Punkt im Rahmen des Mock Crime-Paradigmas ist die Motivation der Probanden. Maßnahmen, die geeignet sind, die Täuschungsmotivation zu erhöhen, sind von solchen abzugrenzen, die die Motivation zur Teilnahme an der Untersuchung betreffen. Für die Teilnahme erhalten die Probanden in den meisten Untersuchungen eine Belohnung, die unabhängig vom Verhalten ist. Um die spezielle Motivation, d.h. die Täuschungsmotivation, zu erhöhen, werden meist zusätzlich Belohnungen (z. B. Honts, Raskin & Kircher, 1994) eingesetzt. In manchen Untersuchungen erhalten

diejenigen Probanden etwas, die erfolgreich täuschen, was die Nichttäter von den Motivationskontingenzen ausschließt. In anderen Studien werden die Nichttäter explizit in die Motivationsbedingung einbezogen, indem ihnen eine Belohnung versprochen wird, wenn sie erfolgreich als unschuldig klassifiziert werden (Waid, Wilson & Orne, 1981). Nur wenige Untersuchungen arbeiten mit angedrohten oder tatsächlich verhängten Strafen für den Fall, dass die Probanden als Täter identifiziert werden. Solche negativen Konsequenzen können im Verlust der angekündigten Belohnung (Barland & Raskin, 1975) oder Elektroschocks (Bradley & Janisse, 1981; Lykken, 1959) bestehen.

Das Mock Crime lässt sich leicht und ohne ethische Bedenken durchführen, so lange keine negativen Konsequenzen, wie Elektroschocks, angedroht werden oder Tiere getötet werden, um einen Mord zu simulieren. Indem die Probanden schauspielern, können einige Aspekte der fokussierten Straftat, wie beispielsweise die Ausführung, simuliert werden. Durch die Simulation einer Straftat können Items oder Fragen verwendet werden, die auch im Realfall zur Anwendung kämen. Gleichzeitig verhalten sich die Probanden allerdings durchgängig, als ob sie eine Straftat begingen und als ob sie aufgrund der Beschuldigung täuschen müssten. Das Ganze hat für die Probanden eher eine spielerische Qualität oder Wettkampfcharakter. Damit weicht das Mock Crime hinsichtlich motivationaler und emotionaler Aspekte vom Realfall ab. Hinzu kommt, dass die Probanden über das Untersuchungsanliegen informiert werden. Damit besteht die Gefahr, dass die Probanden Hypothesen über erwünschtes Probandenverhalten bilden und ihr Verhalten den wahrgenommenen sozialen Erwartungen anpassen werden. Das Mock Crime-Paradigma weist also einige Vorteile hinsichtlich der externen Validität gegenüber den zuvor genannten Paradigmen auf. Es ist dabei leicht praktikabel. Durch den spielerischen Charakter und die informierten Probanden ergeben sich allerdings Einschränkungen hinsichtlich der externen Validität.

4.2.4 Hybridstudien

Eine weitere Gruppe von Studien lässt sich unter dem Begriff Hybridstudie (Ekman, 1985) oder Feld-Analogstudie (Iacono & Patrick, 1987) subsumieren. Sie soll, wie aus der Bezeichnung hervorgeht, eine Mischform von Feld- und Analogstudien darstellen und die Vorteile beider in sich vereinigen. Von Probandenseite soll eine solche

Untersuchung einer Feldstudie gleichen, während sie sich von Seiten des Untersuchungsleiters wie ein Laborexperiment kontrollieren lassen sollte. Folgende Aspekte lassen sich als typische Bestandteile einer Hybridstudie herauskristallisieren:

- Die Probanden sind sich des wahren Untersuchungsanliegens nicht bewusst.
- Sie nehmen an, eine antisoziale Handlung aus eigener Entscheidung begangen zu haben.
- Sie verneinen diese Handlung während der PAG, weil sie aversive Konsequenzen vermeiden möchten.
- Die Probanden werden weder instruiert, die antisoziale Handlung durchzuführen, noch diese Handlung während der PAG abzustreiten.
- Die Situation wirkt in dem Sinne für die Probanden „echt“, dass sie sich der Bedeutung dieser Situation für die zugrundeliegende Untersuchung nicht bewusst sind.
- Die Probanden werden lediglich für die Untersuchung in diese Situation gebracht.
- Es liegt ein objektives Kriterium dafür vor, ob ein Proband die antisoziale Handlung begangen hat oder nicht.

Bei Ginton, Daie, Elaad und Ben-Shakhar (1982) handelte es sich beispielsweise um Polizeischüler, für die sich die Situation so darstellte, dass ihnen die Möglichkeit gegeben wurde, ihre Leistung bei einem Test durch kleine Betrügereien zu verbessern. Anschließend sollten sie im Rahmen einer PAG zeigen, dass sie diese Möglichkeit nicht genutzt hatten. Für den Fall, dass festgestellt wurde, dass sie tatsächlich betrogen hatten, mussten sie Konsequenzen hinsichtlich ihrer Polizeiaufbahn befürchten. Dass sie Teilnehmer einer Untersuchung zur Validierung der PAG waren, erfuhren die Probanden erst im Nachhinein. Der Untersuchungsleiter hatte diese Situation aufgrund seines Untersuchungsanliegens bewusst herbeigeführt. Die Testbögen waren so präpariert, dass die Untersuchungsleiter feststellen konnten, welche Personen betrogen hatten und welche nicht.

In einer Untersuchung von Heckel, Brokaw, Salzberg und Wiggins (1962) ließ der Untersuchungsleiter sein Portemonnaie auf einem Tisch liegen und verließ den Raum, in dem Patienten der Psychiatrie Fragebögen ausfüllten. Nach seiner Rückkehr behauptete er, dass ihm Geld aus dem Portemonnaie fehle. Das Procedere war so gestaltet, dass die Patienten tatsächlich annehmen mussten, dass sie verdächtigt wurden

und aus diesem Grund an der PAG teilnehmen sollten. Die Autoren geben keine Auskunft darüber, woran das Kriterium festgemacht worden wäre, für den Fall, dass tatsächlich einer der Patienten Geld aus dem Portemonnaie entwendet hätte. Da jedoch nichts fehlte, wussten die Untersuchungsleiter, dass es sich bei allen Probanden um Nichttäter handelte. Wiederum erfuhren die Probanden erst im Nachhinein, dass sie an einer Studie teilgenommen hatten.

In beiden Untersuchungen konnte den Probanden die anschließende Durchführung einer PAG zum Zweck der Täuschungsaufdeckung plausibel gemacht werden. Die Probanden waren also von der „Echtheit“ der Situation überzeugt und verhielten sich situationsgemäß. Das Problem dieser beiden Hybridstudien liegt darin, dass es durchweg wenige oder keine Probanden gab, die das „Vergehen“ begingen (Heckel et al.: niemand; Ginton et al.: 7 von 21) oder die die Täuschung (Ginton: 2 von 7) auch während der PAG aufrechterhielten.

In der Studie von Balloun und Holmes (1979) wurde das Tatszenario folgendermaßen verwirklicht: Zwei Konföderierte des Untersuchungsleiters, die sich als Probanden ausgaben, versuchten die eigentlichen Probanden in einem Intelligenztest zum Betrügen zu verleiten. Sie konnten 50% der Probanden zum Schummeln animieren. Anschließend wurden die Probanden beschuldigt, bei dem Test betrogen zu haben. Diesen Vorwurf konnten die Probanden entkräften, indem sie sich einem Tatwissenverfahren unterzogen. Es wurde vor der Durchführung der TWT daran erinnert, dass Betrug von Studenten an der Universität ein ernsthaftes Problem sei und im Falle von nachgewiesenem Betrug in Examenssituationen auch schwerwiegende universitäre Konsequenzen haben könne. Die Untersuchungsleiter nahmen an, dass dies die Probanden bezüglich ihres Verhaltens während des Intelligenztests ernstlich beunruhigen würde. Auf diese Weise sollten die Probanden dazu gebracht werden, den Untersuchungsleiter während der PAG aus eigener Veranlassung davon zu überzeugen, dass sie nicht betrogen hatten. Der Täterschaftsstatus des jeweiligen Probanden wurde durch die Gehilfen des Untersuchungsleiters festgelegt, da diese wussten, ob ihre Animation Erfolg gehabt hatte oder nicht. Auf diese Weise existierte ein valides Außenkriterium.

Eine Untersuchung von Patrick und Iacono (1989) wird von den Autoren ebenfalls als Hybridstudie („analogue field study“) bezeichnet. Sie machten sich die

Situation von Gefängnisinsassen zunutze, indem sie die Möglichkeit des Gewinns eines für deren Verhältnisse beträchtlichen Geldbetrags an die Bedingung knüpften, dass nicht mehr als zehn Probanden durch die PAG durchfallen durften. Die Autoren nahmen an, dass die Probanden sich dadurch einem erheblichen sozialen Druck ausgesetzt fühlten. Die Probanden waren über das Untersuchungsanliegen jedoch von Beginn an informiert. Sie wurden bezüglich der Durchführung eines Mock Crimes und ihres Verhaltens während der folgenden PAG instruiert. Die Probanden simulierten einen Diebstahl, wurden spielerisch dieses Diebstahls beschuldigt und der PAG unterzogen. Lediglich die Konsequenz des PAG-Ausgangs war besonders bedeutungsvoll für die Probanden.

Ähnlich ist eine Untersuchung von Kugelmass und Lieblich (1966) einzuordnen. Sie führten einen der üblichen Kartentests durch. Es handelte sich bei den Probanden um Polizeianwärter, die glaubten, dass der Test ein Teil einer Auswahlprozedur sei. Auf diese Weise erreichten die Untersuchungsleiter eine hohe Täuschungsmotivation bei den Probanden.

In den beiden letztgenannten Untersuchungen wird auf die Operationalisierung der Täuschungsmotivation besonderer Wert gelegt. Der Gegenstand der Täuschung ist allerdings keine aus eigenen Stücken begangene antisoziale Handlung wie in den zuvor genannten Untersuchungen. Auch besteht bei den Probanden keine intrinsische Täuschungsmotivation in dem Sinne, dass diese beiden Studien ohne Täuschungsinstruktion auskämen.

Da die Probanden in der eigentlichen Version der Hybridstudie über den experimentellen Charakter der Untersuchung im Unklaren gelassen werden, besteht die Möglichkeit, „echte“ Motivation und Emotion im Sinne von situationsangemessen zu erzeugen. Gleichzeitig existiert, im Gegensatz zur Feldstudie, bei den Hybridstudien ein valides Außenkriterium. Es fallen außerdem die restriktiven Bedingungen des Mock Crimes weg, nämlich die Begründung, es handele sich um ein Experiment zur Aufdeckung von Täuschungsverhalten, und die Instruktion zur Begehung der Tat und zum Verhalten während der PAG. Im Vergleich zum zuvor beschriebenen Mock Crime bietet die Hybridstudie die Möglichkeit hinsichtlich motivationaler und emotionaler Aspekte Analogien herzustellen. Allerdings lassen sich nur geringfügige Vergehen durchführen. Ethisch diskussionswürdig sind Aspekte der Probandentäuschung und der

Verleitung zu antisozialen Handlungsweisen.

4.2.5 Bewertung der Forschungsszenarien

Im Folgenden werden die in den vorigen Kapiteln vorgestellten Untersuchungsszenarien anhand der folgenden Aspekte bewertet: Durchführbarkeit und Untersuchungsaufwand, Forschungsethik, Anwendungsbereich, ökologische Validität bezüglich der Verwendung im Bereich der Sexualdelikte und Validität des Außenkriteriums.

Karten- oder Zahlentests lassen sich mit minimalem Aufwand durchführen und wären das Untersuchungsverfahren der Wahl, wenn „Durchführbarkeit und Untersuchungsaufwand“ das einzige Kriterium darstellen würden. Das Gleiche gilt für das Täuschen bezüglich autobiographischer Informationen. Soll eine peinliche Geschichte vom Probanden erhoben werden, ist dies natürlich mit mehr Aufwand verbunden, als wenn der Proband sein Geburtsdatum verheimlichen soll. Die Untersuchungen, die unter den Begriff des Mock Crimes zusammengefasst wurden, stellen bezüglich des Aufwands eine relativ heterogene Gruppe dar. Das Scheinverbrechen kann im Erlernen von Codewörtern bestehen, was einen geringen untersuchungstechnischen Aufwand bedeutet, es kann aber auch im Begehen eines simulierten Mordes bestehen, bei dem auf Schaufensterpuppen geschossen und kompliziertere Tatabläufe auswendig gelernt werden müssen. Die Gestaltung eines Hybridszenarios stellt besonders dann ein aufwendiges Paradigma dar, wenn genügend Probanden erhalten werden sollen, die eine Tat zunächst begehen und sie anschließend auch während der PAG abstreiten. Der Probanden-Dropout wird hoch sein.

Eng mit den Aspekten der Durchführbarkeit und des Forschungsaufwands ist die Forschungsethik verbunden. Hinsichtlich dieses Bereichs muss abgewogen werden, was den Probanden zugemutet werden kann und was nicht. In dieser Beziehung sind Karten- und Zahlentests unproblematisch. Das Täuschen bezüglich autobiographischer Informationen ist auch eher unproblematisch, da nur das Gegenstand der Untersuchung sein kann, was die Probanden freiwillig zugeben. Das Mock Crime ist vom ethischen Standpunkt meist unbedenklich, da die Probanden über den Ablauf und den Zweck der Untersuchung informiert sind und sich des spielerischen Charakters bewusst sind. Bei Androhung von Bestrafungen oder Anleitung zum Töten von Tieren erhält das Vorgehen eine andere Qualität. Hybridstudien werden bezüglich ethischer

Gesichtspunkte, wie in Kapitel 4.2.4 dargestellt, kritisiert. Als diskussionswürdig werden Aspekte der Probandentäuschung und, sofern dies geschieht, der Anleitung zu Vergehen erachtet (z. B. Steller, 1987).

Der Anwendungsbereich des Kartenparadigmas oder des Täuschens bezüglich einfacher autobiographischer Informationen ist besonders in der Theoriebildung und ersten Bewertung eines Verfahrens zu sehen (siehe Kapitel 4.2.1). Das Verfahren, das mit diesen Paradigmen in erster Linie untersucht wird, ist die Tatwissenteknik. Das Täuschen bezüglich peinlicher oder antisozialer Handlungen, das Mock Crime und die Hybridstudien wurden bereits herangezogen, um Aussagen über die Verwendung der Tatwissen- und der Kontrollfragentechnik in der Praxis zu machen (z. B. OTA, 1987).

Was die ökologische Validität angeht, ist es gemeinhin Konsens, dass es zur Anwendung der Ergebnisse experimenteller Untersuchungen nicht erforderlich ist, dass die Untersuchungssituation mit der Anwendungssituation identisch ist. Es kommt vielmehr darauf an, dass psychologische Variablen, die in der Anwendungssituation eine Rolle spielen, im Experiment nachgebildet werden, so dass ihre Auswirkungen dort untersucht werden können (Köhnken & Wegener, 1985).

Ein entscheidender Faktor für die Übertragbarkeit der Ergebnisse von Analogstudien auf die forensische Praxis stellt die Motivation dar, während der PAG glaubhaft zu erscheinen. Für den Bereich der Täuschungsaufdeckung mit Hilfe der P300 belegt die Untersuchung von Allen und Iacono (1997, siehe Kapitel 3.2.3.2) den Einfluss unterschiedlicher Motivation auf die Ausprägung dieser Komponente. Die Bedeutung der Motivation für die PAG lässt sich vergegenwärtigen, wenn man bedenkt, dass motivationale Aspekte Einfluss auf Aufmerksamkeitsprozesse haben. Damit wirken sie sich auch auf die Reizwahrnehmung während der PAG aus (Kleiner, 2002).

Die Täuschungsmotivation, die in den meisten Laborexperimenten erzeugt wird, unterscheidet sich quantitativ und qualitativ von der Motivation des real Beschuldigten. Die Situation in den meisten Laborexperimenten ist primär spielerisch und wettkampforientiert. Im Vordergrund steht die attraktive Motivation im Sinne von Dickinson und Dearing (1979). Die besteht darin, glaubhaft zu erscheinen und damit einen Gewinn zu erzielen. Dies steht im Gegensatz zu der motivationalen Situation des real Beschuldigten, die primär aversiven Charakter hat, da die negativen Konsequenzen fokussiert werden. Bei allen dargestellten Untersuchungsszenarien, mit Ausnahme der

Hybridstudien, werden die Probanden zur Täuschung instruiert und zusätzlich ein motivationaler Anreiz, meist in Form von Belohnungen gegeben. Negative Konsequenzen stehen in Hybridstudien im Vordergrund, aber auch bei manchen Mock Crimes, bei denen z. B. Elektroschocks angedroht wurden (siehe Kapitel 4.2.3).

Im Hinblick auf die ökologische Validität ist auch der Aspekt der Probandenaufklärung vor Durchführung der Untersuchung als kritisch zu betrachten. Wenn die Probanden zuvor über das Anliegen der Untersuchung informiert werden, besteht, wie in Kapitel 4.2.3 erläutert, die Gefahr, dass soziale Erwünschtheits-Effekte sich auswirken. Im Mock Crime stellen solche Erläuterungen den Regelfall dar, um den Probanden die Notwendigkeit des Vorgehens plausibel zu machen. Das gilt ebenso für die meisten Untersuchungen, die Kartentests und das Täuschen bezüglich autobiographischer Informationen verwenden. Bei den Hybridstudien entfällt diese Maßnahme. Auch die Untersuchung von Kugelmass und Lieblich (1966) stellt diesbezüglich eine Ausnahme dar (siehe Kapitel 4.2.4).

Im Hinblick auf die Übertragbarkeit von Ergebnissen auf die Verwendung der PAG im Sexualstraftatsbereich ist noch ein weiterer Aspekt zu berücksichtigen. Die Untersuchung sollte sich thematisch auf solche Delikte beziehen, bei denen die PAG zur Zeit in Deutschland nutzbringend zum Einsatz gebracht werden könnte. Ein Übertrag von Untersuchungen zu anderen Themen auf den Sexualstraftatsbereich ist nur bedingt möglich, worauf Steller und Dahle (1999) bereits hinwiesen. Aspekte wie die Auswirkung sexualbezogener Stimuli oder von Peinlichkeitsgefühlen auf die PAG konnten bisher also noch gar nicht thematisiert werden. Da der Sexualstraftatsbereich als Anwendungsfeld für die PAG diskutiert wird, werden auch experimentelle Untersuchungen in Analogie zu dieser Straftat benötigt. Das Thema „Sexualstraftat“ wurde bisher von keiner Analogstudie aufgegriffen.

Als Vorteil der Analogstudien ist die sichere Zuordnung der Probanden zu den Bedingungen anzusehen. Im Zahlen- oder Kartenparadigma kann man die zu merkende Zahl niederschreiben lassen. Man kann allerdings nicht verhindern, dass die Probanden versuchen, an eine andere Zahl oder Karte zu denken und muss sich diesbezüglich auf die Kooperativität der Probanden verlassen. Bei den zu verheimlichenden autobiographischen Informationen kann es sich um nachprüfbar Daten handeln, aber auch um Ereignisse, bei denen der Proband die einzige Informationsquelle darstellt. In

letzterem Fall ist das Kriterium nicht objektiv. Handelt es sich um begangene Delikte oder peinliche Ereignisse, ist es auch denkbar, dass der Proband motiviert ist, bei seinen Berichten nicht ganz ehrlich zu sein. Beim Mock Crime ist es prinzipiell leicht zu kontrollieren, wer die in Frage stehende Handlung begangen hat und wer nicht. Das Auswendiglernen der Instruktionen kann beispielsweise kontrolliert, die Handlungen beobachtet oder das Vorhandensein von Tatwissen durch entsprechende Befragung gesichert werden. Auch bei den vorgestellten Hybridstudien war jeweils klar, wer die Handlung begangen hatte und wer nicht.

Fasst man die Bewertung zusammen, wird man zu dem Schluss kommen, dass hinsichtlich der externen Validität den Hybridstudien der Vorzug zu geben ist. Gleichzeitig bleibt die Kontrolle des externen Glaubhaftigkeitskriteriums gegeben, was ein Vorteil gegenüber den Feldstudien darstellt. Nachteil der Hybridstudien ist ein größerer Untersuchungsaufwand und der fragliche Status hinsichtlich ethischer Aspekte. Bisher ungelöst ist die Verwirklichung eines thematischen Bezugs zu den Sexualdelikten.

5. Ableitung der Fragestellungen

Die vorgestellten Befunde belegen die potentielle Leistungsfähigkeit der Täuschungsaufdeckung mit Hilfe der P300-Komponente. In zahlreichen Untersuchungen zeigt sich, dass bei Anwendung eines sogenannten „Drei-Kategorien-Oddball-Paradigmas“ Probanden mit Tatwissen von solchen ohne Tatwissen anhand der P300 unterschieden werden können (Kapitel 3.2.3.2). Dabei wird unter der „Tat“ die Handlung verstanden, die in der jeweiligen Untersuchung verheimlicht werden soll. In den dargestellten Untersuchungen handelt es sich bei der Tat um antisoziale Handlungen (z.B. Rosenfeld, Angell, Johnson & Quian, 1991), das Erlernen einer Itemliste (z. B. Allen & Iacono, 1997), eine inszenierte Spionagehandlung (Farwell & Donchin, 1991) oder einen inszenierten Einbruch (z. B. Zhou, Yang, Liao & Zou, 1999).

Folgende Annahmen werden durch die vorliegenden Befunde für eine EEG-Ableitung bei Fz, Cz und Pz gestützt:

1) Wenn tatunspezifische Reize (U) mit geringerer Frequenz als neutrale Reize (N) dargeboten werden, dann tritt – unabhängig von der Täuschungsintention und dem Tatwissen der Probanden – im Vergleich zu den neutralen Reizen auf tatunspezifische Reize hin eine ausgeprägtere P300 auf. Also: $U > N$.

2) Für Probanden ohne Tatwissen gilt: Wenn in einem Drei-Kategorien-Oddball tatspezifische Items (S) mit der gleichen Frequenz wie tatunspezifische Items dargeboten werden, dann

a) gleicht die durch tatspezifische Reize hervorgerufene P300-Komponente eher der, die durch neutrale Items hervorgerufen wird als der, die durch tatunspezifische Items hervorgerufen wird ($S \sim N$) \neq U.

b) ist die Amplitude der P300 hinsichtlich tatspezifischer Reize nicht größer als die bezüglich neutraler Reize: $S \leq N$.

3) Unter der gleichen Voraussetzung weisen Probanden mit Tatwissen hinsichtlich der P300 die folgenden zwei Charakteristika auf:

a) Die durch tatspezifische Reize hervorgerufene P300 ist eher vergleichbar mit derjenigen, die durch tatunspezifische Reize hervorgerufen wird, als mit derjenigen, die durch neutrale Reize hervorgerufen wird: ($S \sim U$) \neq N.

b) Bei ihnen rufen tatspezifische Reize eine größere P300-Amplitude hervor als

neutrale Reize: $S > N$.

Die Untersuchung von Zhou, Yang, Liao und Zou (2000) stellt einen – wenn auch schwachen – Beleg für die Robustheit der Täuschungsaufdeckung mit Hilfe der P300 gegenüber „unschuldig erworbenem Tatwissen“ dar. Daraus lässt sich die folgende Annahme ableiten:

4) Wenn in einem Drei-Kategorien-Oddball tatspezifische Items mit der gleichen Frequenz wie tatunspezifische Items dargeboten werden, dann gleicht bei Probanden, die die in Frage stehende Tat nicht begangen haben, die P300 hinsichtlich tatspezifischer Reize der bezüglich neutraler Reize aufgetretenen P300 auch dann, wenn diese Probanden über tatspezifisches Wissen verfügen. Also soll gelten:

a) $(S \sim N) \neq U$ und

b) $S \leq N$.

Die Leistungsfähigkeit der Täuschungsaufdeckung mit Hilfe der P300 wurde anhand von Analogstudien überprüft. Die ökologische Validität der vorliegenden Untersuchungen wurde hinsichtlich der praktischen Anwendung im Bereich der Sexualstraftaten kritisiert (Kapitel 3.2.3.2 und 3.2.3.3). Es würde sich anbieten, ein geeignetes Untersuchungsszenario aus dem Bereich der Forschung zu anderen Verfahren der PAG (KFT, TWT) zur Überprüfung der Leistungsfähigkeit der Täuschungsaufdeckung mit Hilfe der P300 heranzuziehen. Aus der Betrachtung der vorliegenden Szenarien hinsichtlich verschiedener Bewertungsaspekte ergab sich, dass den Hybridstudien der Vorzug zu geben sei (Kapitel 4.2.5). Eine Hybridstudie wurde für die vorliegende Arbeit generiert. Das Vorgehen wird „Bonner Szenario“ genannt und wurde nach einer Idee von Prof. O. Berndt Scholz im Rahmen dieser Arbeit unter seiner Leitung entwickelt.

Folgende Fragestellung ergibt sich: Lässt sich im Rahmen einer Hybridstudie die Leistungsfähigkeit der Täuschungsaufdeckung mit Hilfe der P300 bestätigen? Lassen sich die Annahmen 1, 3 und 4 in diesem Kontext bestätigen? Annahme 2 wird in dieser Untersuchung nicht explizit überprüft.

Begründet werden die Annahmen damit, dass a) Täuschung eine kognitive Leistung ist, die sich mittels des Drei-Kategorien-Oddball-Paradigmas zuverlässig aufdecken lässt und b) das Oddball-Paradigma den Bedeutungsgehalt der tatspezifischen Reize in Abhängigkeit von der Täterschaft eindeutig festlegt. Annahme 1)

ergibt sich zwangsläufig aus dem Umstand, dass der P300-Komplex mit einem parieto-centralen Amplitudenmaximum dann auftritt, wenn die von einem sensorischen Ereignis vermittelte Information gezielt verarbeitet werden muss (Rösler, 1982). Die gezielte Verarbeitung der tatunspezifischen Reize resultiert aus der Aufgabenstellung, die darin besteht auf diese Reize in besonderer Weise zu reagieren. Annahme 2) wird in der vorliegenden Untersuchung durch Annahme 4) ersetzt, da sich das Verfahren für die Praxis besser etablieren ließe, wenn es auch dann zu diskriminieren in der Lage wäre, wenn eine Person, zwar kein Täter ist, jedoch über tatspezifische Informationen verfügt. Begründet werden kann diese Annahme damit, dass die P300 den Aspekt der Täuschung aufdeckt und nicht den Umstand, dass eine Person über Tatwissen verfügt. In diesem Sinne wäre die Rekognition eines tatspezifischen Reizes eine notwendige aber keine hinreichende Bedingung zur Ausbildung einer verstärkten P300-Amplitude. Dies lässt sich aufgrund der vorliegenden Befunde bisher nicht sicher bestätigen. Annahme 3) a) lässt sich damit begründen, dass tatspezifische Reize für einen Täter bedeutsam und damit mit tatunspezifischen Reizen vergleichbar sind. Annahme 3) b) lässt sich damit begründen, dass tatspezifische Reize für den Täter bedeutsamer als neutrale Reize sind. Tatunspezifische Reize werden aufgrund ihrer Aufgabenrelevanz selektiv verarbeitet. Tatspezifische Reize werden dadurch selektiv verarbeitet, dass der Täter sie erstens durch sein Wissen um ihren diagnostischen Charakter als bedrohlich wahrnimmt und zweitens weil er tatspezifische Reize per Knopfdruck wahrheitswidrig als für ihn unbedeutend beziehungsweise neutral kennzeichnet. Der Nichttäter sollte tatspezifische Reize eher nicht selektiv verarbeiten, da für ihn die Mitteilung relevant ist, dass er sie als neutral wahrnehmen wird und sie außerdem wahrheitsgemäß per Knopfdruck als unbedeutend und damit neutral kennzeichnen kann.

Eine zweite Fragestellung betrifft die Deliktsspezifität der Verwendung des Oddball-Paradigmas zur Täuschungsaufdeckung. Wie bereits gezeigt wurde, kam es in den verschiedensten Kontexten bereits erfolgreich zur Anwendung. Es sollte also gelten: Wenn Probanden dazu gebracht werden, über den Bedeutungsgehalt tatspezifischer Items zu täuschen, dann sollte dies unabhängig davon sein, wessen sie zuvor beschuldigt wurden.

Für die vorliegende Untersuchung wurde angenommen, dass sich die Deliktthematik nicht auf die P300-Amplitude auswirken würde. Es sollte also gelten:

- 5) $U(A) \sim U(B)$ sowie
 $S(A) \sim S(B)$ sowie
 $N(A) \sim N(B)$

wobei durch A und B zwei unterschiedliche inkriminierte Handlungen bezeichnet werden sollen. Unterscheiden sich die P300-Komponenten von A nicht von den P300-Komponenten von B, kann daraus auch geschlossen werden, dass für Täter gilt:

- a) $A(S > N) \sim B(S > N)$ und
b) $A[(S \sim U) \neq N] \sim B[(S \sim U) \neq N]$.

Für Nicht-Täter sollte gelten:

- a) $A(S \leq N) \sim B(S \leq N)$ und
b) $A[(S \sim N) \neq U] \sim B[(S \sim N) \neq U]$.

Die intendierte Anwendung des Oddball-Verfahrens bezieht sich besonders auf den Bereich der Sexualstraftaten (siehe Kapitel 1). Hybridstudien wurden bisher jedoch nicht mit der Intention generiert einen Bezug zum Bereich der Sexualstraftaten herzustellen. Das Bonner Szenario ermöglicht es, Probanden eine inkriminierte Handlung in Analogie zu einer Sexualstraftat vorzuwerfen. Die sexualstraftatsbezogene Handlung wird einer gewaltbezogenen Handlung gegenübergestellt.

Empirisch lässt sich diese Annahme damit begründen, dass deliktthemenübergreifend sowohl bei einer Einbruchsthematik (Zhou, Yang, Liao, & Zou, 1999) als auch bei einer Spionagethematik (Donchin, 1991) gute Trefferquoten erzielt wurden. Allerdings liegen bisher keine direkten Vergleiche von P300-Amplituden eines Probanden bezüglich verschiedener Deliktthemen vor. Aus der theoretischen Begründung der Annahmen 3) und 4) ergibt sich, dass die selektive Verarbeitung der tatspezifischen Reize aus ihrer diagnostischen Funktion und der Täuschungsabsicht der Täter resultieren soll. Die selektive Verarbeitung der tatunspezifischen Reize soll aus ihrer Aufgabenrelevanz resultieren. Die neutralen Reize sollen dagegen nicht bedeutsam sein. Deliktspezifische Faktoren sollten sich dagegen nicht auf die P300-Komponente auswirken.

6. Methodik

6.1 Untersuchungsvariablen und Versuchsplan

In diesem Kapitel werden zunächst die für die Beantwortung der Fragestellung relevanten unabhängigen und abhängigen Variablen erläutert (6.1.1) und anschließend das Design vorgestellt (6.1.2)

6.1.1 Die Untersuchungsvariablen

In diesem Abschnitt werden nacheinander die unabhängigen Variablen (UV), die Gegenstand der zugrundeliegenden Fragestellung sind (6.1.1.1), die abhängigen Variablen (AV) (6.1.1.2) und die Kontrollvariablen (6.1.1.3) eingeführt.

6.1.1.1 Die unabhängigen Variablen.

Eine unabhängige Variable ist zweifach gestuft und kennzeichnet zwei *thematisch unterschiedliche deliktbezogene Regelverletzungen*. Bei der ersten handelt es sich um eine eher auf eine Sexualstraftat bezogene Thematik, bei der zweiten um eine eher gewaltbezogene Thematik. Dies wird dadurch operationalisiert, dass einer der Filme, bezüglich dessen die Regelverletzung auftritt, eine Vergewaltigungsszene, der andere Film eine Mordszene darstellt. Der Vorwurf bezüglich der Vergewaltigungsszene umfasst die Beschäftigung mit sexuellen Aspekten, die in dem vorgegebenen Kontext unangebracht waren. Der Vorwurf bezüglich der Mordszene thematisiert die lustvolle Beschäftigung mit gewaltbezogenen Aspekten, die in der vorgegebenen Situation ebenfalls regelverletzend waren.

Als unabhängige Variable wird außerdem der *Täterschaftsfaktor* variiert. Den Probanden, die als *Täter* gelten, wird in der vorliegenden Untersuchung suggeriert, sie hätten sich beim Betrachten eines Films regelwidrig und unmoralisch verhalten. Den Probanden, die als *Nichttäter* gelten, wird suggeriert, sie hätten sich an die Regeln gehalten und erfolgreich beobachtet. Diese Suggestion erfolgte anhand eines zu diesem Zweck generierten psychometrischen Messinstruments, das die Beobachtungsgenauigkeit zurückmelden soll. Korrekt erinnerte Informationen werden gewichtet und zu einem Gesamtscore aufaddiert. Gewichtet wurde nach der geschätzten Relevanz, die die erinnerten Informationen für die Aufklärung der in den Filmen dargestellten Straftaten im Realfall hätten. Den Probanden der Tätergruppe wurde zurückgemeldet,

dass sie ein großes Detailwissen bezüglich gewalt- oder sexbezogener Aspekte des jeweiligen Films hätten, aber nur wenig ermittlungsrelevantes Wissen. Den Probanden der Nichttätergruppe wurde zurückgemeldet, dass sie eher ein großes Detailwissen bezüglich ermittlungsrelevanter Aspekte des Films hätten und sich folglich nicht durch gewalt- und sexbezogene Inhalte hätten ablenken lassen.

Durch die Verwendung des Oddball-Paradigmas variierte intraindividuell als weitere unabhängige Variable der *Itemtyp*. Dieser Faktor war dreigestuft in tatspezifische, tatunspezifische und neutrale Reize. Tatspezifische Reize stellten Beschreibungen der sexuellen beziehungsweise der gewaltbezogenen Aspekte der Filme dar. Neutrale Reize stellten nicht bedeutsame Reize dar, die mit keinem der beiden Filme in Zusammenhang standen. Tatunspezifische, also aufgabenrelevante Reize, wurden vor einer EEG-Ableitung auswendig gelernt und sollten instruktionsgemäß mit einem abweichenden Tastendruck beantwortet werden. Gleichzeitig stellten tatunspezifische Reize solche Aspekte der Filme dar, die als relevant im Ermittlungskontext bezeichnet und von einem Nichttäter vorgeblich vorrangig beobachtet worden waren.

6.1.1.2 Die abhängigen Variablen.

Als abhängige Variable wurden die Amplitudenwerte der P300-Komponenten verwendet. Je nach Auswertungsmethode wurde entweder der maximale Amplitudenwert in einem Zeitfenster von 350 bis 1000 ms nach dem jeweiligen Reiz als abhängige Variable verwendet oder es wurden alle Amplitudenwerte der EKP-Kurve in einem Bereich von 0 bis 1000 ms nach dem Reiz in die Auswertung einbezogen. Wie in Kapitel 3.2.3 erläutert wurde, stellen die Amplitudenwerte der P300 ein sensitives Maß für die Registrierung von Bedeutsamkeitsunterschieden der im Oddball-Paradigma dargebotenen Reize dar. Schlussfolgerungen lassen sich zum einen aus Unterschieden bezüglich der maximalen Amplitudenwerte ziehen, zum anderen können die EKP-Verläufe insgesamt verglichen werden, um ein Maß für die Bedeutsamkeitsunterschiede der zugehörigen Reizkategorien zu erhalten.

6.1.1.3 Zusätzliche Kontrollvariablen.

Als Kontrollvariable wurde erstens die Anzahl der korrekt wiedererkannten tatspezifischen Reize erhoben (1-6 Items), um *Rekognitionsfähigkeit* als Störfaktor auszuschließen. Zweitens wurden die in den *Feedbackbögen* erzielten Ergebnisse als Kontrollvariable festgehalten. Drittens wurden die Probanden anhand ihrer jeweiligen *Berufe* in für das Untersuchungsanliegen in motivationaler Hinsicht relevante und nicht relevante Berufe unterteilt. Zu den relevanten wurden Polizisten, Bundesgrenzschutzangehörige, Juristen und Detektive gezählt, die von Berufs wegen motiviert sein sollten, bei dem durchgeführten Beobachtungstraining gute Leistungen zu erzielen.

6.1.2 Versuchsplan

Entsprechend der Kombination der jeweils zweistufigen Faktoren „Täterschaft“ und „Thematik der Regelverletzung“ ergaben sich vier experimentelle Gruppen (siehe Tabelle 4). Weil sich ereigniskorrelierte Potentiale interindividuell stark unterscheiden, müssen die in Frage stehenden unabhängigen Variablen als Within-Faktoren variiert werden. Aus diesem Grund wurde jeder Proband bezüglich des einen Films der Täter- und bezüglich des anderen Films der Nichttäterbedingung zugeordnet. Für die Frage nach der Diskriminierbarkeit von Tätern und Nichttätern wurde der Täterschaftsfaktor als Within-Faktor betrachtet. Die Variation der Deliktthematik wurde als Between-Faktor behandelt. Für die Frage nach der Deliktspezifität wurde die Deliktthematik als Within-Faktor betrachtet und der Täterschaftsfaktor als Between-Faktor.

Tabelle 4
Versuchsplan der Untersuchung.

	Thematik der Regelverletzung	
Täterschaft	Mord	Vergewaltigung
Regelwidriges Verhalten	n = 20	n = 23
Regelgerechtes Verhalten	n = 23	n = 20

Durch die Verwendung des Oddball-Paradigmas variierte als weiterer Within-Faktor der *Itemtyp*. So ergab sich ein 2 (Täterschaft) x 2 (Delikttyp) x 3 (Itemtyp/Within)-faktorieller Gruppenplan, bei dem alternativ der Täterschaftsfaktor oder der Deliktfaktor als Within-Faktor betrachtet wurde.

6.2 Die Untersuchungsstichprobe

An der Untersuchung nahmen 45 Männer im Alter zwischen 20 und 43 (Mittelwert (M) = 27,16; Standardabweichung (S) = 5,4) Jahren teil. 38 (84,8%) Probanden gaben sofort an, aus Interesse am Untersuchungsanliegen, das in der Erprobung bestimmter Beobachtungsregeln bestand, teilzunehmen. Aus Neugier nahmen 4 (8,9%) und aus Hoffnung auf finanzielle Entlohnung nahmen 3 (6,7%) Personen teil. Nachdem das vorgebliche Untersuchungsanliegen erläutert worden war, gaben alle Probanden an, an einer Verbesserung ihrer Beobachtungsleistung interessiert zu sein. Eine akademische Ausbildung der unterschiedlichsten Fachrichtungen (Jura, Chemie, Agrarwissenschaften, Physik, BWL, Informatik) angefangen oder beendet hatten 40 (89%) Probanden, 5 (11%) Probanden hatten eine Ausbildung als Dachdecker, Steuerfachgehilfe, Polizist, Grenzschutzbeamter oder Sozialarbeiter abgeschlossen. Alle Probanden hatten normale oder zu normal korrigierte Sehstärken. Sie verfügten bezüglich der deutschen Sprache über zumindest durchschnittliche sprachliche Ausdrucksfähigkeit.

Ein Proband erschien nicht zum zweiten Termin, ein weiterer Proband musste aufgrund der Ausschlusskriterien aus der Stichprobe ausgesondert werden. Es gingen damit die Daten von 43 Probanden in die Auswertung ein. Zwanzig Probanden wurden der Bedingung zugeordnet, in der sie als Nichttäter bezüglich der Vergewaltigungsthematik und als Täter bezüglich der Mordthematik galten. Dreiundzwanzig Probanden wurden der Bedingung zugeordnet, in der sie als Täter bezüglich der Vergewaltigungsthematik und als Nichttäter bezüglich der Mordthematik galten.

Die Probanden wurden durch Zeitungsartikel, Aushänge und persönliche Ansprache rekrutiert. Zur Vermeidung geschlechtsspezifischer Effekte wurden ausschließlich Männer in die Stichprobe aufgenommen. Einschlusskriterien waren außerdem „Interesse an der Untersuchung“ und „Interesse an der Verbesserung der eigenen Beobachtungsleistung“. Als Ausschlusskriterium galt eine klinisch relevante psychische Störung, sowie eine auffällige Intelligenzminderung. Darüber hinaus wurde auf die Abwesenheit allgemeinmedizinischer und neurologischer Erkrankungen, sowie akuter und chronischer Schmerzzustände geachtet. Auch die Einnahme von Medikamenten führte zum Ausschluss der Probanden.

6.3 Materialien und apparativer Aufwand

In diesem Kapitel werden die verwendeten Untersuchungsmaterialien, also der Bogen mit den Beobachtungsregeln (6.3.1), das Filmmaterial (6.3.2), die Feedbackbögen (6.3.3), die Gudjonsson Suggestibility Scale (6.3.4), die Itemlisten (6.3.5), sowie der Rekognitionstest (6.3.6) und der apparative Teil (6.3.7) beschrieben. Tabelle 5 zeigt die verschiedenen Untersuchungsschritte und die dazugehörigen Untersuchungsmaterialien.

Tabelle 5

Untersuchungsmaterialien und Aufgaben von Untersuchungsleiter (UL) und Proband (Pb) für die unterschiedlichen Untersuchungsschritte.

Untersuchungsschritt	Aufgabe des UL	Aufgabe des Pb	Untersuchungsmaterial
1. Termin			
Vorgespräch	Motivierung des Probanden	Persönliche Informationen geben	-
Erlernen der Beobachtungsregeln	Erläuterung	Erlernen	Bogen mit Beobachtungsregeln
Filmdarbietung 1	Nicht anwesend	Beobachtungsregeln anwenden	Film 1
Detailabfrage 1	Hilfestellung wenn nötig	Fragen beantworten, Bogen auswerten	Feedbackbogen 1
Filmdarbietung 2	Nicht anwesend	Beobachtungsregeln anwenden	Film 2
Detailabfrage 2	Hilfestellung wenn nötig	Fragen beantworten, Bogen auswerten	Feedbackbogen 2
Gudjonsson Suggestibility Scale	Durchführung	Beantwortung	Gudjonsson Suggestibility Scale
2. Termin			
EEG-Ableitung: Übungsdurchgang	Elektroden kleben, Instruktion	Reize zuordnen	Itemliste 1 (Übungsdurchgang)
Vorwurf	Vorwurf	Keine bestimmte Reaktion erwartet	-
Experimentaldurchgang 1	Nicht anwesend	Reize zuordnen	Itemliste 2
Experimentaldurchgang 2	Nicht anwesend	Reize zuordnen	Itemliste 3
Nachbesprechung	Instruktion	Items zuordnen	Rekognitionstest
Probandenaufklärung	Aufklärung	Information	-

6.3.1 Beobachtungsregeln

Dieser Bogen umfasste Beobachtungsregeln, die dem Probanden ermöglichen sollten, eine gute Beobachtungsleistung zu erzielen. Den Probanden wurde erläutert, dass eine

gute Beobachtungsleistung eine solche wäre, die es ermögliche, alle für die Ermittlung und Überführung eines Täters potentiell wichtigen Details erinnern zu können. Da dies bei der Beobachtung von gewalttätigen Szenen besonders schwierig sei, würden sie hilfreiche Beobachtungstipps erhalten. Sie erhielten diese Tipps in schriftlicher Form. Diese kurze Zusammenfassung bestand aus einer Beschreibung der Regeln und einer allgemeinspsychologischen Begründung dieser Regeln. Der Bogen ist in Anhang B wiedergegeben.

6.3.2 Filmmaterial

In der Untersuchung wurden zwei Filmausschnitte verwendet, anhand derer die Probanden zuvor gelernte Beobachtungsregeln anwenden sollten. Beide Filmausschnitte zeigten eine Straftat. Den Probanden wurden diese Filme als emotional involvierend angekündigt.

Film 1 stellte eine Sexualstraftat dar. Für die Darstellung der Sexualstraftat wurde ein Zusammenschnitt einer Szene aus dem Film „Todesengel“ von Andreas Berthmann ausgewählt. Die geschnittene Szene hat eine Länge von 4 Minuten und 42 Sekunden. Die Sequenz handelt von einer Frau, die eine Autopanne am Waldrand hat. Zwei Männer, die vorgeblich helfen wollen, greifen sie tätlich an. Die Frau kann sich zunächst dem Zugriff durch einen Tritt in die Genitalien eines Angreifers entziehen und flieht. Da ihr jedoch ein Zaun den Weg versperrt, wird sie eingeholt und überwältigt. Einer der Täter entkleidet sie gewaltsam und beißt sie in die Brust. Ob es sich bei der Straftat um eine versuchte oder eine vollzogene Vergewaltigung handelt, ist nicht erkennbar. In der letzten Einstellung liegt die Geschädigte regungslos auf einem Waldweg. Die Täter flüchten mit einem schwarzen Auto der Marke BMW.

Der zweite Ausschnitt (Film 2) stammt aus dem Film mit dem Titel „Ein kurzer Film über das Töten“ von Krzysztof Kieslowski. Die geschnittene Szene hat eine Länge von 5 Minuten und 33 Sekunden. Ein Fahrgast würgt einen Taxifahrer zunächst mit einem Hanfseil. Der Taxifahrer versucht durch lautes Hupen auf sich aufmerksam zu machen, doch niemand hört ihn. Durch Schläge mit einem Metallrohr bringt der Angreifer den Geschädigten dazu, von der Hupe abzulassen. Nach einem Schlag auf den Kopf fällt dem Geschädigten das Gebiss aus dem Mund. Als der Geschädigte sich nicht mehr bewegt, lässt der Täter von ihm ab. Er findet eine Wolldecke im Auto, die er dem

Geschädigten um den Kopf wickelt. An den Füßen zieht er ihn einen Abhang hinunter an ein Flussufer. Da der Geschädigte wieder zu sich kommt, zertrümmert der Täter schließlich durch Schläge mit einem Felsbrocken den Kopf des Geschädigten. Die letzte Einstellung zeigt, wie das Blut durch die Woldecke sickert, die der Geschädigte um den Kopf gewickelt hat.

6.3.3 Feedbackbögen

Die Feedbackbögen wurden konzipiert, um die Probanden den jeweiligen Täterschaftsbedingungen randomisiert zuteilen zu können. Sie stellten somit die Operationalisierung der Variation der Täterschaftsbedingung dar. Einerseits musste zu diesem Zweck das Ergebnis, das die Probanden in dem Fragebogen erzielen würden, bereits feststehen, andererseits mussten die Probanden den Eindruck gewinnen, dass sie selbst verantwortlich für dieses Ergebnis waren. Beim Probanden sollte der Eindruck entstehen, dass er anhand dieses Bogens ein zuverlässiges Feedback bezüglich seiner Beobachtungsleistung während des jeweiligen Films erhalten hatte. Ein weiterer Zweck bestand darin, dem Probanden die Informationen, die aus dem Film bekannt waren, auch in dem Wortlaut, in dem sie dann während der EEG-Ableitung auftauchen würden, vor Augen zu führen.

Jeder Bogen bestand aus 12 Fragen zu Details aus dem jeweiligen Film. Es war vorgesehen, dass die Probanden, nachdem sie alle Fragen beantwortet hatten, die Auswertung selbst vornahmen, damit sie das Ergebnis überzeugend fanden. Sie sollten dazu zunächst die mit dem Antwortschema übereinstimmenden Lösungen kennzeichnen. Anschließend wurde die Wertung vorgenommen. Dazu wurde zwischen Fragen nach relevanten Details und Fragen nach sogenannten „reißerischen Details“ unterschieden. Damit wurden solche Aspekte des Films bezeichnet, die sexuelle oder gewalttätige Aspekte des jeweiligen Films bezeichneten. Die richtige Beantwortung einer Frage nach einem solchen „reißerischen Detail“ wurde negativ gewertet. Um den Probanden diese Entscheidung plausibel zu machen, wurde darauf hingewiesen, dass solche Details in unmittelbarem Zusammenhang mit den Aspekten des Films standen, die nicht im Zentrum der Aufmerksamkeit hätten stehen sollen. Das Wissen um diese Details belege, dass sie sich durch sexuelle und gewalttätige Details hätten ablenken lassen. Eine entsprechende Frage lautete beispielsweise für die Sexualstraftatsszene:

„Wo wurde die geschädigte Person durch einen Biss verletzt?“ Eine richtige Antwort darauf lautete: „Im Brustbereich“. Eine Frage nach einem relevanten, also ermittlungstechnisch bedeutsamen, Detail lautete z. B.: „Was taten die Täter nachdem sie die geschädigte Person auf dem Weg zurück gelassen hatten?“. Die richtige Antwort lautete: „Sie fuhren in einem schwarzen BMW weg!“. Die richtige Beantwortung wurde positiv bewertet. Begründet wurde dies damit, dass diese Information für die Täterüberführung oder -ermittlung hätte hilfreich sein können. Für jede richtig beantwortete Frage wurde ein Punkt vergeben. Handelte es sich um eine richtige Antwort auf eine relevante Frage, wurde ein Punkt hinzuaddiert. Handelte es sich um eine Antwort bezüglich eines Details, das nicht beachtet werden sollte, wurde ein Punkt abgezogen. Es konnte also ein Gesamtwert zwischen -6 und +6 erreicht werden. Alle Ergebnisse kleiner 0 führten zu der Wertung „schlechte Beobachtungsleistung“, alle größer oder gleich 0 zu der Bewertung „gute Beobachtungsleistung“. Probanden, die eine schlechte Beobachtungsleistung bescheinigt bekamen, galten als Täter. Probanden mit guter Beobachtungsleistung galten als Nichttäter.

Entsprechend den Versuchsbedingungen gab es vier Versionen des Feedbackbogens: Eine, in der eine schlechte Leistung bezüglich der Beobachtung des Mordfilms zurückgemeldet wurde, und eine, die eine gute Leistung bescheinigte. Eine weitere Version informierte den Probanden über eine schlechte Leistung bei der Beobachtung der Sexualstraftatsszene und die vierte meldete diesbezüglich eine gute Leistung zurück. Fragebögen, die nach der Beantwortung und Auswertung mit hoher Wahrscheinlichkeit eine schlechte Beobachtungsleistung rückmeldeten, unterschieden sich von den beiden, die mit hoher Wahrscheinlichkeit eine gute Leistung rückmeldeten, durch die Verteilung der Schwierigkeitsgrade auf die Fragen. Die Schwierigkeit wurde variiert durch die Differenziertheit der geforderten Antworten und durch die Eindeutigkeit der Fragen. Da man um ein Ergebnis kleiner 0 zu erhalten möglichst viele Antworten bezüglich „reißerischer“ Aspekte und möglichst wenig relevante Fragen korrekt beantworten musste, waren die Fragebögen, die dies intendierten, so gestaltet, dass die Fragen bezüglich relevanter Aspekte schwer und die bezüglich der anderen Aspekte leicht zu beantworten waren. Bei Feedbackbögen, die eine gute Beobachtungsleistung zurückmelden sollten, verhielt sich die Verteilung der Schwierigkeitsgrade umgekehrt. Die Bögen sind in Anhang C abgedruckt.

6.3.4 Gudjonsson Suggestibility Scale (GSS)

Der Einsatz der GSS in der vorliegenden Arbeit war in erster Linie durch die Untersuchungslogik begründet (siehe Kapitel 6.4). Ergebnisse der GSS gehen nicht als Maße für die Suggestibilität in die Auswertung ein. Dies kann folgendermaßen begründet werden:

Auf der theoretischen Basis des Suggestibilitätsmodells von Gudjonsson und Clark (Gudjonsson, 1993) entwickelte Gudjonsson die Gudjonsson Suggestibility Scale (GSS). Die GSS1 besteht aus 20 Fragen zu einer Geschichte, die zuvor vorgelesen wird. Die Geschichte handelt von einem Handtaschenraub. Der Test wird als Gedächtnistest vorgestellt. Die Versuchsperson wird um eine freie Wiedergabe gebeten und soll anschließend 20 Fragen beantworten. Der Proband erhält ein kritisches Feedback („Sie haben eine ganze Reihe von Fehlern gemacht. Wir müssen deshalb die Fragen noch einmal durchgehen und Sie sollten sich bemühen, dieses Mal zutreffendere Antworten zu geben“). Dann werden die Fragen wiederholt.

Als Maß für die Erinnerungsleistung wird die Zahl der in der freien Wiedergabe richtig berichteten Informationen erhoben. Die Anzahl der in die suggerierte Richtung beantworteten Fragen wird als „Yieldscore“⁹ 1 bezeichnet. Dieser Score soll also ein Maß dafür darstellen, wie nachgiebig Probanden gegenüber Suggestivfragen sind. Die Anzahl der veränderten Antworten beim zweiten Durchgang stellt den Shiftscore¹⁰ dar. Dieser Score soll nach Gudjonsson (1993) ein Maß dafür darstellen, inwieweit ein Proband sozialem Druck nachgibt. Als Yieldscore 2 gilt die Anzahl der in die suggerierte Richtung beantworteten suggestiven Fragen beim zweiten Durchgang. Da die beiden Skalen als voneinander unabhängig gelten, wird aus der Summe von Yieldscore 1 und Shiftscore ein Gesamtwert berechnet.

Interne Konsistenz (Yield: $\alpha = .77$; Shift: $\alpha = .67$) und Konstruktvalidität der GSS gelten als zufriedenstellend. Die GSS-Werte korrelieren signifikant mit Gedächtnis-, Intelligenz- und Persönlichkeitsvariablen (Gudjonsson, 1983). Gudjonsson entwickelte die GSS für den Einsatz bei polizeilichen Verhören mit der Intention, die Suggestibilität der Befragten einschätzen zu können. Befragungstechniken sollten entsprechend angeglichen werden und falsche Geständnisse verhindert werden.

⁹ yield (engl.) - nachgeben

¹⁰ shift (engl.) - verändern

Validierungsversuche von Gudjonsson beziehen sich entsprechend auf den Untersuchungskontext falscher Geständnisse (Gudjonsson, 1984). Insgesamt gesehen gibt es für das Konzept der Messung der Aussagesuggestibilität mit derartigen Skalen nur sehr schwache Belege (siehe Endres, Scholz & Summa, 1997; Scullin & Ceci, 2001), beziehungsweise Hinweise darauf, dass es keinen Zusammenhang zwischen Skalenwerten und der Reaktion auf beeinflussende Faktoren in Interviews gibt (Jager, 1997).

6.3.5 Itemlisten

Für die Realisierung des Oddball-Paradigmas wurden neutrale, tatunspezifische und tatspezifische Items benötigt. Zu diesem Zweck wurde eine Itemliste bezüglich des Mordfilms und eine Itemliste bezüglich der Sexualstraftatsszene generiert. Jede der beiden Listen umfasste 6 tatspezifische, 6 tatunspezifische und 24 neutrale Items, so dass zwischen tatspezifischen, beziehungsweise tatunspezifischen Items und neutralen Items ein Häufigkeitsverhältnis von 1:4 bestand.

Tatspezifische Items stehen in direktem Tatzusammenhang. Da die Tat in der vorliegenden Untersuchung darin bestand, während des Betrachtens eines Filmausschnitts die zuvor vereinbarten Beobachtungsregeln zu vernachlässigen und stattdessen auf die gewalt- beziehungsweise sexbezogenen Aspekte der Filmausschnitte zu achten, bezogen sich diese Items entweder auf gewalt- oder sexbezogene Aspekte. Für die Vergewaltigungsszene stellt beispielsweise „Biss in Busen“ ein tatspezifisches Item dar. Für den Mordfilm stellte „Woldecke“ ein solches Item dar, weil man in einer Szene das Blut durch diese sickern sieht, während der Täter mit einem Felsbrocken auf den mit dieser Decke umwickelten Kopf des Geschädigten einschlägt.

Jedem der sechs tatspezifischen Items war ein tatunspezifisches Item zugeordnet. Dieses hatte mit dem tatspezifischen Item eine semantische Kategorie gemeinsam und wies die gleiche Anzahl an Wörtern und Silben auf. Die tatunspezifischen Items bezogen sich auf diejenigen Aspekte des jeweiligen Films, die zuvor mittels Feedbackbogen als ermittlungsrelevant klassifiziert worden waren. Dem tatspezifischen Item „über Zaun hängen“ – in dieser Position war die blanke Brust der Geschädigten zum ersten Mal in dem Film zu sehen – wurde beispielsweise das tatunspezifische Item „in's Auto steigen“ zugeordnet. Diese Information war

unmittelbar mit der Tatsache verknüpft, dass die Täter in einem Auto flüchteten, dessen Beschreibung ermittlungstechnisch interessant sein könnte. Dem tatspezifischen Reiz „Gebiss“ – dieses fiel dem Geschädigten nach einem Schlag auf den Kopf aus dem Mund – wurde das tatunspezifische Item „Warze“ – diese kennzeichnete den Täter – gegenübergestellt.

Den tatspezifischen Items waren außerdem jeweils vier neutrale Items zugeordnet. Diese Items erfüllen die oben genannten Bedingungen hinsichtlich Silben- und Wortzahl. Sie glichen außerdem inhaltlich dem korrespondierenden tatspezifischen oder tatunspezifischen Item. Neutral waren diese Items in der Hinsicht, dass sie zuvor weder durch die entsprechenden Filme noch durch die Feedbackbögen thematisiert worden waren. Beispiele für neutrale Items stellten in Anlehnung an das tatspezifische Item „Biss in Busen“ weitere körperteilspezifische Angriffe dar. Sie lauteten „Kniff in Hintern“ oder „Stich in Schenkel“. Bezüglich des Mordes stellten die Items beispielsweise in Anlehnung an das genannte tatspezifische Item „Gebiss“ weitere mögliche Auffälligkeiten im Gesichtsbereich dar. Sie lauteten z. B. „Ekzem“ oder „Brille“.

Um sicherzustellen, dass sich tatspezifische Items nicht durch andere Merkmale als ihr Vorkommen im Film von den neutralen Items unterscheiden, waren die Itemlisten gemischt und vor Beginn der eigentlichen Untersuchung 5 Studenten mit der Bitte vorgelegt worden, diejenigen Items herauszusuchen, die sie für tatspezifisch hielten. Zuvor waren den Studenten beide Filme und die zu Grunde liegenden Straftaten kurz umschrieben worden, ohne die fraglichen Items zu erwähnen. Keinem der Probanden gelang es, tatspezifische Items als solche zu identifizieren. Die beiden Itemlisten für die Experimentaldurchgänge sind in Anhang D wiedergegeben.

Für einen Übungsdurchgang (siehe Tabelle 5) wurde eine zusätzliche Itemliste erstellt. Dieser Übungsdurchgang diente lediglich dem Vertrautwerden mit dem Aufgabentyp, so dass keine besonderen Anforderungen an die Items gestellt wurden. Weder tatspezifische noch tatunspezifische oder neutrale Items waren den Probanden durch die gezeigten Filmausschnitte oder die Feedbackbögen bekannt. Tatspezifische und tatunspezifische Items waren willkürlich durch den Untersuchungsleiter als solche klassifiziert worden.

6.3.6 Rekognitionstest

Das Anliegen dieses Tests bestand darin zu überprüfen, ob die Probanden tatspezifische Items als solche identifizieren konnten. Die Rekognition der tatspezifischen Items durch die Probanden stellt eine notwendige Bedingung für die Registrierung einer ausgeprägten P300-Komponente dar. Der Rekognitionstest stellte eine alphabetische Auflistung der während der Experimentaldurchgänge dargebotenen Items dar, die in Anhang D wiedergegeben werden. Die Aufgabe der Probanden bestand darin, sowohl diejenigen Items zu kennzeichnen, die sie für tatspezifisch hielten, als auch diejenigen, die sie für tatspezifisch hielten.

6.3.7 Apparatur zur Stimulusdarbietung und Ableitung des EEGs

Die zufällige Auswahl der Items wurde mit Hilfe von Experimental Run Time System (ERTS) generiert. Die Items wurden auf einem 15-Zoll-Monitor präsentiert. Der Bildschirm war ca. 1,20 m entfernt von den Probanden platziert. Zur Registrierung der Antwortreaktionen wurde ein Brett mit zwei adjustierten leichtgängigen Antworttasten jeweils so positioniert, dass der Proband die Antworttasten bequem mit den Zeigefingern seiner beiden Hände bedienen konnte. Die beiden Zeigefinger wurden in Ruhe über den Tasten gehalten, so dass zur Bedienung der Tasten ein Wegsehen vom Bildschirm nicht erforderlich war.

Das kontinuierliche EEG wurde unipolar abgeleitet. Nach dem internationalen 10-20-System für die Applikation von Elektroden (Jasper, 1958) wurde von folgenden Positionen abgeleitet: Fz, Cz und Pz. Als Referenz dienten die verbundenen Mastoide. Das Elektroofokulogramm (EOG) zur Kontrolle von Blink- und Augenbewegungen wurde von zwei Elektrodenpositionen abgeleitet. Die eine befand sich ca. 1 cm über dem rechten Auge und die andere ca. 1 cm neben dem rechten Auge. Die Erdung erfolgte über eine Elektrode auf der Stirn. Verwendet wurden nichtpolarisierbare Silber/Silberchlorid-Napfelektroden (MES). Als Elektrolyt und Haftpaste wurde Elefix (Nihon-Kohden) verwendet. Die Kopfhaut wurde vor Applikation der Elektroden mit der Peelingpaste SkinPure (Nihon-Kohden) vorbehandelt.

Die Signale wurden mit Hilfe eines DC-Verstärkers (Synamps) 1000fach verstärkt. Während der Ableitung kamen ein Tiefpassfilter von 50 Hz und ein Hochpassfilter von 0,1 Hz zum Einsatz. Alle Daten wurden mit einer

Digitalisierungsrate von 1000 Hz aufgenommen und zur weiteren Analyse samt Triggerinformation gespeichert. Der A/D-Wandler verwendete 16 bits. Die Reichweite betrug 5,5mV. Das heißt, die Auflösung der Signale lag bei 0,084 $\mu\text{V}/\text{bit}$.

Vor Beginn der jeweiligen EEG-Aufzeichnung wurde eine Kalibrierung durchgeführt. Die Widerstände betragen jeweils weniger als 10k Ω , in vielen Fällen weniger als 5k Ω .

6.4 Die Untersuchungsdurchführung

Im Folgenden wird die Durchführung des sogenannten „Bonner Szenarios“ beschrieben. Die Untersuchung wurde auf zwei Termine verteilt, die ungefähr eine Woche auseinander lagen. Nach Ankunft der Probanden im Labor unterzeichneten diese eine Einverständniserklärung (siehe Anhang A). Ein männlicher Versuchsleiter führte mit den Teilnehmern ein einführendes Gespräch in Form eines halbstandardisierten Interviews. Dieses Gespräch diente erstens dazu, soziodemographische Informationen abzufragen, zweitens dazu, dem Probanden das vorgebliche Untersuchungsanliegen zu erläutern und drittens dazu, seine Teilnahmemotivation zu erfragen. Durch dieses Gespräch sollte darüber hinaus zusätzliche Leistungsmotivation geschaffen werden. Das vorgebliche Untersuchungsanliegen bestand in der Überprüfung eines Trainings, das die Beobachtungsleistung von Zeugen einer Straftat verbessern soll.

Anschließend erhielt jeder Proband die Zusammenfassung der sogenannten Beobachtungsregeln. Nachdem der Proband sich die Informationen durchgelesen hatte, stellte der Untersuchungsleiter drei Regeln ausführlicher vor, auf die sich der Proband bei den folgenden Filmsequenzen konzentrieren sollte. Die relevanten Regeln wurden sinngemäß wie folgt vom Untersuchungsleiter zusammengefasst:

- 1) „Die emotionale Verfassung während der Enkodierung sollte der Verfassung während des Abrufs der Informationen gleichen. Versuchen Sie, sich in emotional erregenden Situationen zu entspannen. Eine der Tatsituation ähnliche Verfassung lässt sich dann später leichter herstellen. Atmen Sie z.B. ein und lang und tief aus. Also, versuchen Sie, sich zu entspannen!“
- 2) „Versuchen Sie, Gefühle so weit wie möglich auszuschalten, konzentrieren Sie sich auf das Beobachten!“
- 3) „In emotional erregenden Situationen neigen Menschen dazu, auf zu wenige

Details zu achten, weil sie sich von der Handlung mitreißen lassen. Dadurch werden periphere Details schlechter erinnert als in neutralen Situationen. Sie sollten also periphere Merkmale ganz bewusst beachten und sich nicht von gewalt- oder sexbezogenen Aspekten der Situation ablenken lassen.“

Diese Regeln wurden mit den Probanden erörtert und dabei besonders die dritte Regel fokussiert. Mit der Einigung auf die Regeln für die Beobachtungsweise während der Filme wurde die Grundlage gelegt, auf der später der Vorwurf den Probanden gegenüber aufgebaut werden konnte.

Bevor die Probanden die Videosequenzen ansahen, wurden vier Elektroden zur Ableitung eines Elektrokulogramms (EOGs) appliziert. Die Elektrodenpositionen befanden sich super- und suborbital am rechten Auge, sowie auf dem linken Außenrand des linken Auges und auf dem rechten Außenrand des rechten Auges. Die registrierten Augenbewegungen gingen nicht in die Ergebnisse der vorliegenden Untersuchung ein. Vielmehr wurde den Probanden verdeutlicht, dass den Untersuchungsleitern mit der Registrierung der Augenbewegungen neben den Ergebnissen der Feedbackbögen ein weiteres Kriterium für das gezeigte Beobachtungsverhalten zur Verfügung stand.

Die Instruktion zum Ansehen der Filme lautete: „Stellen Sie sich vor, sie haben diese Straftat beobachtet und sollen anschließend als Zeuge aussagen. Verwenden Sie die eben erhaltenen Beobachtungstipps“. Es folgte die Darbietung des ersten Films. Die Darbietungsreihenfolge der Filme wurde permutiert. Während der Darbietung waren die Probanden allein im Laborraum. An die Videopräsentation schloss sich das Bearbeiten des ersten Feedbackbogens, sowie die selbständige Korrektur und Auswertung desselben an. Den Probanden wurde erklärt, dass diese Bögen lediglich dazu dienen, dem Probanden eine Rückmeldung zu liefern. Es folgte die Darbietung des zweiten Films und im Anschluss daran die Bearbeitung des zweiten Feedbackbogens. Damit war jedem Probanden klar, dass er in einem Film eine gute und in dem anderen Film eine schlechte Beobachtungsleistung gezeigt hatte (siehe Kapitel 6.3.3).

Die Reihenfolge der Bögen war thematisch abhängig von der Reihenfolge der dargebotenen Filme. Die Operationalisierung der Täterschaftsbedingung durch die Feedbackbögen wurde jedoch ebenfalls permutiert, so dass jeder Proband einer von vier Bedingungen zugeordnet war (siehe Tabelle 6).

Tabelle 6

Für die Kombination der Darbietungsreihenfolge ergaben sich vier Bedingungen, die hier wiedergegeben sind. „Mord“ und „Vergewaltigung“ kennzeichnen die Themen der Filmausschnitte. „<0“ steht für ein schlechtes Ergebnis im Feedbackbogen und suggeriert damit dem Probanden Täterschaft. „> oder = 0“ steht für ein gutes Ergebnis im Feedbackbogen und meldet dem Probanden eine gute Beobachtungsleistung und damit keine Täterschaft hinsichtlich des späteren Vorwurfs zurück.

1. Film	Mord	Mord	Vergewaltigung	Vergewaltigung
1.Feedbackbogen	< 0	> oder = 0	< 0	> oder = 0
2. Film	Vergewaltigung	Vergewaltigung	Mord	Mord
2.Feedbackbogen	> oder = 0	< 0	> oder = 0	< 0

Nachdem beide Filme und Bögen bearbeitet worden waren, beantworteten alle Probanden die Fragen der Gudjonsson Suggestibility Scale (GSS) (Gudjonsson, 1993). Anschließend wurden die Probanden mit der Bitte entlassen, vorläufig mit niemandem über die Untersuchung zu sprechen.

Eine Untersuchungsleiterin führte zu Beginn des zweiten Termins eine Befragung bezüglich des Befindlichkeitsstatus, eingenommener Medikamente und etwaiger Krankheiten durch. Die EEG- und EOG-Elektroden wurden wie in Kapitel 6.3.7 dargestellt appliziert. Die Probanden wurden instruiert, sich zu entspannen und ruhig zu sitzen.

Es folgte ein Übungsdurchlauf. Die Aufgabe bestand für jeden Probanden zunächst darin, die tatunspezifischen Items der Itemliste für den Übungsdurchgang (siehe Kapitel 6.3.5) auswendig zu lernen. Diese wurden dem Probanden auf dem Bildschirm (siehe Kapitel 6.3.7) präsentiert. Als Kriterium für das Beherrschen der Liste galt die freie und korrekte Wiedergabe. Die Versuchspersonen sollten nun, während ein EEG abgeleitet wurde, die Items so schnell wie möglich den Kategorien „gelerntes Wort“ (rechte Taste) oder „nicht-gelerntes Wort“ (linke Taste) durch entsprechenden Tastendruck zuordnen. Während des Übungsdurchgangs wurde den Probanden in zufälliger Reihenfolge jedes Item der Übungsliste viermal dargeboten. Das heißt, es wurden insgesamt 144 Items präsentiert. Die Darbietungszeit pro Item betrug 300 ms, die Intervalle zwischen den Items betragen 3000 ms. Die Darbietungszeit betrug also insgesamt ca. 8 Minuten. Die Funktion des Übungsdurchgangs war es, die Probanden vor Durchführung der Experimentaldurchgänge mit der Aufgabenstellung des Oddball-

Paradigmas vertraut zu machen. Außerdem wurden während dieses Durchgangs die EEG-Kurven kontrolliert und, falls erforderlich, Aspekte (schlecht geklebte Elektroden, eingeschaltete Handies, Bewegungen auf Grund unbequemer Sitzhaltung etc.), die offensichtlich Artefakte verursachten, beseitigt.

Nach dem Übungsdurchgang wurde den Probanden das vorgebliche Anliegen der EEG-Ableitung erläutert. Es erfolgte der implizite Vorwurf an die Teilnehmer sich während des Betrachtens der Videos nicht an die Beobachtungsregeln gehalten und sich zusätzlich moralisch bedenklich verhalten zu haben, da sie an sexuellen Reizen oder Gewaltdarstellungen trotz des Kontextes Freude gehabt hätten. Der Wortlaut wird im Folgenden sinngemäß wiedergegeben:

„Wie ich gesehen habe, waren Ihre Werte im Suggestibilitätstest ziemlich gut, sie lagen unter dem Durchschnitt einer Vergleichstichprobe. Demnach lassen Sie sich nicht leicht durch suggestive Einflüsse irreleiten. Das spricht für eine gute allgemeine Wahrnehmungsleistung. Die Augenmessung legt allerdings nahe, dass Sie sich teilweise doch sehr mit den reißerischen Details beschäftigt haben. Eigentlich sollten beide Ergebnisse in eine Richtung deuten, aber es kann natürlich sein, dass Ihr Verhalten nichts mit Ihrer generellen Ablenkbarkeit zu tun hatte. Vielleicht ist das nur themenspezifisch. Es gibt ja solche Effekte, manche sehen sich gerne Sexfilmchen an oder manche können mit Gewaltdarstellungen nicht angemessen umgehen. Dafür benötigen wir die EEG-Ableitung, hier wird sich zeigen, ob Sie das beachtet haben, was Sie sollten oder ob andere Gewohnheiten oder Eigenschaften Sie davon abgehalten haben und welche Punkte genau Ihr Problem waren. Sie wundern sich vielleicht, aber es gibt tatsächlich Leute, die sich trotz der Instruktion und der Tatsache, dass es sich bei den Darstellungen um schwere Straftaten handelt, an den nackten Tatsachen oder den Gewaltdarstellungen erfreuen. Um festzustellen, ob diese Erklärung auch für Sie zutrifft führen wir nun diese EEG-Ableitung durch.“

Diese Vorhaltung sollte von den Probanden differenziert wahrgenommen werden, je nachdem bezüglich welches Films ihnen eine schlechte Beobachtungsleistung suggeriert worden war. Da die EOG-Registrierung als ein Verfahren bezeichnet worden war, das nur ansatzweise Hinweise auf das Verhalten der Probanden liefern konnte und die Feedbackbögen nicht von den Untersuchungsleitern zur Beurteilung

herangezogen worden waren, bestand für den Probanden eine Chance, den Verdacht abzuweisen, indem er den Anforderungen der folgenden PAG gerecht wurde.

Die Probanden erfuhren, dass die Aufgabe derjenigen ähneln würde, die sie während des Übungsdurchgangs hatten. Der Unterschied bestehe darin, dass sich einige der Items auf die Filme beziehen würden. Während eines ersten Durchgangs würden sich die Items auf den einen, während des zweiten Durchgangs auf den anderen Film beziehen. Neben einer Reihe von Wörtern, die nicht aus den Filmen bekannt waren, würden sie solche sehen, die sie instruktionsgemäß während der jeweiligen Filmdarbietung hätten beobachten sollen. Diese Items sollten sie mit einem Druck auf die rechte Taste als Item, dass sie beachtet hätten, bestätigen. Diese Items sollten die Probanden vor der EEG-Ableitung wiederum auswendig lernen. Damit stellten diese die tatunspezifischen Items dar. Die Probanden erfuhren außerdem, dass solche Items auftauchen würden, die mit denjenigen Aspekten des jeweiligen Films in Zusammenhang stünden, mit denen sie sich nicht hätten beschäftigen sollen. Damit wurde Bezug genommen auf die gewalt- beziehungsweise sexbezogenen Detailinformationen der Filme, die hier die tatspezifischen Items darstellten. Diese würden sie aber, wenn sie sich zuvor instruktionsgemäß verhalten hätten, auch während der EEG-Ableitung ohne Schwierigkeit als irrelevant beurteilen können. Wenn sie diese Items nicht beachtet hatten, sollten die Probanden entsprechend den linken Knopf drücken und das entsprechende tatspezifische Item den irrelevanten oder neutralen Items zuordnen. Je schneller und korrekter sie also die tatspezifischen Items herausfiltern könnten, ohne sich von den tatspezifischen Items beeinflussen zu lassen, um so mehr spreche dies dafür, dass sie sich während der Filmdarbietung instruktionsgemäß verhalten hätten. Umgekehrt gelte, dass derjenige, der sich während der Filme von sex- oder gewaltbezogenen Aspekten habe ablenken lassen, diese Reaktionen eventuell auch auf die tatspezifischen Items zeigen würde. So könne von den Reaktionen auf die tatspezifischen Reize auf die Verhaltensweise bei den Filmen geschlossen werden.

Dies implizierte, dass Probanden, die sich eines instruktionsgemäßen Verhaltens aufgrund der Feedbackbögen sicher waren, den tatspezifischen Items gelassen entgegensehen konnten. Probanden, denen eine schlechte Beobachtungsleistung zurückgemeldet worden war, sollten annehmen, dass sie auf die tatspezifischen Items

stark reagieren würden. Da sie jedoch den Eindruck vermeiden wollten, sie hätten sich nicht instruktionsgemäß verhalten, mussten sie versuchen tatspezifische Items möglichst nicht zu beachten. Dies kann man nach der Definition von Köhnken (1990, siehe Kapitel 2) als Täuschungsverhalten klassifizieren. Dieses Täuschungsverhalten sollte anhand der P300-Komponenten identifiziert werden.

Es folgten sechs Blöcke mit je 144 Itemdarbietungen. Das heißt, jedes Item der jeweiligen Itemliste wurde pro Block viermal in zufälliger Reihenfolge dargeboten. Insgesamt wurden dementsprechend 864 Items dargeboten, was insgesamt einer Darbietungszeit von ca. 48 Minuten entspricht. Drei aufeinanderfolgende Blöcke bezogen sich auf die Vergewaltigungsszene, drei weitere aufeinanderfolgende Blöcke auf die Mordszene. Die Reihenfolge der Blockgruppen wurde permutiert. Vor jedem Block musste die dazugehörige Lernliste, bestehend aus sechs tatunspezifischen Items, bis zum Erreichen des Kriteriums „fehlerlose, freie Wiedergabe“ wiederholt werden. Während der Ableitung waren die Probanden allein im Untersuchungsraum, die Untersuchungsleiterin kam jeweils in den Pausen hinzu, um die jeweilige Liste abzufragen.

Nach der Ableitung wurden die Elektroden entfernt. Die Probanden bearbeiteten den Rekognitionstest. Das Nachgespräch erfolgte in Form eines halbstandardisierten Interviews in Anlehnung an Aronson (1994), um dem Probanden adäquat zu vermitteln, dass er sich nicht einfältig oder dumm verhalten habe. Es wurde vielmehr verdeutlicht, dass sich unter den Anforderungen der Untersuchung alle Probanden täuschen lassen und dies eine notwendige Bedingung für den Erkenntnisgewinn der Untersuchung sei.

Zu Beginn der Nachbesprechung wurde jeder Proband ermutigt, seine allgemeinen Eindrücke vom Experiment zu formulieren und jede noch offene Frage zu stellen. Er wurde gebeten, das Anliegen der Untersuchung mit eigenen Worten zu beschreiben und anzugeben, ob ihm irgendein Teil der Untersuchung merkwürdig, verwirrend oder störend vorkam. Dies geschah, um herauszufinden, ob der Proband misstrauisch bezüglich des tatsächlichen Untersuchungsanliegens gewesen war. Anschließend wurde die Aufklärung eingeleitet, indem der Proband gefragt wurde, ob er sich noch einen weiteren Zweck der Untersuchung vorstellen könne. Das tatsächliche Untersuchungsanliegen wurde ausführlich erläutert. Im Rahmen des Nachgesprächs wurde außerdem eruiert, ob die Probanden korrekt erinnerten, bezüglich welches Filmes

sie eine bessere Leistung im Feedbackbogen hatten. Außerdem wurden den Probanden vier Fragen gestellt, die eine Einschätzung darüber ermöglichen sollten, inwiefern die Probanden durch die Untersuchung belastet wurden, beziehungsweise inwiefern sich die Täuschung der Probanden über das Untersuchungsanliegen negativ auswirkte. Die Fragen lauteten: „Reichte Ihnen die gegebene Erklärung aus, oder finden Sie das Vorgehen trotzdem nicht gerechtfertigt?“, „Gab es einen Zeitpunkt, zu dem Sie das Experiment richtig unangenehm fanden?“, „Wie beurteilen Sie Ihre Teilnahme im Nachhinein?“ und „Würden Sie noch einmal an einem psychologischen Experiment teilnehmen, auch wenn Sie annehmen müssten, dass Sie wieder getäuscht würden?“. Anschließend wurden den Probanden eventuell aufgetretene Fragen beantwortet. Die Probanden erhielten unabhängig von ihrer Leistung eine Fahrkostenerstattung.

6.5 Validierung der Methodik

6.5.1 Überprüfung der Annahme: Das Bonner Szenario erzeugt eine andere emotionale und motivationale Voraussetzung als ein Mock Crime

Im Rahmen einer Diplomarbeit (Sartory, 1999) wurde das Bonner Szenario einer ersten Untersuchung unterzogen. Es wurde den typischen Elementen eines Mock Crimes gegenüber gestellt. Zum einen sollte geklärt werden, ob man in Bezug auf die Bedingung „Bonner Szenario“ im Sinne von Aronson (1994) eher von experimenteller Realitätsnähe sprechen kann als in Bezug auf das Mock Crime, d. h., inwieweit zwingt das jeweilige Szenario den Probanden dazu die Angelegenheit ernst zu nehmen und sich intensiv mit den Abläufen zu befassen.

An der Untersuchung nahmen insgesamt 19 Probanden teil. Es handelte sich um eine männliche Zufallsstichprobe (Alter zwischen 19 und 47 Jahren). Ein Proband aus der Bedingung „Bonner Szenario“ erschien nicht zum zweiten Termin, zwei weitere Probanden mussten aufgrund von unvollständigen Daten ausgeschlossen werden.

Der erste Faktor bezieht sich auf das experimentelle Szenario, also Bonner Szenario vs. Mock Crime, der zweite Faktor auf die Bedingung “regelwidriges Verhalten durch schlechte Beobachtungsleistung” vs. “regelgerechtes Verhalten durch gute Beobachtungsleistung”. Die Probanden wurden zufällig einer von vier Bedingungen in einem 2 x 2 Gruppen-Design zugeteilt (siehe Tabelle 7). Bei beiden unabhängigen Variablen handelte es sich um Between-Faktoren.

Tabelle 7

Versuchsplan der Untersuchung von Sartory (1999).

	Art des Szenarios	
Täterschaft	Bonner Szenario	Mock Crime
Regelwidriges Verhalten	n = 4	n = 4
Regelgerechtes Verhalten	n = 4	n = 4

Als abhängige Variablen wurden verschiedene Verhaltensmaße erhoben: Erstens ein globales Maß für das Ausdrucksverhalten. Grundlage dafür waren Urteile von je 5 Ratern pro Proband. Hierzu wurden den Ratern jeweils ca. 30 bis 40 Sekunden dauernde Videosequenzen von den Probanden der Untersuchung gezeigt. Dafür wurde ein an Heinrich (1995) angelehntes semantisches Differential verwendet. Am sinnvollsten ließ sich dieses als einfaktorielle Skala auffassen (Cronbach's Alpha = .88). Die Pole lassen sich als „Entspannung, Spontaneität“ vs. „Anspannung, Kontrolle“ beschreiben. Zweitens wurde die Häufigkeit des „echten“ Lächelns, auch als „Duchennes smile“ bezeichnet (Frank, Ekman & Friesen, 1997), erhoben. Als dritte abhängige Variable wurden die verbalen Reaktionen der Probanden auf die Konfrontation mit dem impliziten Vorwurf registriert.

Die Durchführung des Bonner Szenarios entsprach der unter 6.4 geschilderten Vorgehensweise mit dem Unterschied, dass alle Probanden nur die Vergewaltigungsszene sahen. Außerdem wurde ein Kontrollfragentest unter Ableitung peripher-physiologischer Parameter durchgeführt. Die Mock Crime Bedingung unterschied sich vom Bonner Szenario in folgenden wesentlichen Punkten: Die Probanden wurden zu Beginn des Experiments über den gesamten Ablauf aufgeklärt. Sie erfuhren, dass die Feedbackbögen manipuliert seien und sie beschuldigt werden würden. Sie wurden aufgefordert, sich entsprechend der Versuchsbedingung als gute oder schlechte Beobachter im Sinne der Beobachtungsregeln zu verhalten. Für die Beschuldigung wurde beiden Gruppen aufgetragen, sich so unschuldig wie möglich zu präsentieren. Für den Fall, dass sie als Nichttäter klassifiziert werden würden, wurde ihnen eine monetäre Belohnung von 10 DM versprochen. Die Probanden wurden während der gesamten Untersuchung auf Video aufgezeichnet.

Es wurde eine zweifaktorielle Varianzanalyse berechnet, bei der das Ausdrucksrating als abhängige Variable verwendet wurde. Die Auswertung bezüglich der beiden anderen abhängigen Variablen erfolgte deskriptiv.

Im globalen Ausdrucksrating zeigte sich ein signifikanter Gruppeneffekt für die Beschuldigung ($p < .01$). Probanden mit negativem Feedback werden insgesamt als angespannter beurteilt als Probanden mit positivem Feedback. Ein signifikanter Interaktionseffekt zwischen Szenario und Feedbackbedingung der Probanden besteht jeweils vor und während der Beschuldigung ($p < .05$). Mit anderen Worten: Vor der Beschuldigung werden Probanden im Bonner Szenario, die gut beobachtet haben, im Vergleich zu den Probanden, die schlecht beobachtet haben, als wesentlich geringer angespannt beurteilt als dies in der Mock Crime Bedingung der Fall ist.

Unter der Bedingung „Bonner Szenario/negatives Feedback“ zeigte keiner der Probanden während der Beschuldigungssequenz und dem Kontrollfragentest ein „echtes“ Lächeln, unter der Bedingung „positives Feedback“ zeigten es dagegen fünf Probanden. Im „Mock Crime“ trat „echtes“ Lächeln 13 mal bei Probanden auf, die ein positives Feedback erhalten hatten, und 22 mal bei Probanden mit negativem Feedback. Die verbalen Reaktionen der Probanden auf die Beschuldigung sind in Tabelle 8 aufgeführt. Während sich Probanden, die ein positives Feedback erhalten hatten, nicht spontan zu dem Vorwurf äußerten, unterschieden sich die Reaktionen der Probanden, die sich als Täter betrachteten in der Mock Crime Bedingung von denen im Bonner Szenario. Die Reaktionen der Probanden des Bonner Szenarios reichten von Rechtfertigungsverhalten bis zum Geständnis.

In der Nachbefragung gaben die Probanden der Bedingung „Bonner Szenario“ an, die Feedbackbögen als Grundlage ihrer Leistungseinschätzung akzeptiert zu haben. Sie äußerten ferner, durch die Anschuldigung betroffen gewesen zu sein. Nach der ausführlichen Aufklärung beurteilten sie das Experiment anhand eines Ratings tendenziell jedoch positiver als die übrigen Probanden.

Tabelle 8

Verbale Reaktionen der acht Probanden in der Bedingung „Täter“.

Mock Crime	Bonner Szenario
Keine Entgegnung „Gut! Soll ich jetzt gleich antworten?“ Keine Entgegnung Keine Entgegnung	„Also, ich konnte das gut. An meinen Augenbewegungen kann man das nicht ablesen. Ich bin es beruflich gewohnt, aus den Augenwinkeln zu beobachten.“ „Wie bitte? Also, sie konnten an meinen Augenbewegungen ablesen, dass ich mich sexuell erregen lassen habe?“ „Nee, also ich war schon betroffen von dem Film. Der Film war aber gemacht wie ein Sexfilmchen, aber erregt hat der Film mich nicht.“ „Ja, das war wohl so!“ (Geständnis)

Die Mittelwerte bezüglich des globalen Ausdrucksratings verdeutlichen, dass das Bonner Szenario im Vergleich zur Mock Crime Bedingung geeignet ist, vor Beginn der Beschuldigung deutlichere Verhaltensunterschiede zwischen Probanden mit guter und schlechter Beobachtungsleistung aufzubauen. Die Tatsache, dass „echtes“ Lächeln im Mock Crime häufiger beobachtet wurde, spiegelt vermutlich die geringere Ernsthaftigkeit der experimentellen Situation für die Probanden wider. Für die Probanden im Bonner Szenario gestaltete sich die Situation dagegen eher ernsthaft. Dies spiegeln auch die verbalen Reaktionen wieder. Die Ergebnisse werden durch die geringe Stichprobengröße relativiert.

6.5.2 Validierung einzelner Vorannahmen des Bonner Szenarios

6.5.2.1 Misstrauen gegenüber dem vorgegebenen Untersuchungsanliegen.

Um in Bezug auf die Durchführung des Bonner Szenarios von Realitätsnähe sprechen zu können, ist es notwendig, dass die Probanden das vorgebliche Untersuchungsanliegen als tatsächliches Untersuchungsanliegen akzeptieren. Zu Beginn der Nachbefragung wurden die Probanden daher zu diesem Aspekt befragt (siehe Kapitel 6.4). Alle Probanden nannten als Untersuchungshintergrund sinngemäß „Beobachtungs-

verbesserung unter erschwerten Bedingungen“ beziehungsweise die „unterschiedliche Beeinflussung durch Mord- und Sexthematik“. Auf die Frage, ob den Probanden etwas merkwürdig, verwirrend oder störend vorgekommen sei, machten die Probanden bezüglich des Untersuchungsanliegens nur allgemeine Äußerungen, wie z. B. „Das Ganze war etwas undurchsichtig für mich“ oder „Es war teilweise ganz schön schwierig!“, die nicht darauf schließen ließen, dass sie den Verdacht hatten getäuscht worden zu sein.

6.5.2.2 Feedbackbögen als Kriterium für die Beobachtungsleistung.

Da der Feedbackbogen die Operationalisierung der Täterschaftsbedingung darstellte, war es wichtig, dass die Probanden diesen als Kriterium für die Beurteilung ihrer Beobachtungsleistung heranzogen.

Alle Probanden erinnerten korrekt, in welchem Feedbackbogen sie ein besseres und in welchem sie ein schlechteres Ergebnis erzielt hatten. Damit war die erste Voraussetzung erfüllt. Von 43 Probanden griffen 17 den Feedbackbogen und seine Aussagefähigkeit während des Nachgesprächs an. Kritik richtete sich sowohl gegen die Auswahl der Fragen („Andere Details habe ich mir gemerkt, die sind aber nicht abgefragt worden“) als auch gegen die Auswertung („Die Auswertung war zu streng, im Prinzip hätte ich das richtig gehabt“, „Die Auswertung war unlogisch“). Eine zweifaktorielle Varianzanalyse (Itemtyp, Feedbackbogen angegriffen oder nicht) mit Messwiederholung auf dem ersten Faktor und der maximalen P300-Amplitude als abhängiger Variable ergab keine signifikanten Ergebnisse für den „Feedbackbogen-Faktor“ (Pz: $F(2, 83) = .204$, $p = .816$).

Ungefähr ein Jahr nach der Durchführung der Untersuchung wurden die Probanden aufgerufen, an einer Befragung bezüglich der damaligen Untersuchung teilzunehmen. Eine Stichprobe von 11 Probanden konnte dafür gewonnen werden. Die Befragung wurde als halbstandardisiertes Interview durchgeführt. Die Probanden lieferten zunächst einen freien Bericht. Anschließend wurden in Anlehnung an das Kognitive Interview (Fisher & Geiselman, 1992) mit Hilfe der Technik „Wiederherstellen des Kontextes“ weitere Informationen erhoben. Gegenstand dieser Befragung war unter anderem auch die Beobachtungsleistungen bei der Filmdarbietung. Insgesamt stimmten die Erinnerungen von sieben Probanden mit den Ergebnissen der

Feedbackbögen überein. Das heißt, diese Probanden erinnerten sich an schlechte Beobachtungsleistungen bezüglich desjenigen Films, zu dem ihnen der Feedbackbogen ein negatives Feedback gegeben hatte (Probanden (Pbn.) 2, 8, 32, 34, 38, 40, 41). Zwei Probanden gaben an, sich nicht mehr erinnern zu können (Pbn. 17, 21). Ein Proband erinnerte sich daran, bezüglich beider Filme eine gute Leistung gehabt zu haben (Pb. 6). Die Erinnerungen eines weiteren Probanden verkehrten die damals erzielten Feedbackbogenergebnisse ins Gegenteil (Pb. 39).

6.5.2.3 Wirksamkeit des Vorhalts und resultierende Täuschungsmotivation.

Der Vorhalt gegenüber den Probanden, sie hätten sich möglicherweise an sexuellen oder gewaltbezogenen Aspekten der Filme erfreut, wurde auf seine Wirksamkeit geprüft. Das Ziel des Vorhalts war, die Probanden dazu zu motivieren während der EEG-Ableitung tatunspezifische Items möglichst schnell herauszusuchen und gleichzeitig tatspezifische Items nicht zu beachten. Das Nichtbeachten tatspezifischer Items erforderte von den Tätern, hier vertreten durch Probanden, denen ein schlechtes Ergebnis im Feedbackbogen rückgemeldet worden war, Täuschungsverhalten (siehe Kapitel 6.4).

Unmittelbar nach der EEG-Ableitung waren die Probanden im Anschluss an die Aufklärung über Ziel und Anliegen der Untersuchung gefragt worden, ob sie durch den Vorhalt motiviert worden seien, bei der EEG-Untersuchung gut abzuschneiden. Alle Probanden bejahten diese Frage. Es ist jedoch anzunehmen, dass bei der Beantwortung dieser Frage Aspekte der sozialen Erwünschtheit eine entscheidende Rolle spielten.

Daher wurde in der im letzten Kapitel beschriebenen Nachbefragung das Thema der Motivation aufgegriffen. Die Annahme war, dass für Probanden, die sich ein Jahr nach der Untersuchung an den Vorhalt erinnern können, diese Vorhaltung bedeutsam gewesen sein muss. Unbedeutsame Informationen werden schnell vergessen (Ebbinghaus, 1885). Eine gedankliche Beschäftigung mit dem Vorwurf macht es wahrscheinlicher, dass die Information auch nach einem Jahr noch erinnert werden kann (Craik & Lockhart, 1972). Um einen Eindruck von der Motivation der Probanden zu erhalten, wurden Erinnerungen bezüglich der Reaktion auf den Vorwurf wörtlich festgehalten.

Zehn von elf Probanden erinnerten den Vorwurf inhaltlich korrekt. Davon berichtete 1 Proband spontan im freien Bericht davon. Im weiteren Verlauf des

Interviews berichteten 7 Probanden auf die Frage: „Welchen Zweck hatte die EEG-Ableitung“ und 2 Probanden nach weiteren Zusatzfragen von dem Vorwurf. Ein Proband konnte sich daran nicht mehr erinnern.

Anhand der Statements der Probanden 6, 8, 32, 34, 38, 39, 40 und 41 lässt sich ablesen, dass die Probanden sich mit der Möglichkeit des Zutreffens des Vorwurfes beschäftigten. Sie waren daher wahrscheinlich motiviert, bei der EEG-Ableitung gut abzuschneiden. Die Probanden 2 und 21 gaben an, von dem Vorhalt eher nicht betroffen gewesen zu sein, erinnerten sich aber daran. Proband 17 erinnerte sich nicht an den Vorhalt.

6.5.2.4 Probandenbelastung.

Auf die Frage „Reichte Ihnen die Erklärung für das Vorgehen der Untersuchung aus?“ antworteten 39 (90,7%) Probanden mit „Ja“ und 4 Probanden (9,3%) mit „Nein“. Von 43 Probanden bejahten 21 (48,8%), dass es einen Zeitpunkt gegeben habe, zu dem das Experiment richtig unangenehm gewesen sei, 22 (51,2%) verneinten dies. Lediglich 3 (7%) Probanden beurteilten Ihre Teilnahme im Nachhinein negativ, 40 (93%) Probanden dagegen positiv. Probanden empfanden ihre Teilnahme positiv, weil sie den Untersuchungshintergrund begrüßten, weil sie das Vorgehen interessant fanden und/oder weil sie es für eine interessante Erfahrung hielten. Negativ beurteilten die Probanden ihre Teilnahme, weil sie den Untersuchungsgegenstand, die PAG, ablehnten. Ein Proband empfand seine Teilnahme auf Grund der Täuschung als negativ. Ohne Einschränkung bejahten 31 (72%) Probanden die Frage „Würden Sie noch einmal an einem psychologischen Experiment teilnehmen, auch wenn Sie annehmen müssten, dass Sie vielleicht wieder getäuscht würden?“, 11 (25,6%) Probanden würden mit Einschränkungen noch einmal teilnehmen und ein (2,3%) Proband gab an, auf keinen Fall wieder teilnehmen zu wollen.

6.6 Datenaufbereitung und statistische Auswertung

6.6.1 Datenaufbereitung

Die Auswertung der EEG-Daten erfolgte mit Hilfe des „Konstanzer Formats“ (Berg, 1999). Aus dem kontinuierlich aufgezeichneten EEG wurden pro Bedingung 432 Epochen von 1100 ms Länge, das bedeutet 100 ms vor bis 1000 ms nach der

Itemdarbietung, ausgeschnitten. Berücksichtigt wurden nur Roh-EEG-Epochen, die aus korrekten Antworten resultierten. Ausgesondert wurden solche Trials, bei denen der Proband eine falsche Antwort gegeben hatte, keine Antworttaste bedient hatte oder die Reaktionszeit 1000 ms überschritten hatte. Das Minimum an falschen Reaktionen lag in der vorliegenden Untersuchung bei 1, das Maximum bei 35. Der Mittelwert betrug 9,19 bei einer Standardabweichung von 7,86. Insgesamt wurden 395 Fehler bei 37152 Trials gemacht. Das entspricht 1,06%.

Die drei EEG-Kanäle wurden mit Hilfe eines Regressionsverfahrens nach Berg (1999) hinsichtlich der EOG-Artefakte (Blinkbewegungen) bereinigt. Zusätzlich wurden alle Epochen nach der Korrektur visuell inspiziert und diejenigen manuell eliminiert, die sonstige Artefakte, z.B. Bewegungsartefakte, enthielten. Schließlich wurden die korrigierten Daten nochmals gefiltert (Tiefpassfilter 10 Hz).

Die artefaktfreien EEG-Kurven wurden für jeden Probanden, jede Bedingung, jeden Ableitort und jeden Itemtyp gemittelt, so dass aus der Mittelung für jeden Probanden 18 EKP-Kurven resultierten. Im Folgenden wird mit der Bezeichnung „EEG-Kurven“ auf die noch nicht gemittelten aber korrigierten EEG-Epochen Bezug genommen. Mit „EKP-Kurven“ werden die gemittelten Signale bezeichnet.

6.6.2 Auswertung zur Überprüfung der Annahme 1

Um zu zeigen, dass unabhängig von der Täuschungsintention und dem Tatwissen der Probanden im Vergleich zu den neutralen Reizen auf tatspezifische Reize hin eine ausgeprägte P300 auftritt, wurde zunächst ein t-Test für abhängige Stichproben berechnet. Verglichen wurden die maximalen P300-Amplituden, die auf tatspezifische Reize aufgetreten waren, mit denen, die auf neutrale Reize hin aufgetreten waren. Die maximalen P300-Amplituden wurden jeweils in einem Zeitfenster von 350 bis 1000 ms nach dem Reiz bestimmt. Bei einer Stichprobengröße von 86, einem $\alpha = .05$ und der Annahme großer Effekte beträgt die Power bei einseitiger Testung (berechnet nach G-Power, Faul & Erdfelder, 1992) .98. Das heißt, bei der angegebenen Stichprobengröße werden große Effekte mit einer Wahrscheinlichkeit von 98% erhalten.

Wie in Kapitel 3.2.3.1 geschildert, weist die P300-Komponente als ein charakterisierendes Merkmal ein parietales Maximum auf. Um zu zeigen, dass es sich bei den positiven Wellen um die P300-Komponente handelte, wurde eine einfaktorielle

Varianzanalyse mit Messwiederholung auf dem Faktor „Ableitort“ (Fz, Cz, Pz) berechnet. Als abhängige Variablen dienten die maximalen Amplituden der auf neutrale, tatspezifische und tatunspezifische Reize aufgetretenen EKP-Kurven. Die Varianzanalyse wurde getrennt für die unterschiedlichen Itemtypen berechnet.

6.6.3 Auswertung zur Überprüfung der Annahmen 3 und 4

Um der Frage nach der Klassifizierbarkeit der Probanden in Täter und Nichttäter nachzugehen, wurde zunächst eine erste visuelle Inspektion der EKP-Kurven vorgenommen. Entsprechend der Annahmen 3b) ($S > N$ für Täter) und 4b) ($S \leq N$ für Nichttäter) wurden die maximalen Amplituden bezüglich der neutralen Reize mit denen verglichen, die auf die tatspezifischen Reize hin aufgetreten waren. War die Amplitude der tatspezifischen Reize größer als die der neutralen, wurde der Proband als Täter klassifiziert. Wenn die Amplitude der tatspezifischen Reize kleiner oder gleich der der neutralen Reize war, wurde der Proband als Nichttäter klassifiziert. Ließ sich keine eindeutige P300-Komponente oder kein deutliches Maximum ausmachen, wurde der Proband als „nicht klassifizierbar“ beurteilt. Ein Vorzeichentest wurde durchgeführt, um die Annahme zu überprüfen, dass zwischen der Täter- und der Nichttäterbedingung ein Unterschied hinsichtlich der Klassifikationen bestand. Der Vorzeichentest wurde getrennt für diejenigen Probanden berechnet, die Täter bezüglich der Mordthematik beziehungsweise Nichttäter bezüglich der Vergewaltigungsthematik waren, und diejenigen, die Nichttäter bezüglich der Mordthematik beziehungsweise Täter bezüglich der Vergewaltigungsthematik waren.

Um zu testen, ob sich Gruppenunterschiede zwischen Täter- und Nichttäterbedingung ergeben würden, wurde eine 2 (Täterschaft) x 2 (Deliktthema) x 3 (Itemtyp) faktorielle Varianzanalyse mit Messwiederholung auf dem ersten und dritten Faktor durchgeführt. Die Varianzanalyse wurde getrennt für die Ableitungen bei Fz, Cz und Pz berechnet. Als abhängige Variable dienten die maximalen P300-Amplituden. Bei einer Stichprobengröße von 43, einem $\alpha = .05$ und der Annahme mittlerer Effekte beträgt die Power (berechnet nach G-Power, Faul & Erdfelder, 1992) .84. Das heißt, bei der angegebenen Stichprobengröße werden mittlere Effekte für die Variation der Täterschaftsbedingung mit einer Wahrscheinlichkeit von 84% erhalten.

Für die Einzelfallauswertung wurden zwei unterschiedliche Bootstrapverfahren verwendet. Mit Hilfe des Verfahrens 1 wurden die Annahmen 3a) ($(S \sim U) \neq N$ für Täter) und 4a) ($(S \sim N) \neq U$ für Nichttäter), mit Hilfe des Verfahrens 2 die Annahmen 3b) und 4b) für den Einzelfall überprüft. Das Bootstrapverfahren 1 wurde zuerst von Wasserman und Bockenholt (1989) beschrieben. Anhand von Kreuzkorrelationen können mit Hilfe dieses Verfahrens Kurvenverläufe verglichen werden. Das zuerst von Allen und Iacono (1997) beschriebene Bootstrapverfahren 2 vergleicht Amplitudenmaxima. Beide Verfahren liefern sogenannte „Bootstrapindizes“, die einen Wert zwischen 0 und 100 annehmen können. Ein kleiner Wert indiziert Täterschaft, ein großer Nichttäterschaft. Die beiden Verfahren werden in Anhang E ausführlich erläutert.

Um einen Wert für die Diskriminationsfähigkeit bezüglich der Täter und Nichttäter zu erhalten, wurden Receiver-Operating-Characteristic-Curves (ROC) für die Bootstrapindizes verwendet. Der ROC-Graph trägt die Rate der korrekt-positiven Entscheidungen (Maß für die Sensitivität) gegen die Rate der falsch-positiven Entscheidungen (Maß für 1 - Spezifität) anhand der Bootstrapindizes sukzessiv für alle Cut-Off-Points auf (Mossman & Somoza, 1991). Der „Area Under Curve“-Index (AUC-Index), der anhand der Fläche unter der ROC-Kurve berechnet wird, stellt ein Maß für die Diskriminationsfähigkeit einer diagnostischen Methode dar. Der AUC-Index gibt die Wahrscheinlichkeit an, mit der ein diagnostisches Instrument ein Paar von Probanden, das in der vorliegenden Arbeit aus einem Täter und einem Nichttäter bestehen würde, korrekt zuordnet (Green & Swets, 1966). Der AUC-Index kann Werte zwischen .50 und 1.00 annehmen. Ein Wert von .50 indiziert, dass die ROC-Kurve mit der Diagonalen durch die Punkte (0/0) und (1/1) im Koordinatensystem zusammenfällt. Dieser Wert gibt also an, dass ein Verfahren nicht besser als zufällig zwischen Tätern und Nichttätern diskriminiert. Ein AUC-Wert von 1 gibt an, dass das Verfahren perfekt zwischen den Bedingungen trennt.

6.6.4 Auswertung zur Überprüfung der Annahme 5

Um der Frage nach der Deliktspezifität der Täuschungsaufdeckung mittels der P300-Komponente nachzugehen wurde eine 2 (Täterschaft) x 2 (Deliktthema) x 3 (Itemtyp) faktorielle Varianzanalyse mit Messwiederholung auf dem zweiten und dritten Faktor

durchgeführt. Die Varianzanalyse wurde getrennt für die Ableitungen bei Fz, Cz und Pz berechnet. Als abhängige Variable dienten die maximalen P300-Amplituden. Bei einer Stichprobengröße von 43, einem $\alpha = .05$ und der Annahme mittlerer Effekte beträgt die Power (berechnet nach G-Power, Faul & Erdfelder, 1992) .84. Das heißt, bei der angegebenen Stichprobengröße werden mittlere Effekte für die Variation der Deliktthematik mit einer Wahrscheinlichkeit von 84% erhalten.

6.6.5 Auswertung zur Überprüfung des Einflusses der Kontrollvariablen

Um den Einfluss der Kontrollvariablen auf die Klassifikation der Probanden zu beurteilen, wurden entweder Korrelationen zwischen den maximalen P300-Amplituden und der jeweiligen Kontrollvariablen berechnet oder es wurden zweifaktorielle Varianzanalysen (Itemtyp, Kontrollvariable) mit Messwiederholung auf dem ersten Faktor berechnet. Die Power betrug für große Effektstärken, eine Stichprobengröße von 86 und einem $\alpha = .05$.93 für den Interaktionseffekt der zweifaktoriellen Varianzanalyse. Für den Zwischensubjekteffekt betrug sie .97.

Für die Entdeckung signifikanter Korrelationen bei einseitiger Hypothesentestung, einem N von 86, einem $\alpha = .05$ und der Annahme mittlerer Effekte betrug die Power .89. Für die Kontrollvariable „Rekognitionstest“ liegt die Power für die gleichen Parameter bei .82, weil nur 35 Probanden diese Aufgabe bearbeiteten.

Alle AUC-Werte, t-Tests, ANOVA und Korrelationen wurden mit Hilfe der 10. Version des Statistical Package for Social Sciences (SPSS® 10.0 für Windows) berechnet. Die Bootstrapanalysen wurden mit Hilfe selbstgeschriebener Programme in „Practical Extraction and Report Language“ (Perl-Programme) durchgeführt.

7. Ergebnisse

7.1 Annahme 1: Tatunspezifische Reize rufen ausgeprägte P300 hervor

Zunächst sollte gezeigt werden, dass eine größere P300 auf tatunspezifische Items hin auftritt als auf neutrale Reize. Die t-Tests für abhängige Stichproben wurden getrennt nach Ableitort berechnet und ergaben signifikante Unterschiede zwischen den maximalen P300-Amplituden bezüglich tatunspezifischer und neutraler Items (alle $p < .001$).

Zusätzlich sollte gezeigt werden, dass die als P300 bezeichnete Welle ein parietales Maximum aufwies. Eine einfaktorielle Varianzanalyse mit Messwiederholung auf dem Faktor Ableitort (Fz, Cz, Pz) und der maximalen Amplitude der neutralen, tatunspezifischen und tatspezifischen Reize als abhängiger Variablen erbrachte signifikante Haupteffekte (neutrale: $F(2, 84) = 63.53$; tatspezifisch: $F(2, 84) = 47.75$; tatunspezifisch: $F(2, 84) = 106.16$; alle $p < .01$). Tabelle 9 gibt die Mittelwerte und Standardabweichungen für die maximalen Amplituden wieder.

Tabelle 9

Mittelwerte und Standardabweichungen (μV) der maximalen P300-Amplitudenwerte für drei Ableitorte und drei Itemtypen.

	Fz	Cz	Pz
neutral	4.09 (4.71)	7.65 (6.10)	8.80 (5.98)
tatspezifisch	4.12 (5.37)	7.63 (6.77)	8.89 (7.04)
tatunspezifisch	7.56 (5.93)	13.05 (6.71)	14.60 (6.38)

7.2 Annahmen 3 und 4: Klassifizierbarkeit bezüglich Täterschaft

Die 18 EKP-Kurven für jeden der 43 Probanden werden in den Abbildungen 6 bis 16 wiedergegeben. Tabelle 10 gibt die Klassifikationsergebnisse für den Amplitudenvergleich per Augenschein wieder.

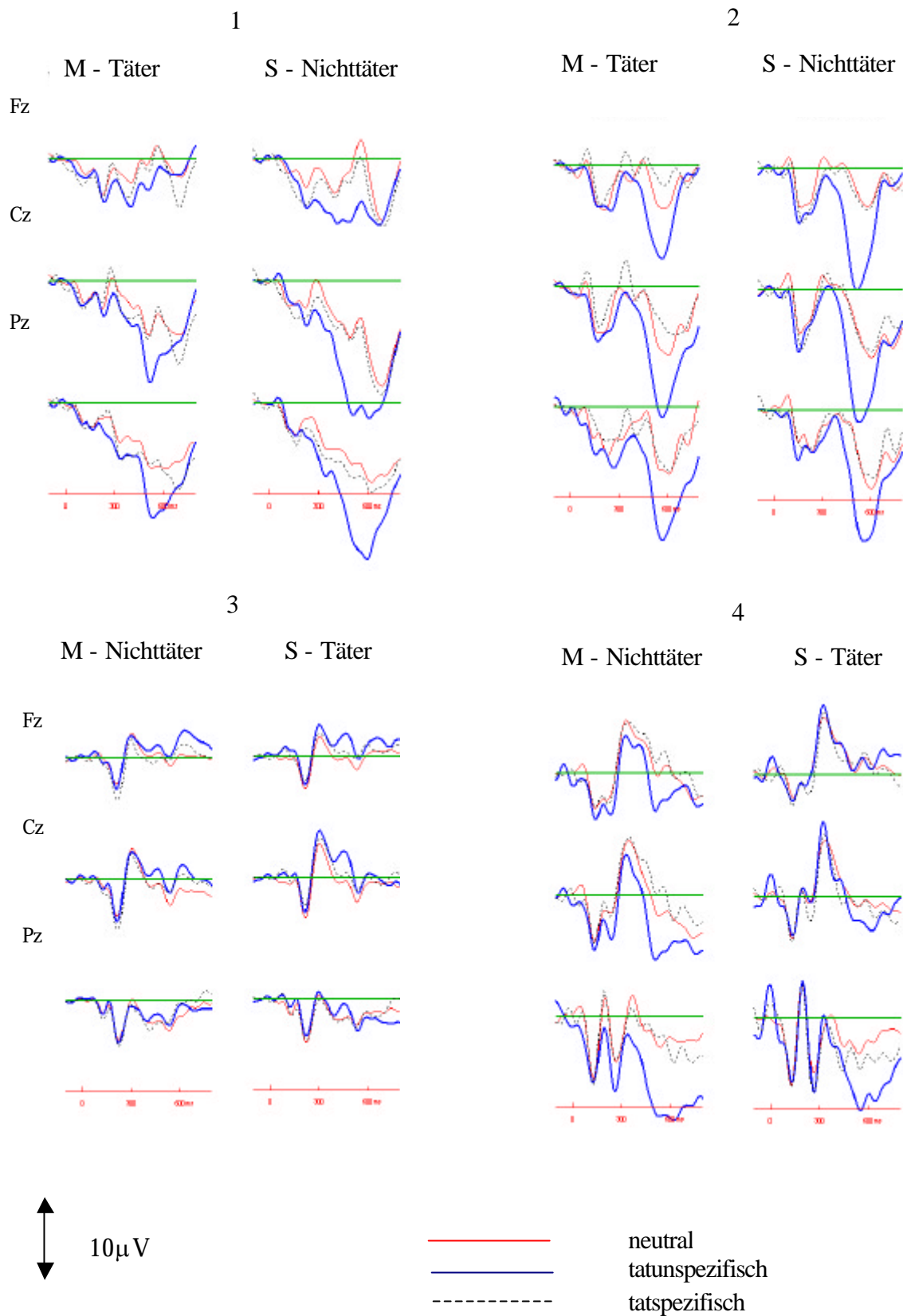


Abbildung 6
 EKP-Kurven von Probanden 1 bis 4 für Täter- und Nichttäterbedingung. Dargestellt sind die Kurven für drei Itemtypen (tatspezifisch, tatunspezifisch, neutral) und drei Ableitorte (Fz, Cz, Pz). (M = Mord, S = Sexualstraftat).

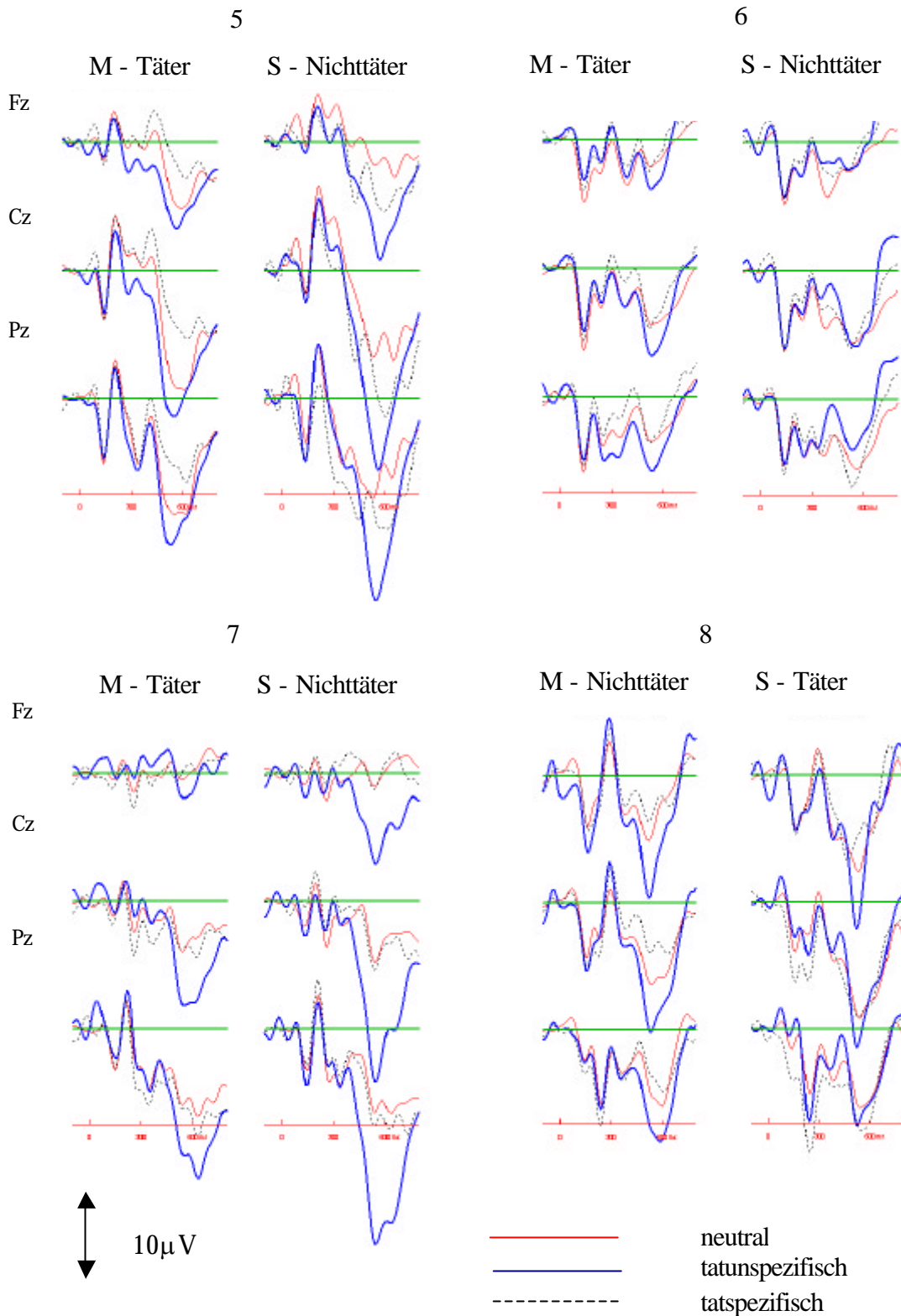


Abbildung 7
 EKP-Kurven von Probanden 5 bis 8 für Täter- und Nichttäterbedingung. Dargestellt sind die Kurven für drei Itemtypen (tatspezifisch, tatunspezifisch, neutral) und drei Ableitorte (Fz, Cz, Pz). (M = Mord, S = Sexualstraftat).

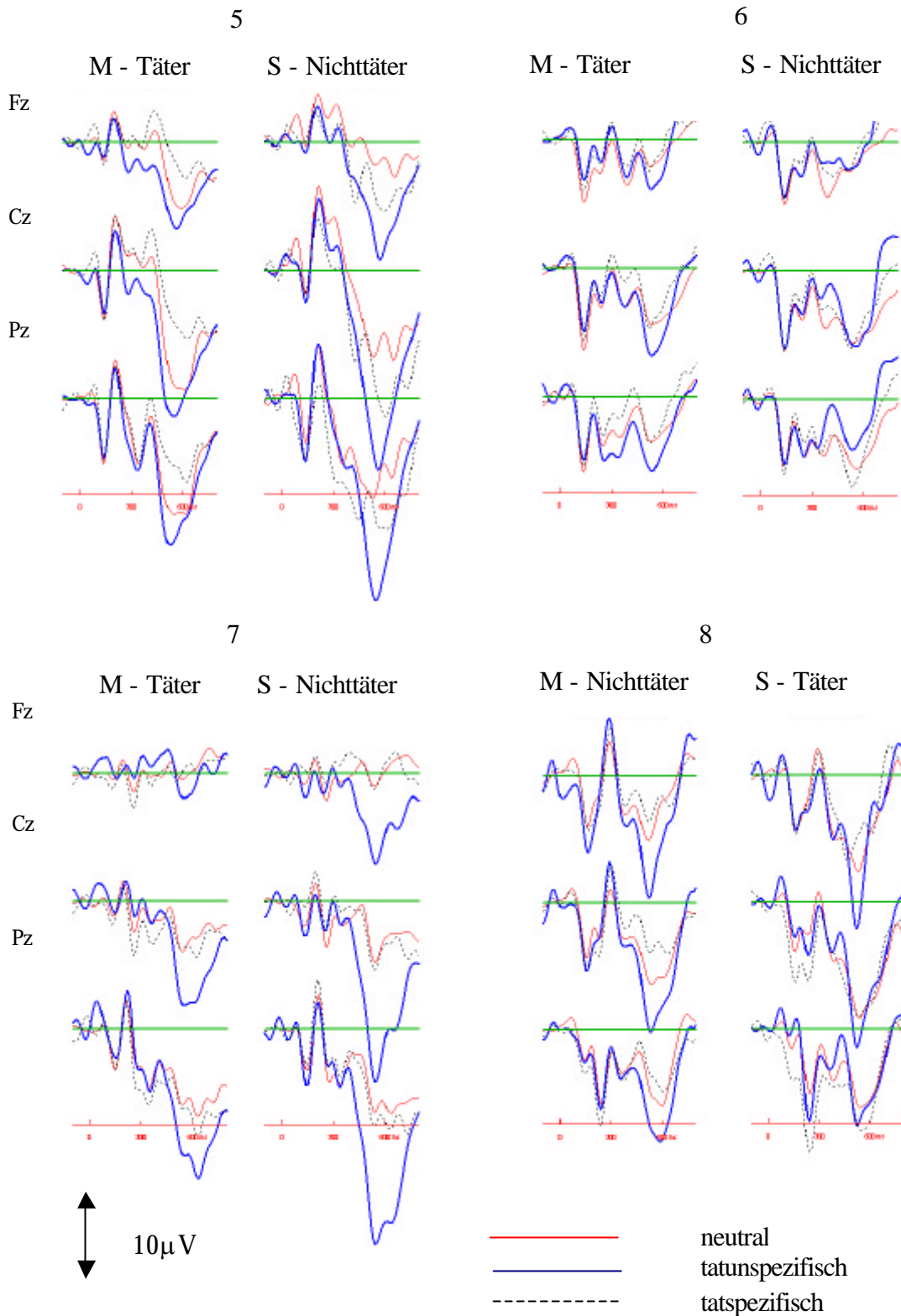


Abbildung 7
 EKP-Kurven von Probanden 5 bis 8 für Täter- und Nichttäterbedingung. Dargestellt sind die Kurven für drei Itemtypen (tatspezifisch, tatunspezifisch, neutral) und drei Ableitorte (Fz, Cz, Pz). (M = Mord, S = Sexualstraftat).

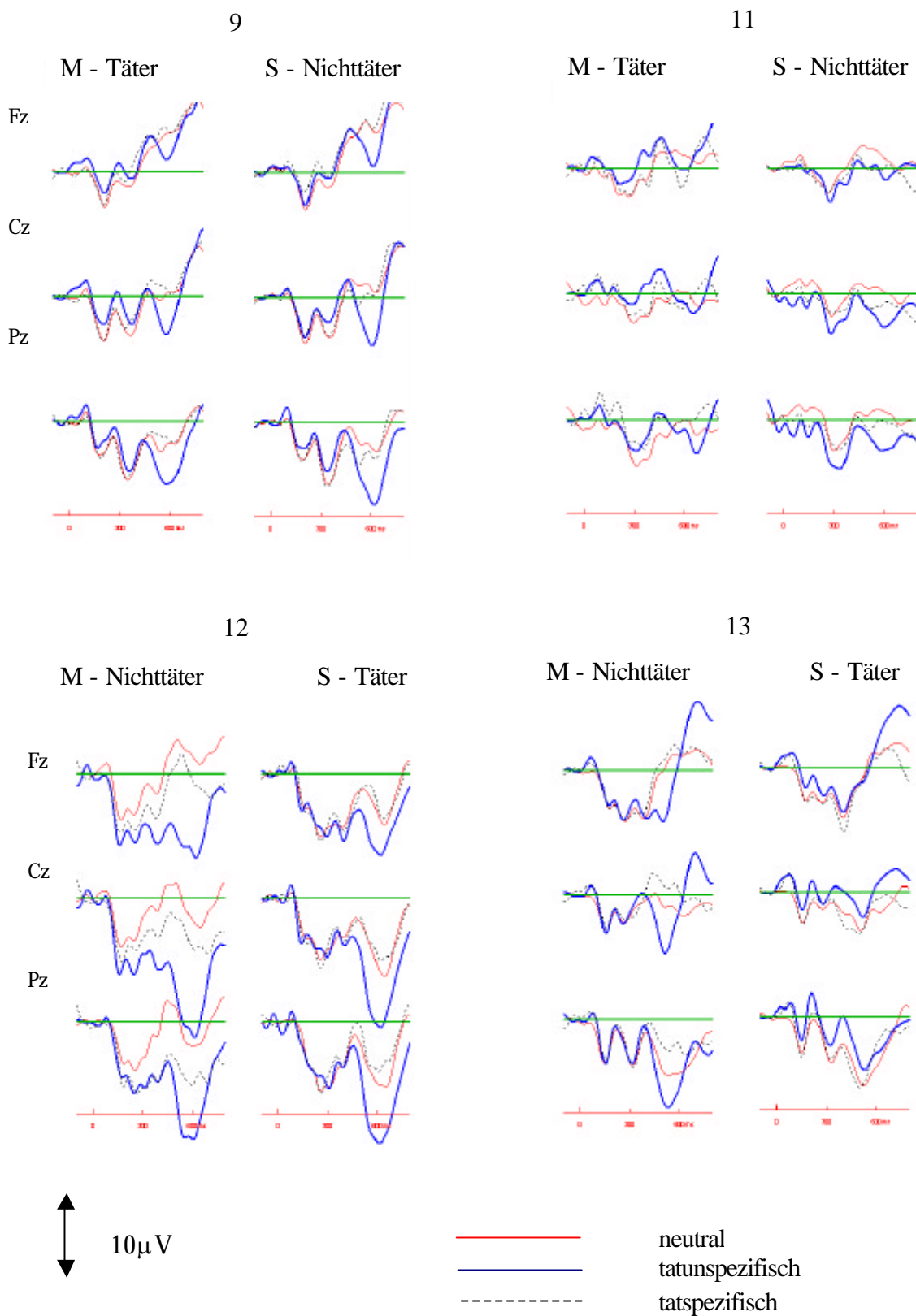


Abbildung 8

EKP-Kurven von Probanden 9 bis 13 für die Täter- und die Nichttäterbedingung. Von Proband 10 wurden keine EEG-Daten registriert. Dargestellt sind die Kurven für drei Itemtypen (tatspezifisch, tatunspezifisch, neutral) und drei Ableitorte (Fz, Cz, Pz). (M = Mord, S = Sexualstraftat).

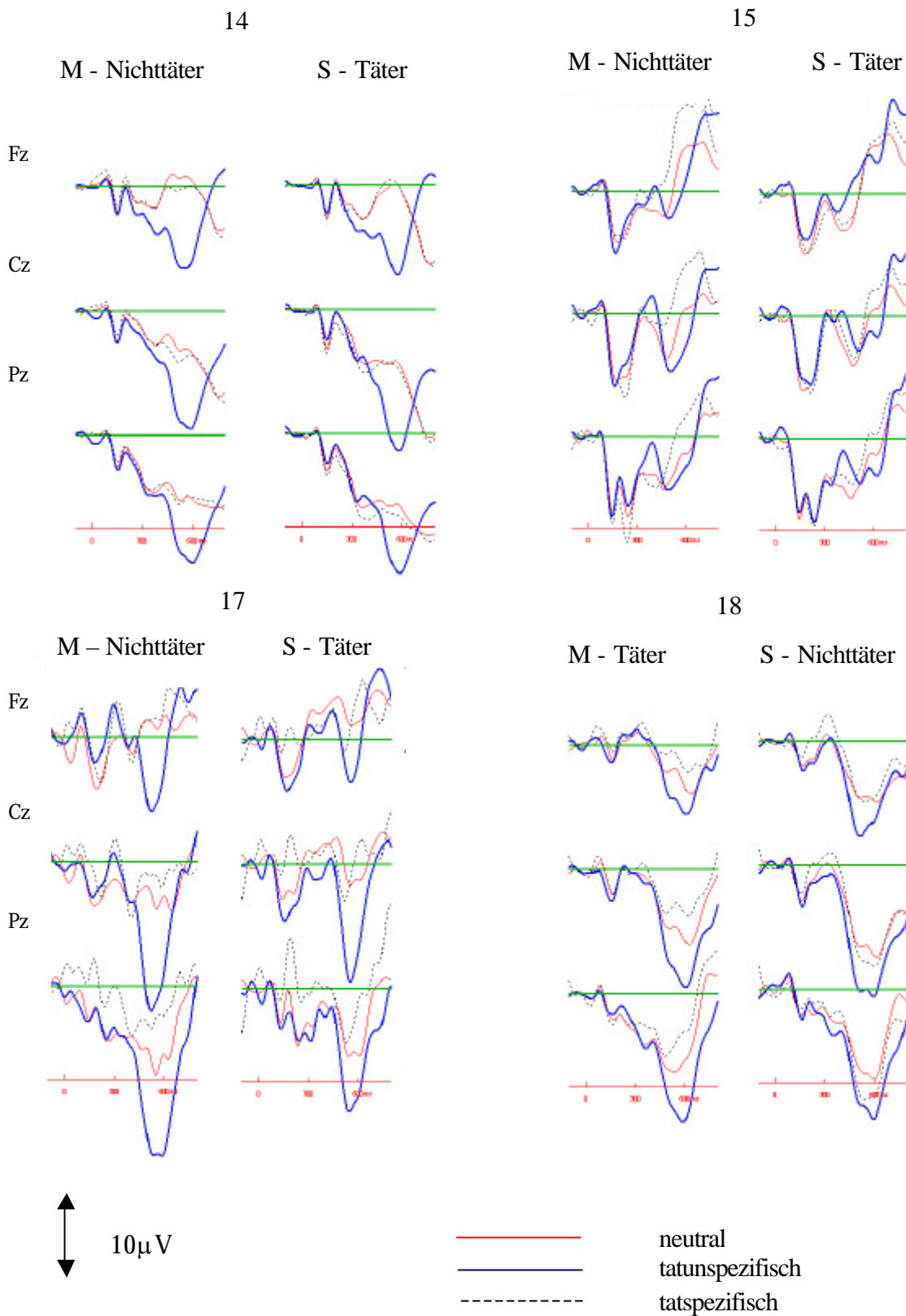


Abbildung 9

EKP-Kurven von Probanden 14 bis 18 für Täter- und Nichttäterbedingung. Proband 16 erschien nicht zum 2. Termin. Dargestellt sind die Kurven für drei Itemtypen (tatspezifisch, tatunspezifisch, neutral) und drei Ableitorte (Fz, Cz, Pz). (M = Mord, S = Sexualstraftat).

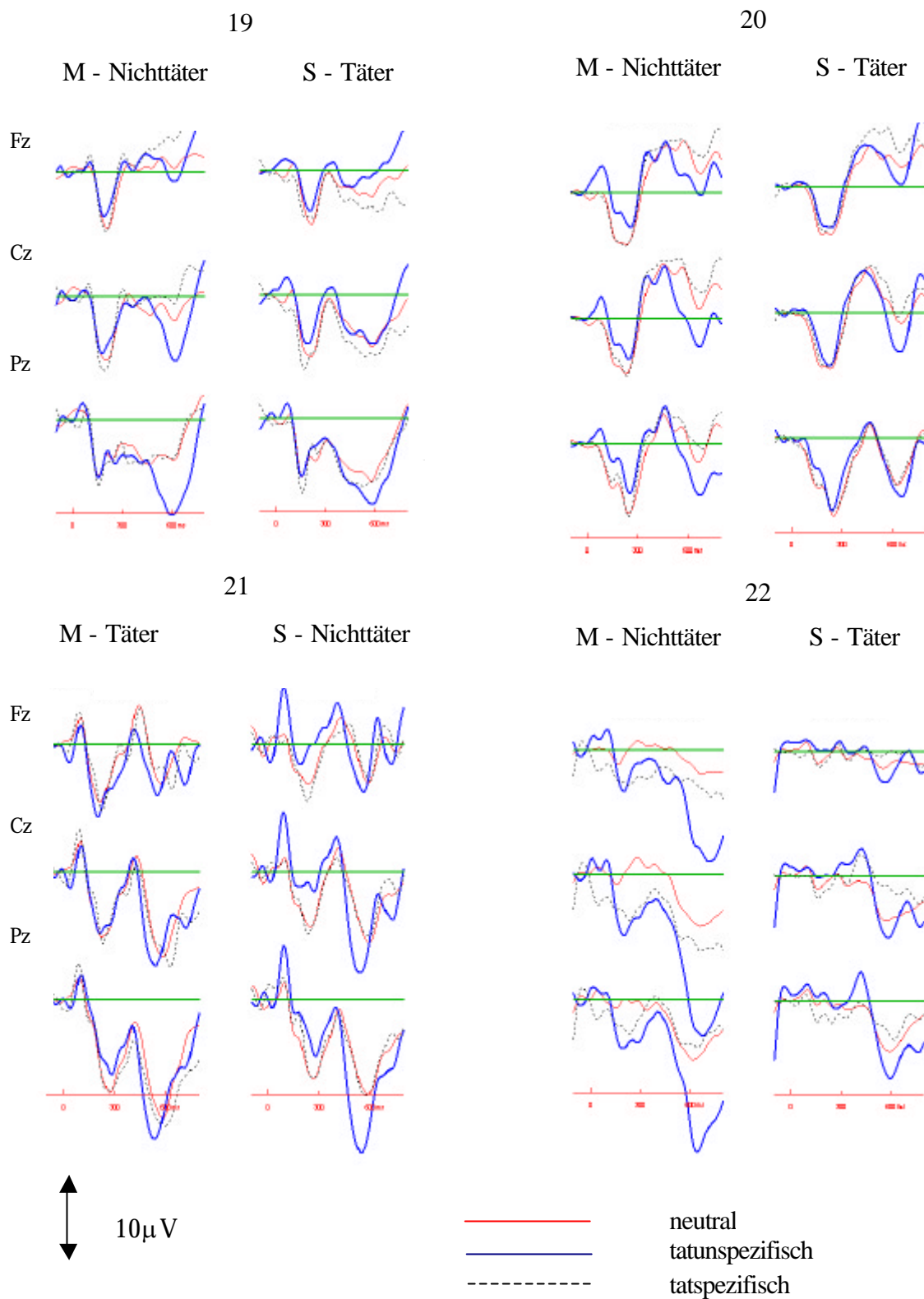


Abbildung 10
 EKP-Kurven von Probanden 19 bis 22 für Täter- und Nichttäterbedingung. Dargestellt sind die Kurven für drei Itemtypen (tatspezifisch, tatunspezifisch, neutral) und drei Ableitorte (Fz, Cz, Pz). (M = Mord, S = Sexualstraftat).

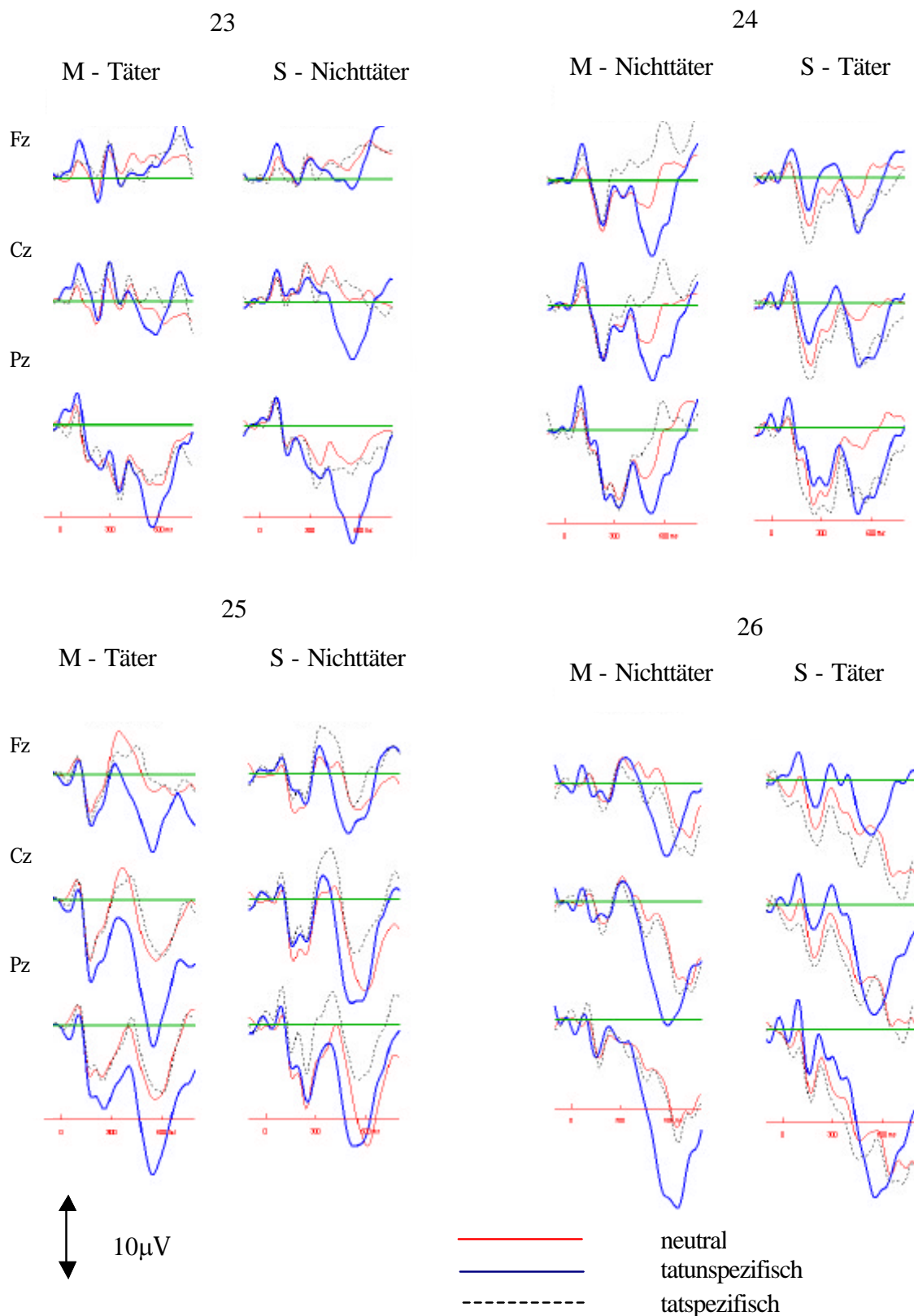


Abbildung 11
 EKP-Kurven von Probanden 23 bis 26 für Täter und Nichttäterbedingung. Dargestellt sind die Kurven für drei Itemtypen (tatspezifisch, tatunspezifisch, neutral) und drei Ableitorte (Fz, Cz, Pz). (M = Mord, S = Sexualstraftat).

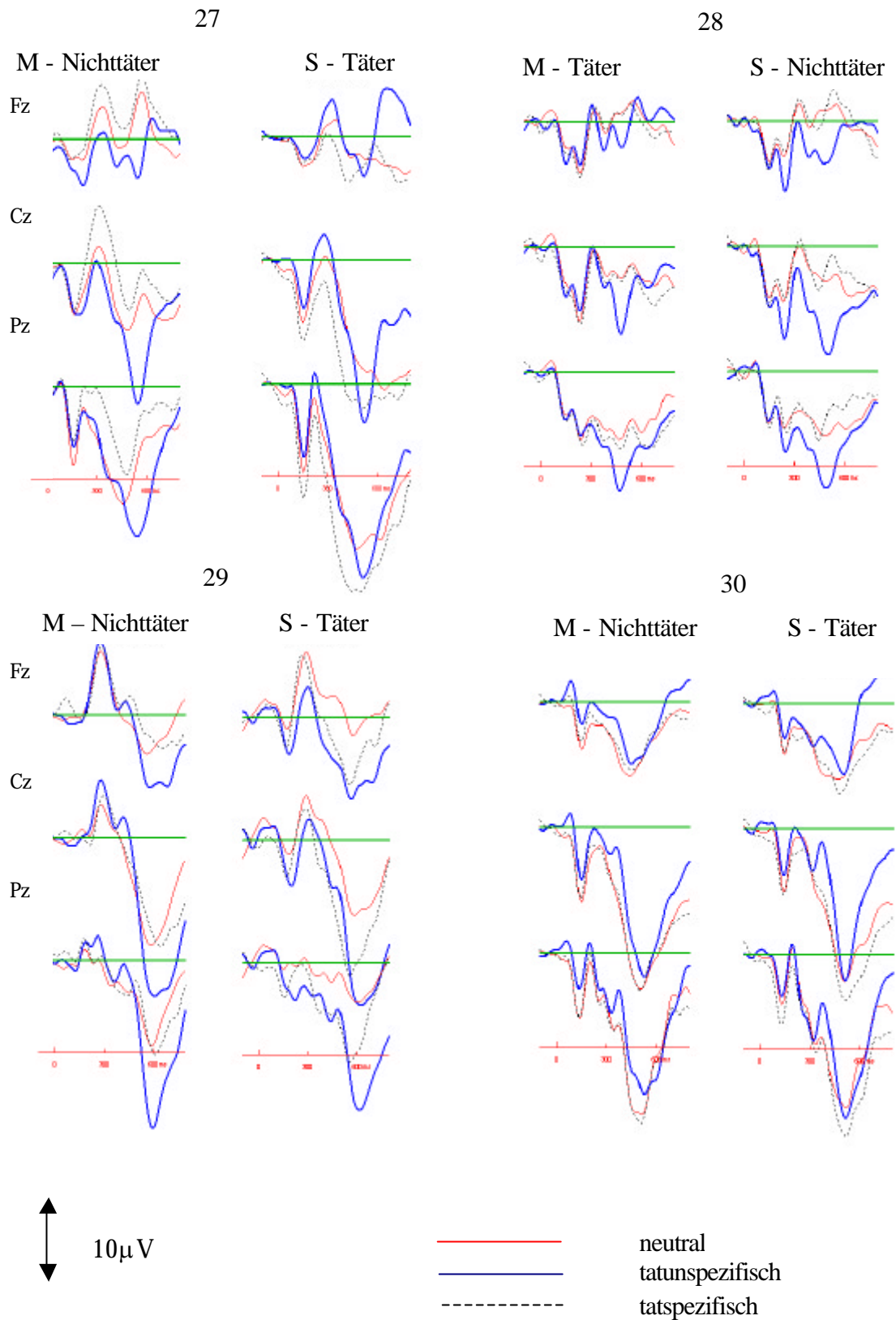


Abbildung 12
 EKP-Kurven von Probanden 27 bis 30 für Täter- und Nichttäterbedingung. Dargestellt sind die Kurven für drei Itemtypen (tatspezifisch, tatunspezifisch, neutral) und drei Ableitorte (Fz, Cz, Pz). (M = Mord, S = Sexualstraftat).

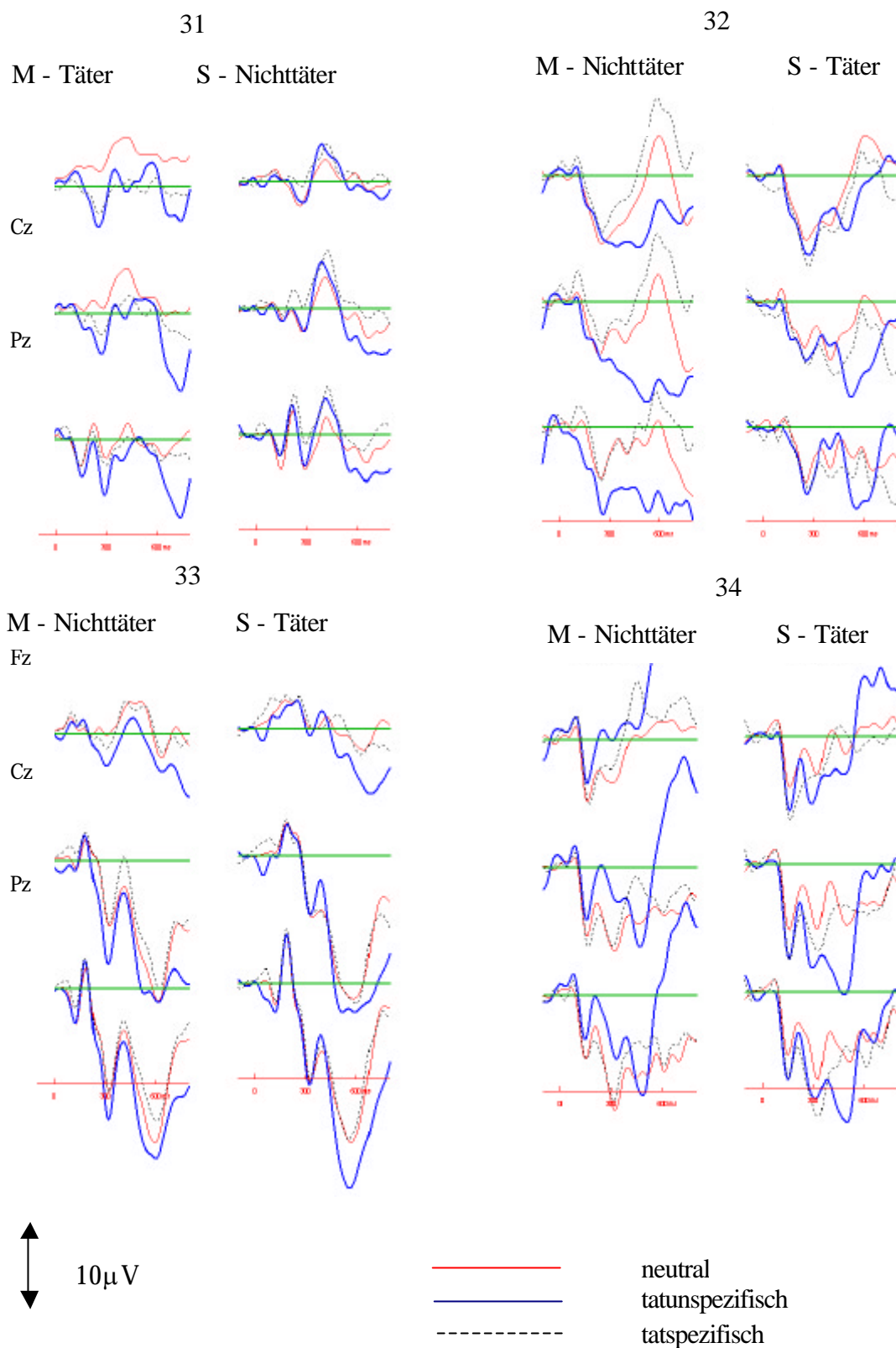


Abbildung 13
 EKP-Kurven von Probanden 31 bis 34 für Täter- und Nichttäterbedingung. Dargestellt sind die Kurven für drei Itemtypen (tatspezifisch, tatunspezifisch, neutral) und drei Ableitorte (Fz, Cz, Pz). (M = Mord, S = Sexualstraftat).

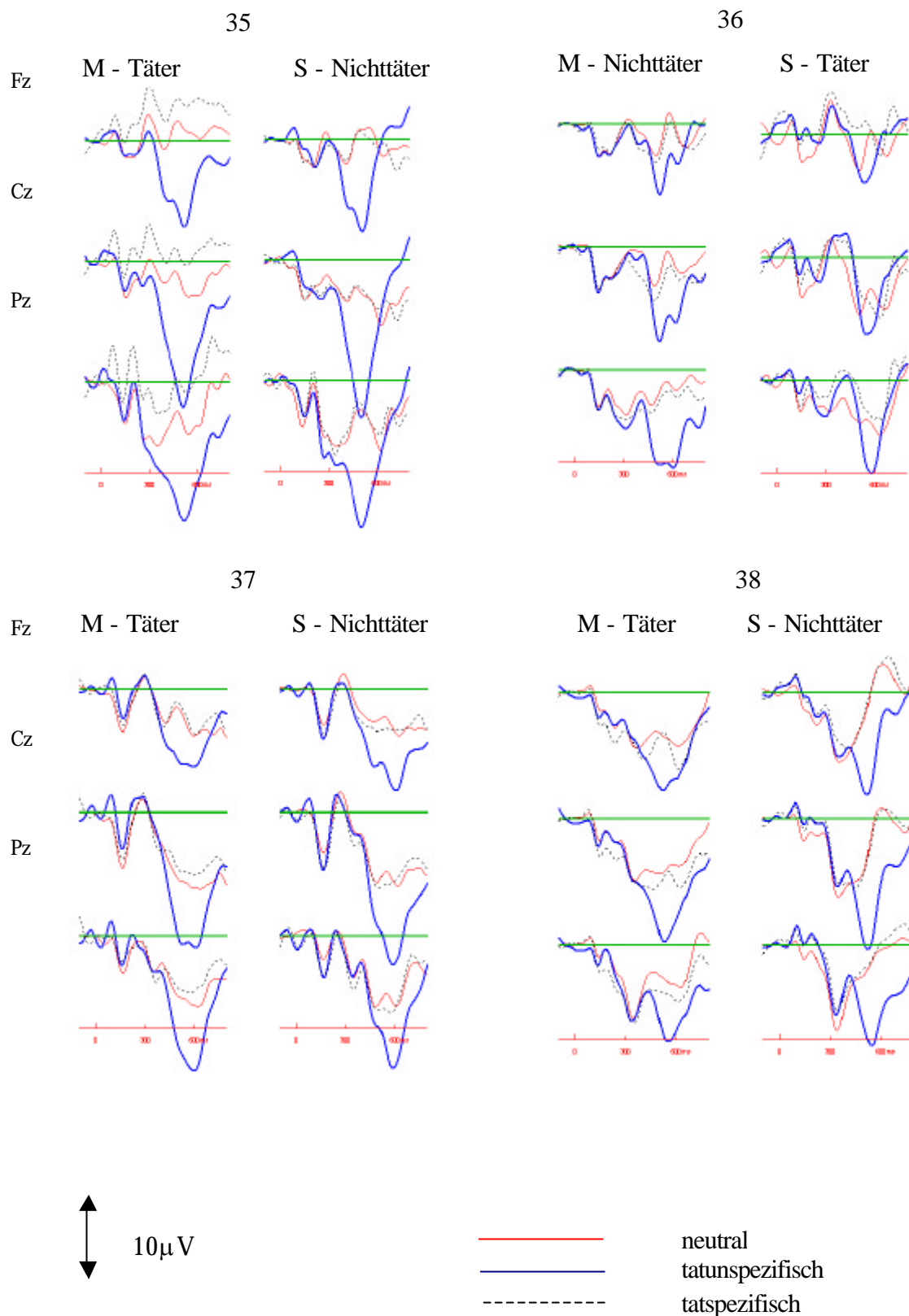


Abbildung 14
 EKP-Kurven von Probanden 35 bis 38 für Täter- und Nichttäterbedingung. Dargestellt sind die Kurven für drei Itemtypen (tatspezifisch, tatunspezifisch, neutral) und drei Ableitorte (Fz, Cz, Pz). (M = Mord, S = Sexualstraftat).

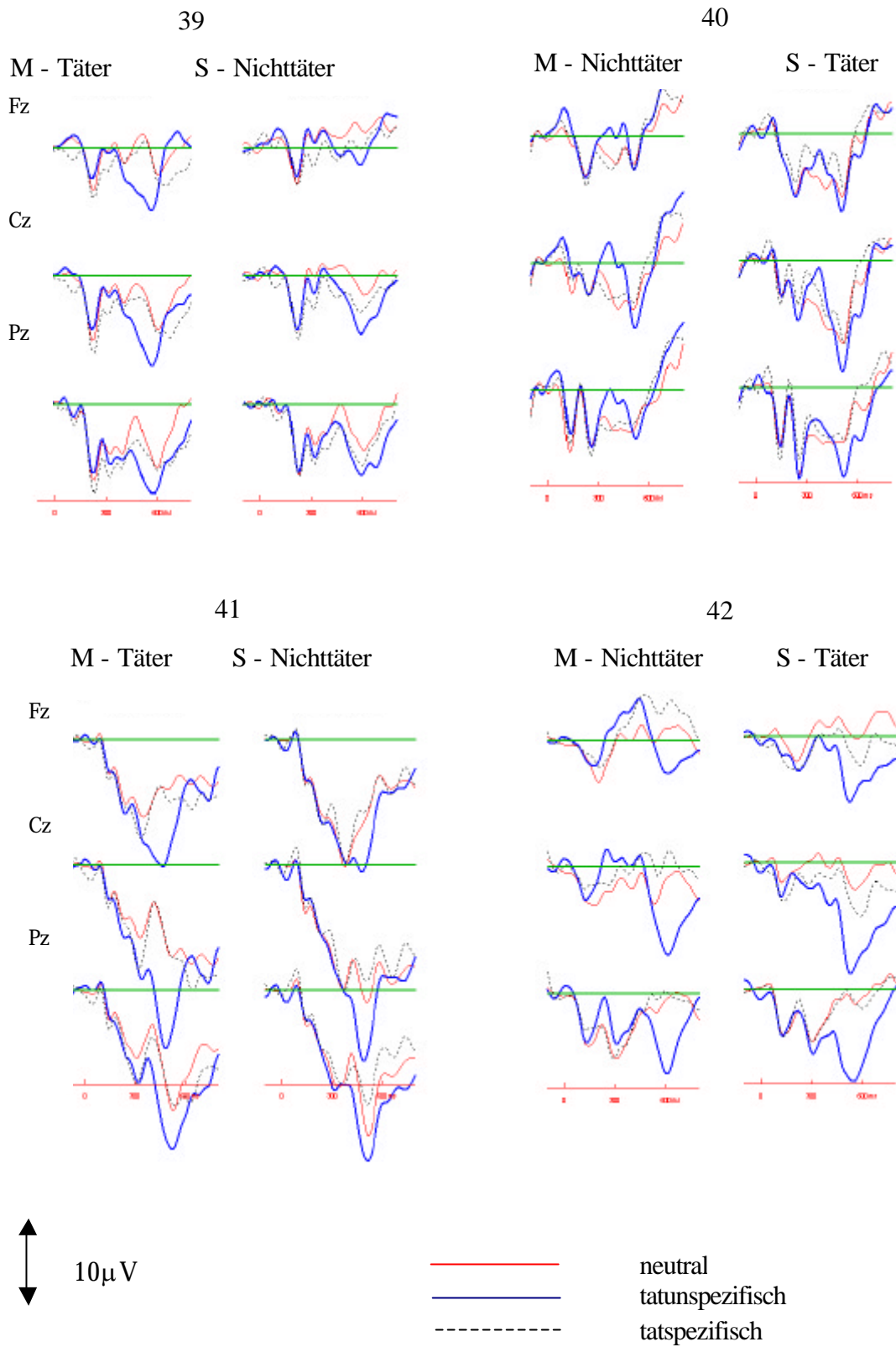


Abbildung 15
 EKP-Kurven von Probanden 39 bis 42 für Täter- und Nichttäterbedingung. Dargestellt sind die Kurven für drei Itemtypen (tatspezifisch, tatunspezifisch, neutral) und drei Ableitorte (Fz, Cz, Pz). (M = Mord, S = Sexualstraftat).

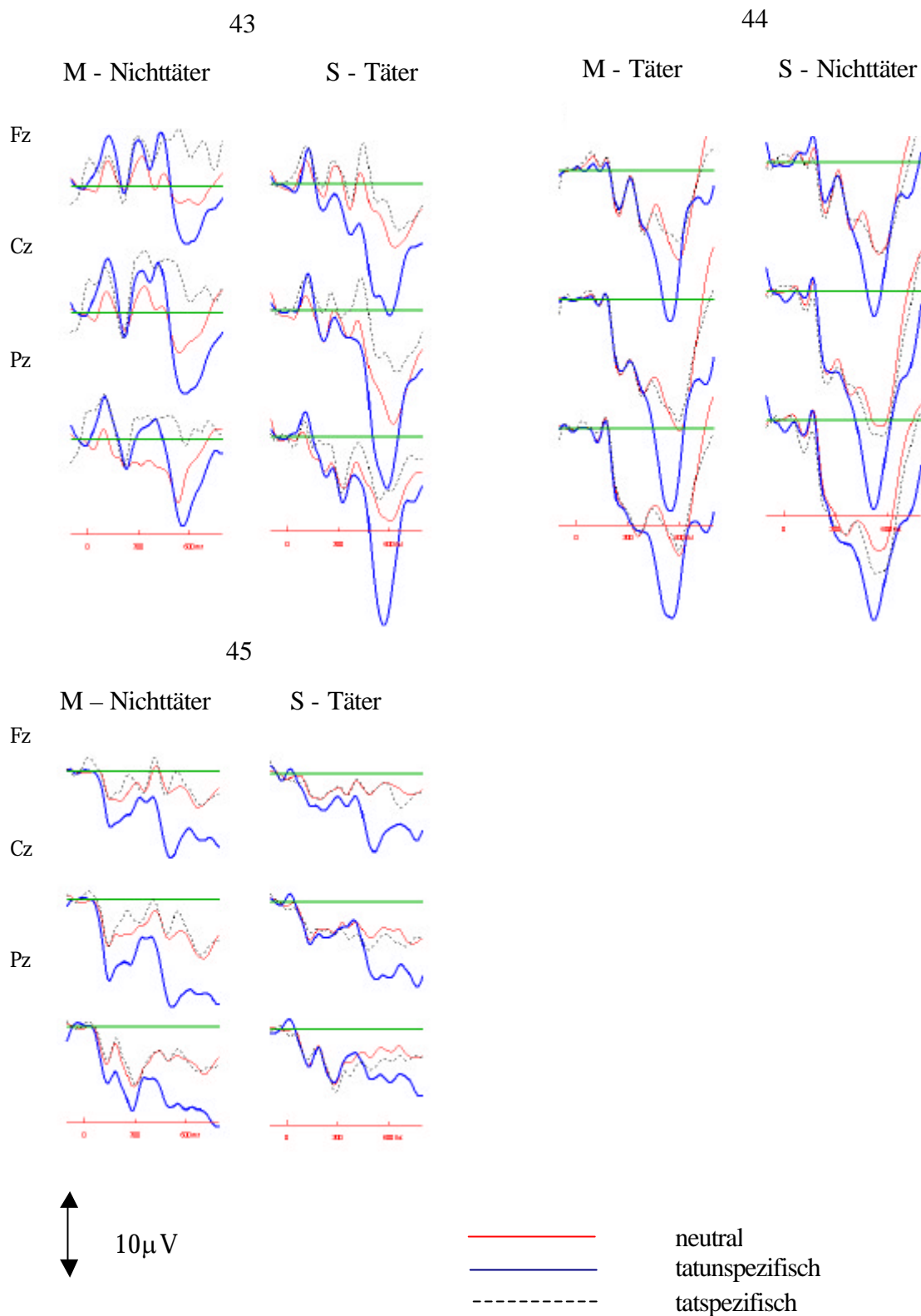


Abbildung 16
 EKP-Kurven von Probanden 43 bis 45 für Täter- und Nichttäterbedingung. Dargestellt sind die Kurven für drei Itemtypen (tatspezifisch, tatunspezifisch, neutral) und drei Ableitorte (Fz, Cz, Pz). (M = Mord, S = Sexualstraftat).

Tabelle 10

Klassifikation anhand der Amplitudenvergleiche bezüglich der dargestellten EKP-Kurven. Die grau unterlegten Felder kennzeichnen korrekte Zuordnungen.

Sexualstraftat	S > N			S ≤ N			Unentscheidbar		
	Fz	Cz	Pz	Fz	Cz	Pz	Fz	Cz	Pz
Täter	10	15	14	8	5	7	5	3	2
Nichttäter	5	11	12	11	9	8	4	0	0
Mord									
Täter	9	9	6	9	10	14	2	1	0
Nichttäter	5	5	7	13	16	14	5	2	2

Mit Hilfe eines Vorzeichentests wurde überprüft, ob es signifikante Unterschiede zwischen Täter- und Nichttäterbedingung hinsichtlich der per Augenschein anhand der EKP-Kurven vorgenommenen Klassifizierungen gab. Der Vorzeichentest ergab für diejenigen Probanden, die Täter bezüglich der Mordthematik beziehungsweise Nichttäter bezüglich der Vergewaltigungsthematik waren, keine signifikanten Unterschiede (Fz: $p = .50$; Cz: $p = .77$; Pz: $p = .15$). Bezüglich der Probanden, die Nichttäter bezüglich der Mordthematik beziehungsweise Täter bezüglich der Vergewaltigungsthematik waren, musste die Annahme, dass beide Stichproben hinsichtlich der Klassifikation einer Grundgesamtheit mit gleicher Verteilung entstammen, zurückgewiesen werden. Hinsichtlich der Ableitung bei Fz ergaben sich keine signifikanten Abweichungen ($p = .30$), hinsichtlich der bei Cz und Pz abgeleiteten Daten waren die Abweichungen signifikant (Cz: $p < .01$; Pz: $p < .10$).

Betrachtet man lediglich die EKP-Kurven derjenigen Probanden, die an der Nachbefragung ein Jahr später teilnahmen, lässt sich Folgendes feststellen:

Der Proband mit der Probandennummer 39, der sich daran erinnerte bei der Vergewaltigungsthematik schlecht und bezüglich der Mordthematik gut abgeschnitten zu haben, obwohl der Feedbackbogen entgegengesetzte Ergebnisse lieferte, lässt sich bezüglich der Vergewaltigungsthematik als Täter und bezüglich der Mordthematik als Nichttäter klassifizieren. Das heißt, die Klassifizierung entspricht den Erinnerungen des Probanden. Proband 6, der sich erinnerte, bei beiden Filmen eine gute Beobachtungsleistung gezeigt zu haben, lässt sich bezüglich der Täterbedingung als Nichttäter und bezüglich der Nichttäterbedingung als Täter klassifizieren.

Zieht man als Kriterium für die Täterschaftsbedingung die Erinnerungen der Probanden bezüglich der Beobachtungsleistungen hinzu und betrachtet lediglich die 8 Probanden, die sich in der Nachbefragung motiviert zeigten, können anhand der bei Pz abgeleiteten EKP-Kurven korrekte Klassifikationen in 13 von 16 (81%) Fällen erfolgen. Alle Falschklassifikationen sind falsch-negative Klassifikationen. Zwei der drei nicht-motivierten Probanden werden hinsichtlich beider Bedingungen als Nichttäter klassifiziert. Proband 21 wird hinsichtlich beider Bedingungen korrekt klassifiziert.

Eine 2 (Täterschaft) x 2 (Deliktthema) x 3 (Itemtyp) faktorielle Varianzanalyse mit Messwiederholung auf dem ersten und dritten Faktor wurde durchgeführt, um festzustellen, ob sich anhand der gemessenen Amplitudenunterschiede Gruppenunterschiede bezüglich der Täterschaftsbedingung ergeben würden. Es ergab sich kein signifikanter Haupteffekt für die Täterschaftsbedingung. Ein Haupteffekt für den Itemtyp ergab sich für alle Ableitorte (alle $p < .001$). Ein Interaktionseffekt trat zwischen der Täterschaftsbedingung und der Deliktthematik für alle Ableitorte auf (Fz: $F(1, 41) = 3.68, p < .10$; Cz: $F(1, 41) = 10.07, p < .05$; Pz: $F(1, 41) = 4.02, p < .10$). Für die Ableitung bei Pz ergab sich eine signifikante Interaktion zwischen Täterschaftsbedingung, Itemtyp und Deliktthematik ($F(2, 40) = 4.21, p < .05$).

Abbildung 17 gibt Grand Averages für alle Probanden der Täterbedingung im Vergleich zur Nichttäterbedingung getrennt für die beiden Deliktthemen wieder. Die dargestellten Kurven spiegeln die Mittelwerte der maximalen P300 Amplituden für die verschiedenen Bedingungen wieder, wie sie in Tabelle 11 dargestellt sind. Der Übersicht halber gibt Tabelle 11 lediglich die Werte für die bei Pz abgeleiteten Werte wieder.

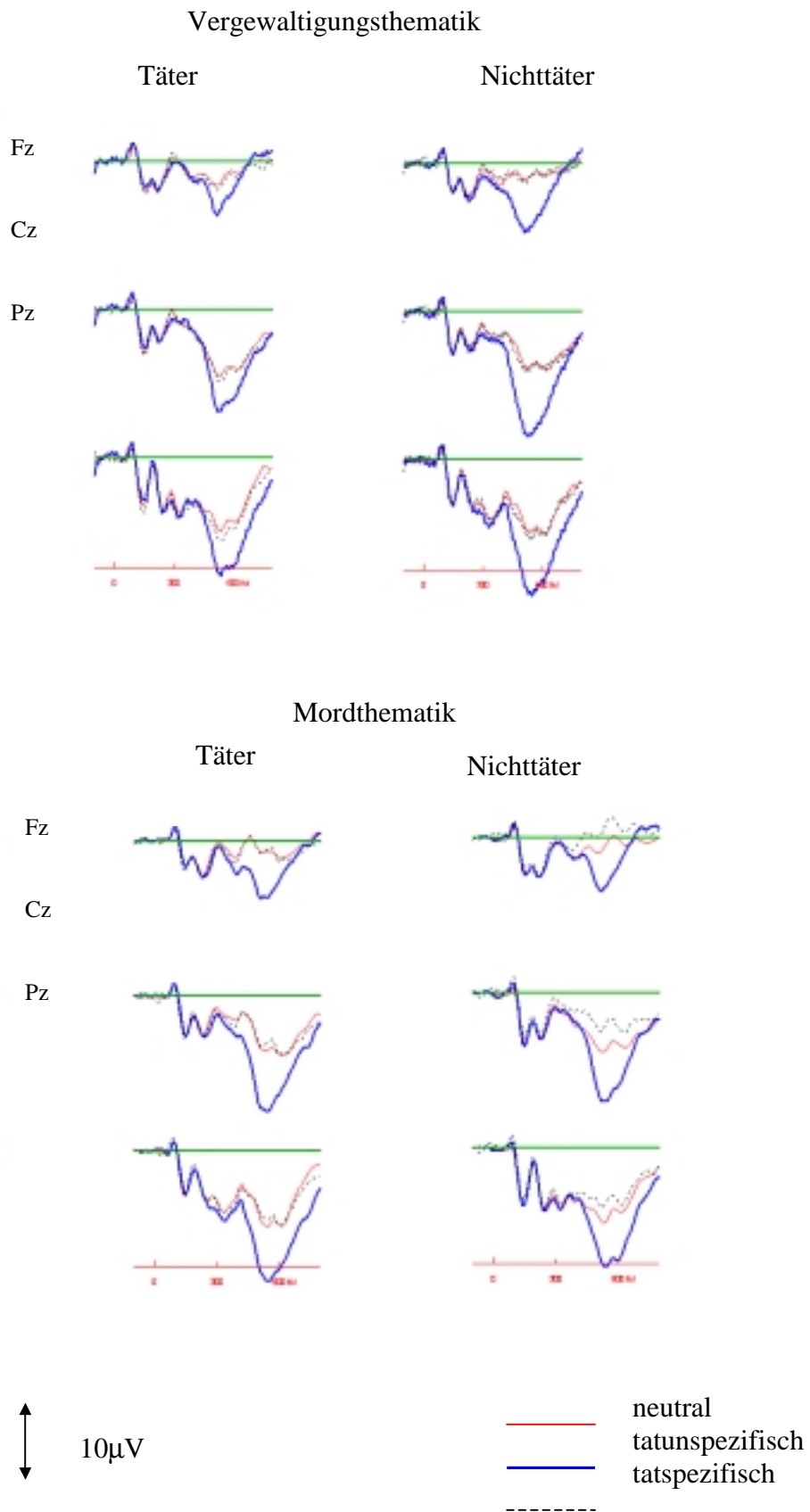


Abbildung 17
 Grand Averages für die Täter- und Nichttäterbedingung getrennt nach Deliktthemen.

Tabelle 11

Mittelwerte und Standardabweichungen (in Klammern) der maximalen Amplituden in μV für neutrale, tatunspezifische und tatspezifische Reize für die Ableitung bei Pz.

	neutrale Reize	tatunspezifisch	tatspezifisch
Gesamt (N=86)	8.81 (5.99)	14.60 (6.39)	8.89 (7.04)
Tötungsdelikt (gesamt)	8.89 (6.38)	14.37 (6.42)	7.72 (6.01)
Täter	9.33 (8.18)	14.47 (7.34)	7.80 (7.17)
Nichttäter	8.39 (3.48)	14.24 (5.36)	7.62 (4.50)
Sexualstraftat (gesamt)	8.72 (5.64)	14.84 (6.43)	10.07 (7.84)
Täter	8.67 (6.98)	14.62 (6.59)	10.98 (9.92)
Nichttäter	8.78 (3.73)	15.08 (6.39)	9.03 (4.45)
Täter (gesamt)	8.54 (5.57)	14.45 (5.98)	9.42 (7.97)
Nichttäter (gesamt)	9.08 (6.43)	14.76 (6.84)	8.37 (6.02)

Hinsichtlich der Einzelfallauswertung sind die Verteilungen der per Korrelationsvergleich gewonnenen Bootstrapindizes (Verfahren 1) für alle Probanden nach Filmthema getrennt in Abbildung 18 und 19 wiedergegeben. Für die Amplitudenvergleiche (Verfahren 2) werden sie getrennt nach Filmthemen in Abbildung 20 und 21 wiedergegeben.

Die Verteilungen der Bootstrapindizes weisen nicht die U-förmige Verteilung auf, die im Falle guter Klassifizierbarkeit zu erwarten gewesen wäre. Bootstrapindizes für Nichttäter hätten klein ausfallen sollen, während diejenigen für Täter hoch sein sollten. Diese Unterscheidung lässt sich auf den ersten Blick nicht vornehmen. Tabelle 12 gibt die AUC-Werte der anhand verschiedener Bootstrapindizes gewonnenen ROC-Kurven mit den zugehörigen Konfidenzintervallen für beide Bootstrapalgorithmen wieder. Alle Werte liegen nahe .50. Ein Wert von 1.00 spricht für optimale Klassifizierbarkeit, ein Wert von .50 ist ein Indikator dafür, dass dieses Verfahren nicht zwischen den Bedingungen differenziert.

Sexualstraftat

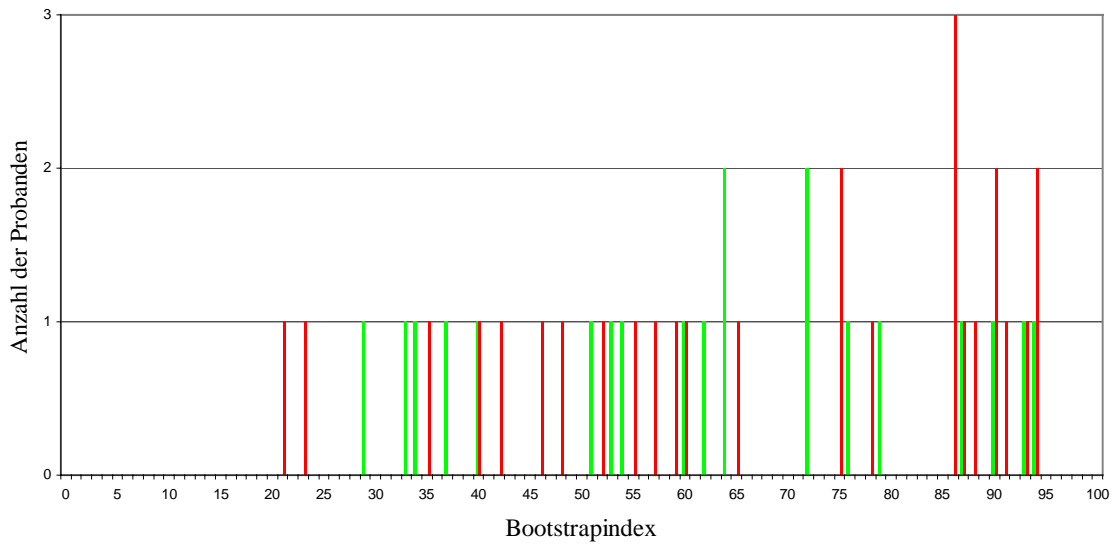


Abbildung 18
 Bootstrapindizes anhand des korrelativen Verfahrens für das Thema „Sexualstraftat“. ■ Nichttäter ■ Täter.

Tötungsdelikt

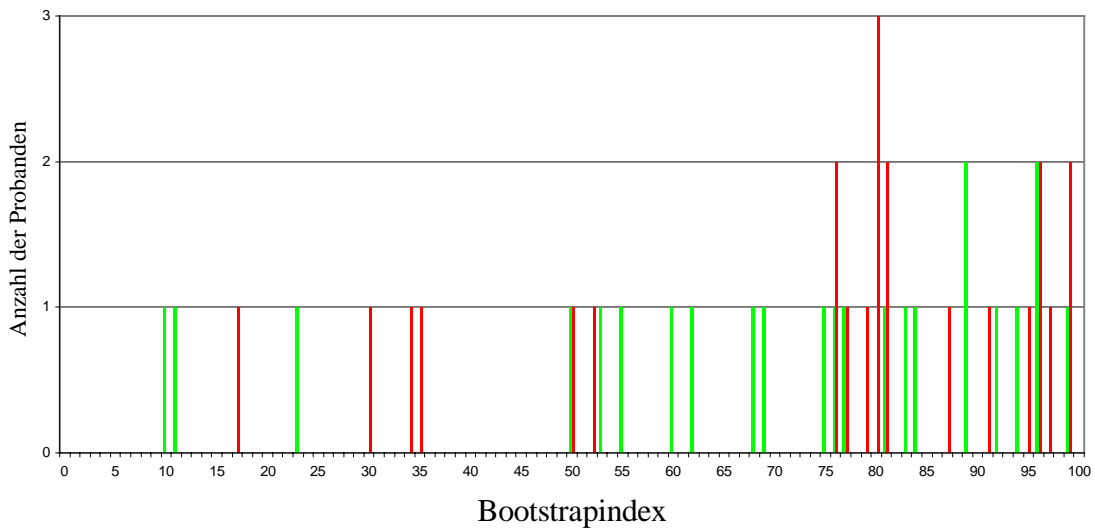


Abbildung 19
 Bootstrapindizes anhand des korrelativen Verfahrens für das Thema „Mord“
 ■ Nichttäter ■ Täter.

Sexualstraftat

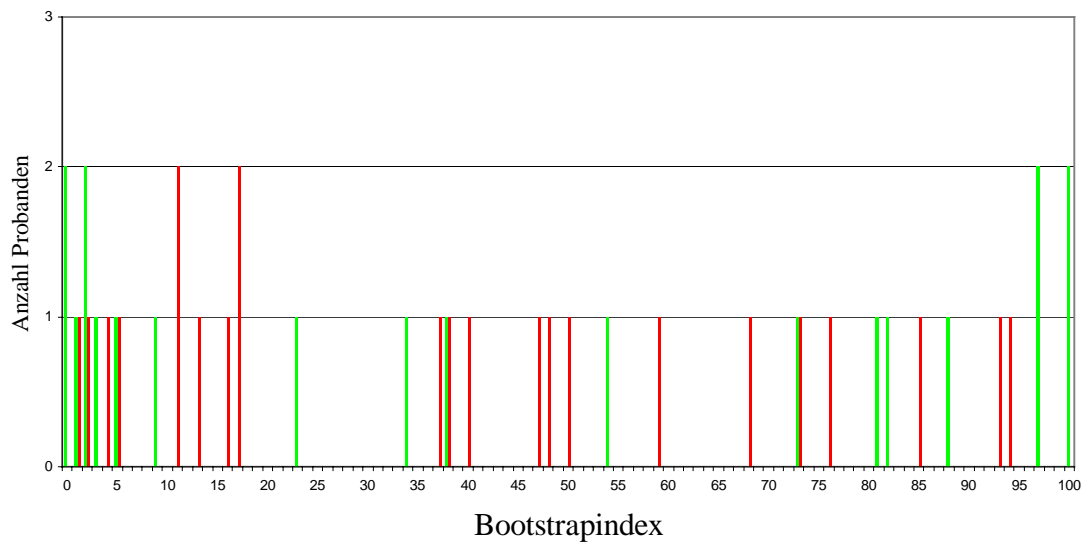


Abbildung 20
 Bootstrapindizes anhand des korrelativen Verfahrens für das Thema „Sexualstraftat“. ■ Nichttäter ■ Täter.

Tötungsdelikt

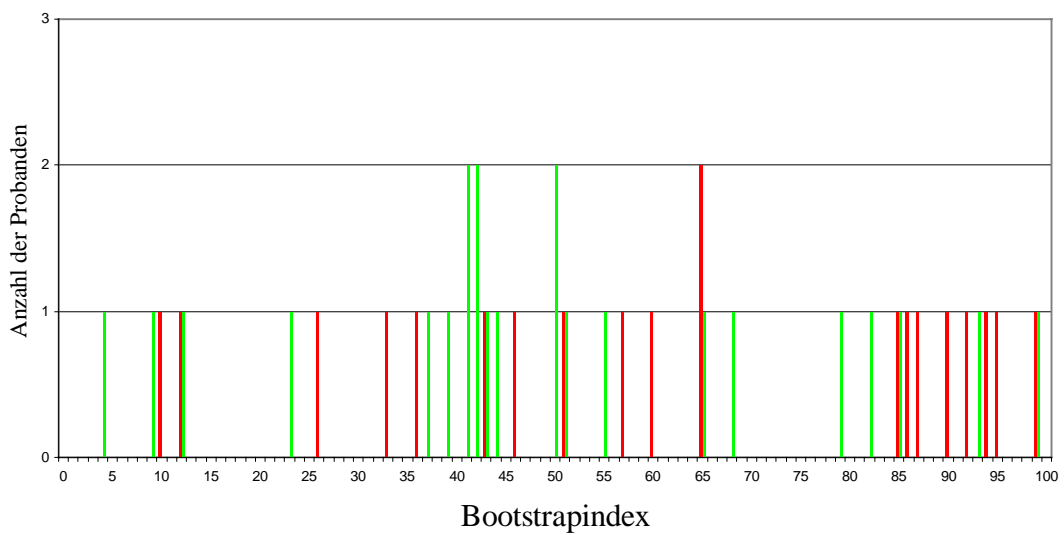


Abbildung 21
 Bootstrapindizes anhand des korrelativen Verfahrens für das Thema „Mord“. ■ Nichttäter ■ Täter.

Tabelle 12

Flächenwerte und Standardfehler der ROC-Kurven für per Amplitudenvergleich und per korrelativem Verfahren gewonnene Bootstrapindizes.

Ableitort	Fläche (AUC)	Standardfehler
Amplitudenvergleich		
Fz	.552	.062
Cz	.537	.063
Pz	.540	.063
korrelatives Vorgehen		
Fz	.565	.063
Cz	.505	.063
Pz	.502	.063

Anhand der Verteilung der Bootstrapindizes (siehe Abbildungen 18 bis 21) lassen sich keine sinnvollen Grenzwerte festsetzen. Zur Veranschaulichung werden im Folgenden die anhand der Ableitung bei Pz gewonnenen Daten exemplarisch klassifiziert. Setzt man für die per Amplitudenvergleich gewonnenen Bootstrapindizes Grenzwerte bei 10 und 70 fest (Farwell & Donchin, 1991), erhält man für die Ableitung bei Pz und das Thema „Sexualstraftat“ 28% korrekte Entscheidungen, 12% falsch negative, 19% falsch positive und 42% unentscheidbare Fälle. Für die gleichen Grenzwerte und den gleichen Ableitort erhält man für das Thema „Mord“ 12% korrekte Entscheidungen, 19% falsch negative, 5% falsch positive und 65% unentscheidbare Fälle (siehe Tabelle 13).

Tabelle 13

Klassifizierung der Probanden anhand der per Amplitudenvergleich gewonnenen Bootstrapindizes (B). Grenzwerte bei 10 und 70. Grau unterlegte Felder beinhalten hypothesenkonforme Fälle.

	$(S \sim U) > N$ bzw. $B < 10$	$(S \sim N) \neq U$ bzw. $B > 70$	Unentscheidbar
Sexualstraftat			
Zu Recht	4	5	14
Zu Unrecht	8	8	4
Mord			
Zu Recht	0	8	12
Zu Unrecht	2	5	16

Setzt man die gleichen Grenzwerte für die per korrelativem Verfahren gewonnenen Bootstrapindizes fest, erhält man für die Ableitung bei Pz und das Thema „Sexualstraftat“ 19% korrekte Entscheidungen, die sich allerdings auf die korrekte Identifizierung von Nichttätern begrenzen. Man erhält 23% falsch negative und 58% unentscheidbare Fälle. Kein Proband wurde als Täter klassifiziert. Für die gleichen Grenzwerte und den gleichen Ableitort erhält man für das Thema „Mord“ 30% korrekte Entscheidungen, die sich wiederum auf die Identifizierung von Nichttätern beschränkt. Man erhält 33% falsch negative und 37% unentscheidbare Fälle (Tabelle 14).

Tabelle 14

Klassifizierung der Probanden anhand der per korrelativem Verfahren gewonnenen Bootstrapindizes (B). Grenzwerte bei 10 und 70. Grau unterlegte Felder beinhalten hypothesenkonforme Fälle.

	$(S \sim U) > N$ bzw. $B < 10$	$(S \sim N) \neq U$ bzw. $B > 70$	Unentscheidbar
Sexualstraftat			
Zu Recht	0	10	13
Zu Unrecht	0	8	12
Mord			
Zu Recht	0	14	6
Zu Unrecht	0	13	10

Tabelle 15 gibt exemplarisch die Verteilung der Klassifikationen pro Person wieder. Die Klassifikation wurde anhand der oben genannten Grenzwerte für per Amplitudenvergleich gewonnene Bootstrapindizes vorgenommen. Dargestellt sind exemplarisch die Ergebnisse für Pz. Es zeigt sich, dass kein Proband hinsichtlich beider Filmthemen für einen Täter gehalten wird. Drei Probanden werden als Nichttäter bezüglich „Mord“ und als Täter bezüglich der „Sexualstraftat“ beurteilt. Ein Proband wird bezüglich „Mord“ als Täter und bezüglich „Sexualstraftat“ als Nichttäter beurteilt. Fünf Probanden werden in beiden Fällen als „Nichttäter“ klassifiziert. Die restlichen Probanden werden zumindest hinsichtlich eines Themas als unklassifizierbar beurteilt. Die Darstellung für die mit Hilfe des korrelativen Verfahrens gewonnenen Daten erübrigt sich, da anhand dieser Methode kein Proband als Täter klassifiziert wurde.

Tabelle 15

Verteilung der Probanden über die verschiedenen Klassifikationsbedingungen. (0 = unentscheidbar; 1 = Nichttäter; 2 = Täter).

		Klassifikation für Mord			
		0	1	2	Ges.
Klassifikation für Sexualstraftat	0	12	5	1	18
	1	7	5	1	13
	2	9	3	0	12
	Ges.	28	13	2	43

7.3 Annahme 5: Deliktthemenspezifische Unterschiede

Weiterhin wurde die Annahme geprüft, dass es keine themenspezifischen Unterschiede bezüglich der P300 und der Klassifizierbarkeit der Probanden gibt. Eine 2 (Täterschaft) x 2 (Deliktthema) x 3 (Itemtyp) faktorielle Varianzanalyse mit Messwiederholung auf dem zweiten und dritten Faktor ergab einen signifikanten Haupteffekt für die Deliktthematik (Fz: $F(1, 41) = 4.02, p < .10$; Cz: $F(1, 41) = 10.07, p < .05$; Pz: $F(1, 41) = 3.17, p < .10$). Weiterhin zeigte sich, dass die maximale Amplitude der P300 signifikant vom Itemtyp beeinflusst wurde (Fz: $F(2, 40) = 25.23, p < .001$; Cz: $F(2, 40) = 41.99, p < .001$; Pz: $F(2, 40) = 54.13, p < .001$). Ein signifikanter Interaktionseffekt zeigte sich für Itemtyp und Filmbedingung für die Ableitung bei Pz ($F(2, 40) = 4.22, p$

< .05). Es ergaben sich keine weiteren signifikanten Haupt- oder Interaktionseffekte. Tabelle 11 und Abbildung 17 spiegeln diese Ergebnisse wieder.

7.4 Ergebnisse bezüglich der Kontrollvariablen

Daten des Rekognitionstests bezüglich des Wiedererkennens der tatspezifischen und tatunspezifischen Items lagen bei 35 der 43 Probanden vor. Der Mittelwert der richtig wiedererkannten tatspezifischen Reize betrug 5.22 mit einer Standardabweichung von 1.00 für das Thema „Sexualstraftat“ und 4.97 mit einer Standardabweichung von 1.09 für das Thema „Tötungsdelikt“. Der Mittelwert der richtig wiedererkannten tatunspezifischen Items betrug 5.49 mit einer Standardabweichung von .95 für das Thema „Sexualstraftat“ und 5.26 mit einer Standardabweichung von 1.04 für das Thema „Tötungsdelikt“. Die Verteilung der Wiedererkennungsraten gibt Tabelle 16 wieder. Die nachfolgend angegebenen Korrelationen beziehen sich auf die anhand der Ableitung bei Pz gewonnenen Daten. Die Anzahl der korrekt erinnerten tatspezifischen Items im Rekognitionstest korrelierte zu $r_{\text{Pearson}} = .17$ mit den maximalen P300-Amplituden der tatspezifischen Reize. Die Anzahl der korrekt erinnerten tatunspezifischen Reize korrelierte zu $r_{\text{Pearson}} = -.12$ mit den maximalen P300-Amplituden der zugehörigen Reize. Die Anzahl der fälschlich für relevant gehaltenen neutralen Items korrelierte zu $r_{\text{Pearson}} = .05$ mit den maximalen P300-Amplituden der neutralen Reize. Die bei Fz und Cz ermittelten maximalen P300-Amplituden ergaben ebenfalls keine signifikanten Korrelationen.

Tabelle 16

Verteilung der Ergebnisse im Rekognitionstest über alle Probanden für die verschiedenen Bedingungen.

Anzahl erkannter tatspezifischer Items	Vergewaltigungsthematik		Mordthematik	
	Täter	Nichttäter	Täter	Nichttäter
2	0	0	0	1
3	1	2	1	3
4	2	3	0	3
5	4	4	5	9
6	13	6	9	4

Von 86 Feedbackbögen wurden 84 Bögen mit den erwarteten Ergebnissen bearbeitet, d. h. sie führten zu dem Ergebnis „schlechte Beobachtungsleistung“, wenn die Schwierigkeit des Bogens hoch gewählt war, und umgekehrt zu dem Ergebnis „gute Leistung“, wenn die Schwierigkeit niedrig gewählt war. Ein Proband erreichte in der „Tötungsdelikt/Täter“-Bedingung ein positives Ergebnis und in der „Sexualstraftat/Nichttäter“-Bedingung ein negatives Ergebnis. Dieser Proband wurde wie ein Proband der Bedingung „Tötungsdelikt/Nichttäter“ und „Sexualstraftat/Täter“ behandelt. Der Mittelwert der Feedbackbogenergebnisse für die Bedingung „Tötungsdelikt/Täter“ betrug -3.15 bei einer Standardabweichung von 2.35 . Der Mittelwert der Feedbackbogenergebnisse für die Bedingung „Tötungsdelikt/Nichttäter“ betrug 1.78 bei einer Standardabweichung von 1.77 . Für die Bedingung „Sexualstraftat/Täter“ betrug der Mittelwert -2.04 und die Standardabweichung von 2.08 , für die Bedingung „Sexualstraftat/Nichttäter“ betrug der Mittelwert 1.45 bei einer Standardabweichung von 2.63 . Die Werte der Feedbackbögen korrelierten zu $r_{\text{Pearson}} = -.057$ mit den bei Pz gewonnenen maximalen P300-Amplituden der bezüglich der tatspezifischen Reize abgeleiteten EKP-Kurven. Die bei Fz und Cz abgeleiteten Werte ergaben gleichfalls keine signifikanten Korrelationen.

Um zu prüfen, ob sich eine möglicherweise bestehende unterschiedliche, berufsbedingte Motivation auf die Amplitude der P300 auswirkte, wurde eine zweifaktorielle Varianzanalyse (Itemtyp, Berufsgruppe) mit Messwiederholung auf dem ersten Faktor berechnet. Die Mittelwerte und Standardabweichungen gibt Tabelle 17 wieder. Die Mittelwerte sind in der motivierten Gruppe größer als in der von Berufs wegen nicht motivierten Gruppe. Die Varianzanalyse ergab bei Verwendung der maximalen P300-Amplituden als abhängiger Variable keine signifikanten Mittelwertsunterschiede für die beiden Berufsgruppen.

Tabelle 17

Mittelwerte und Standardabweichungen in μV für Itemtyp und Berufsgruppe.

Berufsgruppe		neutral	tatunspezifisch	tatspezifisch
Nicht motiviert	Mittelwert	7.78	13.26	8.00
	N	50	50	50
	Standardabweichung	3.48	5.61	5.21
Motiviert	Mittelwert	10.23	16.46	10.12
	N	36	36	36
	Standardabweichung	8.15	6.98	8.92
Insgesamt	Mittelwert	8.80	14.60	8.89
	N	86	86	86
	Standardabweichung	5.98	6.38	7.04

8. Diskussion

8.1 Beantwortung der Fragestellungen

Die erste Annahme war, dass tatunspezifische Reize eine ausgeprägte P300 hervorrufen würden. Diese Annahme konnte bestätigt werden. Die Probanden erfüllten instruktionsgemäß die Aufgabenstellung, nämlich die zuvor auswendig gelernten tatunspezifischen Items aus den neutralen Items herauszufiltern. Entsprechend zeigte der Amplitudenvergleich zwischen neutralen und tatunspezifischen Items, dass die durch tatunspezifische Items hervorgerufene P300-Amplitude größer war. Damit wurde eine erste Voraussetzung für die Klassifizierbarkeit der Probanden geschaffen. Durch das Auftreten des Amplitudenunterschiedes war gewährleistet, dass die Probanden den Items ihre Aufmerksamkeit schenken. Somit hatte das Drei-Kategorien-Oddball-Paradigma diesen Zweck erfüllt.

Gemäß der Annahmen 3) und 4) sollten sich Probanden der Täter- und der Nichttäterbedingung hinsichtlich der P300-Komponente folgendermaßen unterscheiden: Bezüglich der P300-Amplitude sollte sich für Täter zeigen, dass $S > N$ gilt (Annahme 3)b)). Für Nichttäter sollte gemäß Annahme 4)b) $S \leq N$ gelten. Die gruppenstatistische Auswertung zeigt, dass die Unterschiede bei denjenigen Probanden, die Täter bezüglich der Vergewaltigungsthematik und Nichttäter bezüglich der Mordthematik waren, in diese Richtung weisen. Bei den übrigen Probanden lässt sich dieser Gruppenunterschied nicht feststellen. Anhand der Einzelfallauswertung können einige Probanden korrekt ihren Bedingungen zugeordnet werden. Insgesamt zeigt jedoch die Einzelauswertung sowohl mit Hilfe des korrelativen Verfahrens (Annahmen 3)a) und 4)a)), als auch mit Hilfe der Amplitudenvergleiche (Annahmen 3)b) und 4)b)), dass die Klassifizierung der Probanden nicht besser als zufällig erfolgt. Die Annahmen 3) und 4) können demnach anhand dieser Daten nicht belegt werden. Bezüglich der Annahme 4) ist festzustellen, dass falsch-positive Klassifikationen nicht überwiegen. Die hohe Fehlerrate lässt sich also offenbar nicht dadurch erklären, dass Nichttäter über tatspezifische Informationen verfügten und aufgrund dieses Wissens als Täter klassifiziert wurden.

Annahme 5) kann aufgrund der gruppenstatistischen Auswertung im Rahmen dieser Arbeit nicht belegt werden. Es zeigten sich Gruppenunterschiede zwischen den Deliktthemen, die darauf zurückzuführen waren, dass die P300-Amplituden bezüglich der Sexualstraftatsthematik jeweils größer waren. Signifikant war dieser Unterschied

dann, wenn die Variation der Thematik als Messwiederholungsvariable betrachtet wurde. Im Mittel maßen die Probanden der Sexualstraftatsthematik also größere Bedeutung zu. Bezüglich der Deliktthematik ließ sich außerdem, wie oben bereits erwähnt, ein Einfluss auf die Täterschaftsbedingung feststellen. Täter zeigten dann, wenn sie als Täter bezüglich der Vergewaltigungsthematik galten, im Vergleich zur Nichttäterbedingung größere positive Differenzen zwischen tatspezifischen und neutralen Items als Täter, die als Täter bezüglich der Mordthematik galten. Offensichtlich hatte es eine andere Qualität anzunehmen, dass man bezüglich der Vergewaltigungsthematik eine schlechte Beobachtungsleistung gezeigt hatte und bezüglich der Mordthematik eine gute, als umgekehrt.

Hinsichtlich der Kontrollvariablen zeigte sich kein Einfluss der Rekognitionsleistung. Die Probanden zeigten trotz des relativ langen Zwischenintervalls von einer Woche überwiegend eine gute Wiedererkennungslleistung bezüglich tatspezifischer und tatunspezifischer Items. Dies spricht dafür, dass es nicht zwingend notwendig ist, tatspezifische Informationen in einer Simulationsstudie durch Auswendiglernen einzuführen. Die Ergebnisse der Feedbackbögen zeigten ebenfalls keine Zusammenhänge mit P300-Amplituden. Zwar lagen die P300-Amplitudenmittelwerte der Probanden, die auf Grund ihres Berufs zur Gruppe der Motivierteren zugeteilt worden waren, tatsächlich höher als die der übrigen Probanden, der Unterschied war jedoch nicht signifikant. Bei der Betrachtung der Ergebnisse bezüglich der Kontrollvariablen ist zu beachten, dass hier P300-Amplituden interindividuell verglichen werden. Aufgrund der hohen interindividuellen Varianz der EKP-Amplituden sind diese Vergleiche nur bedingt aussagekräftig.

Die Überprüfung der internen Validität des Untersuchungsvorgehens erbrachte folgende Anhaltspunkte:

Eine Voruntersuchung bekräftigte die Annahme, dass das Bonner Szenario geeignet sei, eine andere motivationale und emotionale Täuschungssituation zu erzeugen als das Mock Crime-Paradigma. Probanden des Bonner Szenarios reagierten ernsthafter als Probanden des Mock Crimes. Dies kann als ein Beleg gelten, dass das Vorgehen so gestaltet war, dass Probanden die Situation ernst nahmen und somit Realitätsnähe geschaffen wurde, die natürliches und spontanes Verhalten im Gegensatz zum Mock Crime ermöglicht.

Ferner ergaben sich in der Nachbefragung keine Hinweise darauf, dass die Probanden gegenüber dem vorgeblichen Untersuchungshintergrund misstrauisch geworden wären. Einige Probanden kritisierten die Feedbackbögen während der Untersuchung. In der Stichprobe, die ein Jahr nach der Untersuchung durchgeführt wurde, zeigt sich, dass die Wahrnehmung der Beobachtungsleistung bezüglich der Filme bei den meisten Probanden von den Ergebnissen des Feedbackbogens abhing. Offenbar wirkt sich eine Wahrnehmung der Beobachtungsleistung, die von der durch die Feedbackbögen zurückgemeldeten Leistung abweicht, auf die EKP-Kurven aus. Dies konnte anhand eines Probanden demonstriert werden. Aus den Kommentaren bezüglich der erinnerten Empfindungen, die auf die Beschuldigung folgten, und der Tatsache, dass die Probanden sich auch nach einem Jahr noch an die Abläufe erinnerten, lässt sich schließen, dass ein Großteil der Probanden sich mit dem Vorwurf auseinandersetzt. Dass eine kleine Zahl von Probanden von den Vorhaltungen wahrscheinlich nicht beeindruckt war, lässt sich aus dieser Nachbefragung jedoch ebenso entnehmen. Der Proband, der sich nicht erinnerte, zeigte auch keine EKP-Unterschiede zwischen den beiden Täterschaftsbedingungen. Schließt man diejenigen Probanden aus, die sich durch den Vorhalt nicht beeindruckt zeigen und wählt als Kriterium für die Täterschaftsbedingungen die von den Probanden erinnerten Leistungen, fällt die Probandenklassifikation besser aus, als bei Betrachtung der Gesamtstichprobe.

8.2 Vergleich mit anderen Studien im Bereich der Täuschungsaufdeckung mit Hilfe der P300

Das Prinzip der Täuschungsaufdeckung mit Hilfe der P300 wurde im Rahmen von Simulationsstudien im Hinblick auf verschiedene *Anwendungsbereiche* erprobt.

Einen Anwendungsbereich stellt die Überprüfung von *Aggravation und Simulation neurologischer Symptome* dar. Anhand der P300-Komponente soll zwischen solchen Patienten unterschieden werden, die tatsächlich nicht in der Lage sind Informationen aufgrund amnestischer Symptome zu erinnern, und solchen, die zwar über die entsprechende Erinnerung verfügen, diese aber nicht zugeben. Van Hooff, Brunia und Allen (1996) erwähnen die Möglichkeit, den Anwendungsbereich auch auf die Überprüfung des impliziten Gedächtnisses auszudehnen. Belege dafür, dass die P300 als Indikator für die Rekognition von Informationen dienen kann, liefern

verschiedene Untersuchungen. Exemplarisch wurden bereits die Untersuchungen von Allen, Iacono und Danielson (1992) bzw. von Allen und Iacono (1997) vorgestellt (siehe Kapitel 3.2.3.2). Hier wurden nacheinander zwei Itemlisten von den Probanden, die Patienten mit Täuschungstendenzen simulierten, auswendig gelernt. Die Probanden verheimlichen per Instruktion die Kenntnis der zuerst gelernten Liste. Die Rekognition der zuerst gelernten Liste soll anschließend durch die Registrierung der P300-Komponente überprüft werden. Indiziert die P300 durch einen Amplitudenunterschied zwischen früher gelernten Items und neutralen Items, dass diese wiedererkannt wurden, obwohl der Proband das Gegenteil behauptet, wird damit angezeigt, dass der Proband täuscht. Eine Reihe weiterer Untersuchungen zeigt, dass die P300-Komponente als Index für ein intaktes Rekognitionsgedächtnis fungieren kann (Ellwanger, Rosenfeld, Hannkin & Sweet, 1999; Ellwanger, Rosenfeld, Sweet & Bhatt, 1996; Van Hooff, Brunia & Allen, 1996).

Die Verwendung des Oddball-Paradigmas als Screening-Instrument im Rahmen von Integritätsfragen stellt eine weitere Anwendungsmöglichkeit dar. Eine Reihe unerwünschter Aktivitäten wird als Items verwendet. Die Annahme ist, dass eine begangene Tat als Item eine verstärkte P300 hervorruft. Im Rahmen von Simulationsstudien ließen sich mit Hilfe dieser Methode die tatsächlich begangenen sozial unerwünschten Handlungen identifizieren (Johnson & Rosenfeld, 1992; Rosenfeld, Angell, Johnson & Quian, 1991).

Mit Hilfe typischer Mock Crime Studien untersuchte man die Verwendung der P300 als Täuschungsindikator im Hinblick auf die Beschuldigung hinsichtlich einer Straftat. Als inkriminierte Taten wurden Einbruch und Diebstahl (Zhou, Yang, Liao & Zou, 1999; Zhou, Yang, Liao & Zou, 2000) sowie Spionagehandlungen simuliert (Farwell & Donchin, 1991). Die Klassifikation der Probanden anhand der P300 in Täter und Nichttäter erzielte hohe Trefferquoten.

Die vorliegende Untersuchung wurde im Hinblick auf die Verwendung des Drei-Kategorien-Oddball-Paradigmas im Rahmen einer PAG, die bei inkriminierten Sexualstraftaten durchgeführt werden könnte, gestaltet. Dieses Anwendungsgebiet stellt andere Anforderungen an die Methode zur Täuschungsaufdeckung als ein neurologischer Test, ein Screeningkontext oder auch die Anwendung auf Einbruchdelikte oder Spionage. Aus diesem Grund ergaben sich Unterschiede hinsichtlich der

Gestaltung der Untersuchung zu bisherigen Arbeiten. Im Folgenden werden einige dieser Unterschiede beschrieben. Da die vorliegende Untersuchung niedrige Trefferquoten erzielte, wird diskutiert, ob die Unterschiede in der Untersuchungsgestaltung zu den diskrepanten Ergebnissen beigetragen haben könnten.

Die genannten Untersuchungen unterscheiden sich hinsichtlich der *Zeitspanne*, die zwischen dem Erwerb der Informationen über tatspezifische Items und der EEG-Ableitung liegt (siehe Tabelle 3). Soll das Oddball-Paradigma herangezogen werden, um das Kurzzeitgedächtnis von Patienten zu überprüfen, sind kurze Zeitintervalle von Minuten bis zu wenigen Stunden gerechtfertigt. Im Falle einer Begutachtung im Strafprozess liegen zwischen dem inkriminierten Geschehen und der Durchführung der PAG dagegen Monate oder sogar Jahre. Die Gefahr, dass tatspezifische Informationen vergessen werden, ist hier eher gegeben.

Van Hooff, Brunia und Allen (1996) konnten zeigen, dass bereits ein Zeitabstand von einem Tag zwischen dem Erlernen einer Itemliste und der Ableitung des EEGs dazu führte, dass die P300-Unterschiede zwischen neutralen und gelerntem Material sich verminderten. Zwar waren die Unterschiede noch vorhanden, sie waren aber weniger stark ausgeprägt, als wenn die EEG-Ableitung innerhalb der nächsten Stunde nach Erlernen der Liste erfolgte. In dieser Untersuchung konnten lediglich 3 von 20 Probanden alle 18 gelernten Items der Itemlisten nach der letzten EEG-Ableitung niederschreiben. Weitere numerische Angaben über die Anzahl der Fehler werden nicht gemacht. Auch in den übrigen vorgestellten Untersuchungen werden keine Angaben zu der Überprüfung der erinnerten Informationen gemacht. Damit ist ein Vergleich mit den eigenen Daten nicht möglich.

Der Zeitabstand zwischen Informationserwerb und EEG-Ableitung betrug in der vorliegenden Untersuchung ca. eine Woche. Die Rekognitionsleistungen der Probanden waren trotz des langen Zwischenintervalls überwiegend gut. Zwar zeigte sich kein Zusammenhang zwischen der P300-Amplitude und der Anzahl der korrekt wiederkannten tatspezifischen Items, es ist aber denkbar, dass jedes nicht wiedererkannte tatspezifische Item während der EEG-Ableitung die Wahrscheinlichkeit einer Klassifikation als Täter vermindert.

Ein weiterer Unterschied zwischen der vorliegenden Untersuchung und den bisher genannten Arbeiten auf dem Gebiet der Täuschungsaufdeckung mit Hilfe der

P300 besteht im *Informationserwerb*. In einem Großteil der dargestellten Studien wird tatspezifisches Wissen durch Auswendiglernen erworben. In der vorliegenden Untersuchung erhielten die Probanden dieses Wissen zwecks größerer Praxisnähe beiläufig während des Betrachtens der Filme und während des Ausfüllens der Feedbackbögen. Auch dies kann dazu beigetragen haben, dass tatspezifische Informationen nicht registriert wurden. Wie bereits erwähnt, liegen jedoch keine brauchbaren Vergleichsdaten aus anderen Studien zur P300 vor.

Ein Schwachpunkt der EKP-Studien, die mit verbalen Reizen arbeiten, besteht darin, dass es kaum Erkenntnisse über die *Anforderung an die Itemformulierung* gibt. Farwell und Donchin (1991) verwendeten sehr homogenes Reizmaterial (siehe Tabelle 2). Das Problem der Itemformulierung stellt sich im Bereich des Einsatzes der P300 zur Aufdeckung von Simulation und Aggravation bei Patienten der Neurologie in sehr abgeschwächter Weise. Soll das Kurzzeitgedächtnis getestet werden, kann mit standardisiertem Material in Form eines Tests gearbeitet werden. Die in Frage stehenden Items werden dann auch bei der praktischen Anwendung des Oddball-Paradigmas auswendig gelernt und ihre Funktion wird dem Patienten per Instruktion verdeutlicht. Da es nicht erforderlich ist, dass thematisch ein bestimmter Deliktbezug deutlich wird, kann ein homogenes Itemmaterial generiert werden, das unabhängig von der Testung keine besondere Bedeutsamkeit aufweist (siehe z. B. Allen & Iacono, 1997). Bei allen genannten Untersuchungen wurden gleiche Wort- und Silbenzahl hinsichtlich relevanter und Vergleichsitems verwendet. Die Zugehörigkeit zu einer semantischen Kategorie innerhalb einer Itemklasse, zu der ein tatspezifisches, ein tatunspezifisches und mehrere neutrale Items gehören, ist ein weiteres Kriterium, das berücksichtigt wird. Das Material der vorliegenden Untersuchung fiel inhomogener aus als in bisherigen Studien, obwohl die zuletzt genannten Kriterien berücksichtigt wurden. In der vorliegenden Untersuchung unterschieden sich z. B. die tatspezifischen Items untereinander stärker hinsichtlich Silben- oder Wortzahl als bei den zuvor genannten Untersuchungen. Dies war notwendig, da solche Items ausgewählt werden mussten, von denen man annehmen konnte, dass diese vom Täter zuverlässig als tatspezifische Items erkannt und erinnert werden konnten. Entsprechend fungierten in der vorliegenden Untersuchung solche Items als tatspezifisch, die unmittelbar mit der inkriminierten Tat, dem lustvollen Betrachten von Gewalttaten, in Zusammenhang standen. Damit bestan-

den in der vorliegenden Untersuchung sehr viel weniger Freiheitsgrade bei der Auswahl und Formulierung von Items. Es ist bisher ungeklärt, inwieweit die Verletzung der Prinzipien, wie z. B. gleiche Wortlänge, Bekanntheitsgrad bzw. Gebrauchshäufigkeit oder Anzahl der Wörter bzw. Silben, sich auf die P300-Komponente auswirken.

Die Items der vorliegenden Untersuchung weisen dadurch, dass sie in direktem Tatzusammenhang stehen, gegenüber denen anderer Untersuchungen einen weiteren Unterschied auf. So lokalisiert beispielweise das Item „über Zaun hängen“ den Ort in der Filmszene, in der der nackte Busen der Geschädigten zum ersten Mal zu sehen ist. Das Item „wollene Decke“ erhält seinen direkten Tatbezug dadurch, dass diese um den Kopf der geschädigten Person gewickelt ist, während diese ermordet wird und das Blut durch die Decke sickert. Die in der vorliegenden Untersuchung verwendeten tatspezifischen Items könnten in der Hinsicht problematisch gewesen sein, dass sie eine *itemspezifische Bedeutsamkeit* unabhängig von ihrer Bedeutung als zu verheimlichende oder nicht zu verheimlichende Items entwickelten. Hier spielen eventuell Peinlichkeits- oder Ekelgefühle bezüglich der Items eine Rolle, die unabhängig vom Täterschaftsstatus des jeweiligen Probanden auftraten und dadurch die Unterschiede zwischen den Bedingungen relativierten.

Solche unspezifischen Bedeutsamkeitseffekte könnten auch bezüglich anderer Itemtypen aufgetreten sein. In diesem Zusammenhang fällt auf, dass in der vorliegenden Untersuchung im Vergleich zu anderen Untersuchungen die P300 auf neutrale Items bei einigen Probanden sehr ausgeprägt ist (vgl. z.B. Farwell & Donchin, 1991; Kapitel 3.2.3.2). Große P300-Amplituden auf neutrale Reize können zu schlechter Klassifizierbarkeit des jeweiligen Probanden führen. Die Reaktionen auf neutrale Reize könnten mit den oben erwähnten Bedeutsamkeitseffekten erklärt werden, die einige Items unabhängig von ihrer Funktion im Oddball-Paradigma entwickelt haben könnten. Items wie „enge Hotpants“ sind offenbar weniger neutral als Items wie „graue Hose“ (Farwell & Donchin, 1991). Eine Untersuchung von Begleiter und Platz (1969) zeigt beispielsweise, dass sogenannte Tabuwörter („fuck“ und „shit“) geeignet sind, EKPs mit größeren Amplituden zu erzeugen als neutrale Wörter.

Durch die Verwendung einer *Hybridstudie* ergab sich ein weiterer Unterschied zu den bisher durchgeführten Untersuchungen: Wie einer Studie von Deegener (1995) zu entnehmen ist, verwenden Personen, die einer Sexualstraftat beschuldigt werden und

bei denen der Tatvorwurf gerechtfertigt ist, häufig Neutralisierungstechniken, um sich vor sich selbst und anderen zu rechtfertigen. Möglicherweise wirkt sich die Verwendung von Rechtfertigungen und Rationalisierungen negativ auf die Möglichkeit der Aufdeckung von Täuschung aus. Neutralisierungstendenzen sind nur dann zu erwarten, wenn Probanden annehmen eine antisoziale Handlung durchgeführt zu haben. Dies ist nur bei den Hybridstudien der Fall. Hinweise auf Rechtfertigungstendenzen traten in der vorliegenden Untersuchung im Nachgespräch unmittelbar nach der Untersuchung und in den Interviews ein Jahr später auf. Bemerkungen der Probanden, wie „Das ist aber normal, dass Männer so reagieren.“ oder „Eigentlich konnte ich das gut, aber im Feedbackbogen wurden nicht die richtigen Fragen gestellt.“ können als Hinweise auf Rechtfertigungstendenzen gewertet werden. In diese Richtung gingen auch Bemerkungen, die die Psychologen und ihre Experimente abwerteten, was als eine Neutralisierungstechnik aufgefasst werden kann, die Deegener (1995) als „*Verdammung der Verdammenden*“ bezeichnet. Die Auswirkungen solcher Rechtfertigungstendenzen sind bisher nicht berücksichtigt worden, obwohl sie auch im Rahmen anderer antisozialer Handlungen eine Rolle spielen (Sykes & Matza, 1957).

Zusammenfassend lassen sich folgende Unterschiede zwischen der vorliegenden Untersuchung und bisherigen Studien zur Täuschungsaufdeckung mit Hilfe der P300-Komponente registrieren: Erstens könnten sich Diskrepanzen hinsichtlich der Zeitspanne zwischen Erwerb der Information und EEG-Ableitung, sowie hinsichtlich des Informationserwerbs dahin gehend auswirken, dass tatspezifische Informationen unterschiedlich gut während der EEG-Ableitung präsent sind. Für die schlechte Klassifizierbarkeit können diese Faktoren in der vorliegenden Untersuchung nur in wenigen Extremfällen verantwortlich sein, da die Wiedererkennungslleistungen bei den meisten Probanden gut bis sehr gut waren. Zweitens ergaben sich Unterschiede hinsichtlich der Itemformulierung, die sich vermutlich sowohl auf die P300-Komponente bezüglich tatspezifischer als auch bezüglich neutraler Items dahin gehend ausgewirkt haben, dass eine korrekte Klassifizierbarkeit erschwert wurde. Eine weitere bisher nicht beachtete Variable stellen Rechtfertigungstendenzen der Probanden dar, die der Untersuchung im Hinblick auf die Klassifizierbarkeit der Täter bedürfen. Da es sich um die erste Studie handelt, die die P300-Komponente im Zusammenhang mit derartigem Itemmaterial im Rahmen einer Hybridstudie untersuchte, sollten weitere

Studien mit dem Ziel, Möglichkeiten und Grenzen der Formulierung „tabubezogener“ Items im Oddball-Paradigma durchgeführt werden. Dies sollte im Besonderen im Hinblick auf interindividuelle Unterschiede geschehen.

Bezüglich der unterschiedlichen Auswertungsstrategien bleibt Folgendes festzuhalten: Das korrelative Bootstrapverfahren bietet den Vorteil, dass in den Vergleich auch die Reaktionen auf tatunspezifische Reize einbezogen werden. Sowohl dieses Verfahren als auch die Amplitudenvergleiche erwiesen sich in einer Studie von Allen und Iacono (1997) als höchst effizient. Das heißt, die anhand dieser Methoden ermittelten Klassifikationen führten zu AUC-Werten, die über .90 lagen. Das korrelative Vorgehen zeigte sich in der Arbeit von Allen und Iacono (1997) gegenüber den Amplitudenvergleichen signifikant überlegen (siehe Kapitel 3.2.3.2). In der vorliegenden Untersuchung ließ sich dieser Unterschied nicht feststellen. Anhand der Signalentdeckungskurven ließ sich kein Unterschied bezüglich der Klassifikationsgüte ausmachen. Dies ist jedoch wahrscheinlich auf einen Bodeneffekt, das heißt auf schlecht klassifizierbare Daten zurückzuführen.

8.3 Diskussion unter methodischen Gesichtspunkten

Erstmals wurde das Oddball-Verfahren in einer *Hybridstudie* getestet. Als problematisch erwies es sich in bisherigen Hybridstudien Probanden zu gewinnen, die als Täter betrachtet werden konnten. Daher erfolgte die Zuteilung zu der Täterbedingung in der vorliegenden Untersuchung durch Suggestion einer schlechten Beobachtungsleistung. Ob diese Suggestion tatsächlich zu Täuschungsmotivation während der EEG-Ableitung führte, konnte lediglich anschließend überprüft, aber nicht im Voraus garantiert werden. Problematisch war diesbezüglich, dass der verhaltensnahe Validitätscheck durch Effekte der sozialen Erwünschtheit nicht in der Lage war, die tatsächliche Täuschungsmotivation der Probanden zu erfassen. Eine objektivere Überprüfung der Wirksamkeit des Vorhalts stellte die Erinnerungsüberprüfung der Probanden nach einem Jahr dar. Nachteilig ist der hohe Probandenverlust und die resultierende geringe Stichprobengröße. Das Procedere wäre dahin gehend zu verbessern, ein zeitnahe Erhebungsinstrument anzuwenden, das zuverlässig Auskunft über die Täuschungsmotivation der Probanden gibt, so dass die Stichprobengröße bei Probandenverlust angepasst werden kann.

Ein Teil der Falschklassifikationen ist wahrscheinlich darauf zurückzuführen, dass die Probanden die *Feedbackbögen nicht als Kriterium* für ihre Beobachtungsleistung heranzogen. Dieser Aspekt des Bonner Szenarios ist verbesserungswürdig. Ein Ausweg wäre, Probanden die Tat – das lustvolle Betrachten von Straftaten – tatsächlich begehen zu lassen. Dies würde jedoch wiederum Schwierigkeiten aufwerfen. Zu nennen wären zusätzliche ethische Bedenken und Probandenverlust, der dadurch noch erhöht wird, dass jeder Proband in einem anderen Kontext als Nichttäter fungieren soll. Ein anderer Weg bestünde darin, die Überzeugungskraft der Suggestion zu steigern.

Neben diesen beiden untersuchungstechnischen Problemen stellte die Notwendigkeit der *Probandentäuschung* ein ethisches Problem dar. Die Verwendung des Bonner Szenarios implizierte, dass die Probanden über das eigentliche Anliegen getäuscht werden mussten. Die Notwendigkeit der Probandentäuschung ergibt sich aus der Untersuchungslogik. Würde auf diesen Aspekt verzichtet, ließen sich Aspekte wie Ernsthaftigkeit oder das Auftreten von Neutralisierungstechniken nicht adäquat abbilden. Gründe, die gegen die Verwendung von Versuchspersonentäuschung vorgebracht werden, sind, dass es generell unethisch sei, Personen zu täuschen, dass diese Täuschung negative Effekte auf die Probanden habe und dass es sich nachteilig auf ihre Haltung gegenüber psychologischen Studien und der Psychologie als Wissenschaft auswirke (Kimmel, 1998). Christensen (1988) fand, dass Probanden, die über das tatsächliche Untersuchungsanliegen getäuscht worden waren, nach der Aufklärung angaben, dass es ihnen nichts ausgemacht habe, getäuscht worden zu sein. Außerdem berichteten sie häufiger, dass sie von dieser Erfahrung profitiert hätten, als Probanden, die nicht getäuscht wurden. Sie fühlten sich auch nicht in ihrer Privatsphäre verletzt. Obwohl bei solchen Angaben sicherlich Effekte kognitiver Dissonanz und sozialer Erwünschtheit eine Rolle spielen, ist die Versuchspersonentäuschung besonders vor dem Hintergrund möglichen Erkenntnisgewinns nicht als einseitig negativ zu bewerten. Bröders (1998) Argumentation, dass Probanden die Täuschung verstehen und sogar akzeptieren, wenn sie anschließend adäquat aufgeklärt werden, kann anhand der eigenen Erfahrung unterstützt werden. Vielleicht liegt auch in einer stärkeren Fokussierung des Gesprächs auf die Vorteile dieser Erfahrung eine Möglichkeit, den Probanden eine positivere Sichtweise des Vorgehens zu eröffnen. Probanden, die im Nachhinein nicht vollständig zufrieden mit der Situation waren, waren in der

vorliegenden Untersuchung solche, die mit besonders hohen Erwartungen teilgenommen hatten. Ein Proband, der als einziger auch nach Abschluss des Nachgesprächs verstimmt reagierte, äußerte schon vor Beginn der Untersuchung seinen Unmut gegenüber psychologischen Untersuchungen und zu einem späteren Zeitpunkt eine ungewöhnliche Sichtweise bezüglich der Bedeutung der PAG und über die Natur des Kindesmissbrauchs. Eventuell ließe sich auf dieser Grundlage in Zukunft eine strengere Vorauswahl der Probanden treffen.

8.4 Theoretische Einbettung der Ergebnisse und Diskussion im Hinblick auf die Anwendung der Methode

Die P300-Komponente wird a) durch methodische Komponenten, wie Darbietungshäufigkeit, Darbietungsdauer oder Ableitort, b) durch kognitive Variablen, wie Bedeutsamkeit oder Gedächtnisaspekte und c) von biologischen Determinanten, wie circadiane Rhythmen oder Drogengenuss, beeinflusst (Polich & Kok, 1995).

In der vorliegenden Untersuchung wurde eine Darbietungshäufigkeit von 1:4 sowohl für tatspezifische als auch für tatunspezifische Items gegenüber den neutralen Items gewählt, da sich dieses Verhältnis beispielsweise in der Untersuchung von Farwell und Donchin (1991) als effizient erwiesen hat. Die Darbietungsdauer betrug 300 ms. Eine Untersuchung von Polich (1990) vergleicht die Verwendung eines Interstimulusintervalls von 3000 ms mit der Verwendung eines Intervalls von 1500 ms. Da die P300-Amplituden für das längere Interstimulusintervall größer waren, wurde ein Intervall von 3000 ms gewählt. Als Ableitorte wurden diejenigen gewählt, von denen gezeigt werden konnte, dass die P300 hier mit der stärksten Positivierung auftritt (Rosenfeld 2002). Diese Parameter wurden über alle Probanden gleich gehalten. Da auch biologische Parameter die P300 beeinflussen, war es wichtig, ein Design zu wählen, in dem die unabhängigen Faktoren intraindividuell verglichen werden konnten. Die Probanden erhielten zwischen den verschiedenen Ableitungen nicht die Möglichkeit etwas zu konsumieren, so dass sich auch solche systematischen Veränderungen nicht einstellen konnten. Variiert wurden dagegen die kognitiven Variablen. Welche kognitiven Variablen sind es aber, die für die Differenzen hinsichtlich der P300-Amplitude in der vorliegenden Untersuchung verantwortlich sind?

Rosenfeld (2002) bezeichnet die P300-Komponente als einen Rekognitionsindex. Im Zusammenhang mit der vorliegenden Studie erscheint diese Bezeichnung nicht angemessen. Wenn die Rekognition der Items hinreichend wäre, um eine P300 mit größerer Amplitude hervorzurufen, hätten in der vorliegenden Untersuchung die meisten Probanden hinsichtlich beider Filmthemen als Täter klassifiziert werden müssen. Sie zeigten im Rekognitionstest, dass sie in der Lage waren tatspezifische Informationen als solche zu identifizieren.

Johnsons „Dreieckmodell“ (1986, 1993) fasst alle experimentellen Variablen, die die P300-Amplitude beeinflussen, in drei Kategorien oder Dimensionen zusammen: 1) Subjektive Wahrscheinlichkeit des Auftretens eines Stimulus (P), 2) Bedeutung des Stimulus (M) und 3) Informationsübertragung (T). Das Verhältnis des Beitrags der drei Dimensionen zur Ausprägung der P300-Komponente wird in folgender Gleichung ausgedrückt: $P300\text{-Amplitude} = f(T \times (1/P + M))$. Die additive Beziehung zwischen Stimulusbedeutung und subjektiver Auftretenswahrscheinlichkeit drückt die Unabhängigkeit des Einflusses der beiden Dimensionen aus. Beide Dimensionen werden dagegen mit der Größe der Informationsübermittlung zu gleichen Anteilen reduziert oder maximiert.

Der Einfluss der subjektiven Auftretenswahrscheinlichkeit auf die P300-Amplitude ist gut belegt. Je seltener ein Reiz oder eine Reizkategorie dargeboten wird, um so ausgeprägter fällt die P300-Amplitude aus (z.B. Duncan-Johnson & Donchin, 1977; Friedman, Simson, Ritter & Rapin; 1975, Johnson & Donchin, 1980). Neben der Darbietungshäufigkeit hat außerdem die Sequenz der Darbietung der Items einen Einfluss. So spielen z. B. Reizwiederholungen eine Rolle (Squires, Wickens, Squires & Donchin, 1976). In der vorliegenden Untersuchung wurde die Auftretenswahrscheinlichkeit über alle Probanden und Bedingungen konstant gehalten, systematische sequentielle Effekte wurden durch zufällige Darbietungsreihenfolge ausgeschlossen.

Die Stimulusbedeutung fasst drei untersuchte Variablen zusammen: 1) Komplexität der Aufgabe, 2) Komplexität des Stimulus und 3) Wert des Stimulus. Johnson (1986) weist darauf hin, dass „Stimulusbedeutung“ als vorläufige Bezeichnung in Ermangelung eines übergreifenden Verständnisses dieser Dimension zu verstehen ist. So bereitet es beispielsweise Schwierigkeiten, die Komplexität der Aufgabe zu quantifizieren. Studien, die den Einfluss der Komplexität belegen, vergleichen

beispielsweise eine Aufgabe, bei der Stimuli gezählt werden müssen, mit einer Reaktionszeit-Aufgabe, bei der die Probanden möglichst schnell eine Taste bedienen müssen. Die Komplexität der Reaktionszeit-Aufgabe gilt als höher. In einer Untersuchung von Wilkinson und Morlock (1967) resultierten größere P300-Amplituden aus dieser Aufgabe. Hinsichtlich der Stimulusqualität fand z. B. Verbatan (1983) größere P300-Amplituden bei Verwendung von komplexeren Mustern als Stimuli als bei einfacheren Stimulismustern. Der Stimuluswert oder die Stimulusbedeutsamkeit wurde in Experimenten als Belohnungsrate operationalisiert. Steinhauer fand z.B. eine hohe positive Korrelation zwischen der Höhe des Geldbetrags, der mit der korrekten Zuordnung eines Stimulus assoziiert war, und der Größe der P300-Amplitude. Unter diesen Punkt werden aber auch Ergebnisse subsumiert, die zeigen dass große P300-Amplituden aus der Darbietung von sehr lauten Tönen resultieren, auch wenn die Probanden diese nicht beachten sollten. Manche Reize, so interpretiert Johnson dieses Ergebnis, haben durch alltägliche Erfahrung Bedeutsamkeit. In der vorliegenden Arbeit bestanden keine systematischen Unterschiede hinsichtlich der Stimuluskomplexität. Jedem tatspezifischen Stimulus wurde ein hinsichtlich der semantischen Kategorie, der Silbenanzahl und Wortzahl passender tatspezifischer Reiz und entsprechende neutrale Reize zugeordnet. Die Komplexität der Aufgabe variierte in zweierlei Hinsicht: Erstens bestanden sie zwischen den Täterschaftsbedingungen, da der Täter auf tatspezifische Reize mit Täuschungsverhalten reagierte, indem er sie als neutral klassifizierte, der Nichttäter sie dagegen aufgrund seiner guten Beobachtungsleistung korrekt als neutral klassifizierte. Dies war ihm zuvor durch den Feedbackbogen bestätigt worden. Zweitens bestanden Unterschiede hinsichtlich der Aufgabenkomplexität bezüglich neutraler, tatspezifischer und tatunspezifischer Items. Ein tatunspezifisches Item erforderte auf jeden Fall einen Knopfdruck abweichend von neutralen Items. Dies war bei neutralen Items nicht der Fall. Sie sollten per Instruktion nicht beachtet werden. Damit war die Aufgabenstellung bezüglich tatunspezifischer Items komplexer, was insgesamt in einer über alle Bedingungen größeren P300-Amplitude bezüglich dieser Reize resultierte. Zwischen den Täterschaftsbedingungen variierte die Bedeutsamkeit der Items. Der Täter weiß um seine schlechte Beobachtungsleistung während des Films durch den Feedbackbogen. Gleichzeitig wurde ihm gesagt, dass das lustvolle Betrachten der Straftat und die daraus resultierende

Beschäftigung mit den entsprechenden Aspekten des Films sich in den verstärkten Reaktionen auf die Items zeigen würde, die solche Aspekte benennen. Ein tatspezifisches Item birgt also für den Täter die Gefahr der Aufdeckung seines unerwünschten Verhaltens. Für den Nichttäter, der weiß, dass er sich von unerwünschten Aspekten des Films nicht beeinflussen ließ, sind die tatspezifischen Items sehr viel weniger bedrohlich. Anhand dieser Bedeutsamkeitsvariable lässt sich auch das Ergebnis einordnen, dass Probanden, die eine Beschäftigung mit den Vorhalten erkennen ließen, eher erfolgreich klassifiziert wurden. Wenn die Probanden sich nicht mit den Inhalten dieses Vorhalts auseinandersetzen, entwickeln tatspezifische Items keine Bedeutsamkeit für einen Täter und er wird in jedem Fall als Nichttäter klassifiziert. Hat der Proband seine Täterschaft anders interpretiert als dies der Feedbackbogen vorgibt, wird sich die Bedeutsamkeit der tatspezifischen Items entsprechend verändern. Auch das bereits erwähnte Problem der itemspezifischen Bedeutsamkeit lässt sich gemeinsam mit den deliktspezifischen Unterschieden unter den Wertaspekt subsumieren.

Mit der P300-Komponente steht ein psychophysiologischer Index für kognitive Vorgänge zur Verfügung. Zwar existiert kein neuropsychologisches Modell der Täuschung in der Weise, dass eine spezifische Täuschungsreaktion abgeleitet werden könnte, doch impliziert die Täuschung häufig andere kognitive Phänomene, die mit Hilfe ereigniskorrelierter Potentiale messbar werden (s.o.). Rosenfeld (2002) stellt allerdings die Möglichkeit einer spezifischen Täuschungsreaktion in Aussicht. Er zieht als Index die Verteilung der P300-Amplitudenunterschiede über die verschiedenen Ableitorte heran. Abweichende Amplitudenprofile werden z. B. dann erhalten, wenn Probanden instruiert werden in einem Zwei-Kategorien-Oddball 50% der relevanten Items als irrelevant zu bezeichnen, obwohl sie sich des relevanten Charakters dieser Items bewusst sind. In diesem Fall unterscheidet sich die Verteilung der P300-Amplitudenwerte auf die Ableitorte zwischen Probanden, die wahrheitsgemäß antworten und solchen, die täuschen (Miller, 1999). Nach Rosenfeld (2002) beschäftigt sich seine Arbeitsgruppe zur Zeit damit, Algorithmen zu entwickeln, die eine Klassifikation anhand solcher Profile im Einzelfall ermöglichen. Es existieren bis jetzt jedoch keine differenzierten theoretischen Ansätze, die das Zustandekommen eines spezifischen Profils auf neuropsychologischer Ebene erklären könnten. Auch steht ein

Vergleich hinsichtlich der Klassifikationsgüte der Verwendung der P300-Profile mit der Verwendung der P300-Amplituden als abhängiger Variable noch aus.

Eine Weiterentwicklung des Ansatzes der Verwendung der P300 kann auch in einer simultanen Interpretation verschiedener EKP-Komponenten bestehen. So verwendeten MacLaren und Taululis (2000) einen Ansatz, der es ermöglichen sollte sowohl Schlussfolgerungen anhand der N400 als auch der P300 zu ziehen. Die P300 sollte bei Tätern mit großer Amplitude auf zutreffende Items wie „I shot Steven“ auftreten. Neutrale Items wurden durch unzutreffende Behauptungen gebildet, wie „I strangled Paul“. Um eine verstärkte N400 beim Täter festzustellen waren sowohl Sätze mit korrekter Endung als auch Sätze mit unzutreffender Endung im Itemsatz enthalten. Die Autoren erläutern nicht, wie genau die Rationale in Bezug auf die Klassifikation anhand der N400 lautet. Es werden auch lediglich die Ergebnisse bezüglich der P300-Komponente berichtet. Von 16 Probanden konnten in diesem Mock Crime mit Hilfe der P300 11 Probanden korrekt klassifiziert werden, eine Zuordnung fiel falsch-positiv und 4 Fälle unentscheidbar aus. Interessant wäre es, wenn die zusätzliche Betrachtung der N400 zu einer höheren Trefferrate führen würde als die Betrachtung der P300 allein.

Abschließend lässt sich feststellen, dass die Ergebnisse der vorliegenden Untersuchung sich mit Hilfe bestehender theoretischer Erklärungsansätze interpretieren lassen. Weitere Untersuchungen, die die genannten problematischen Aspekte sowohl bezüglich des Bonner Szenarios als auch des Oddball-Paradigmas aufgreifen, sind erforderlich, um festzustellen, inwieweit die Täuschungsaufdeckung mit Hilfe der P300 eine in der Praxis anwendbare, zuverlässige Methode darstellt. Die Möglichkeit, erweiterte Paradigmen, wie die Verwendung der P300-Profile als abhängige Variablen zur Täuschungsaufdeckung oder die Hinzuziehung anderer endogener Potentiale, zu untersuchen, stellt eine weitere Herausforderung für die Forschung dar.

9. Zusammenfassung

Die Forschung im Rahmen der Psychophysiologischen Aussagebegutachtung untersucht neben herkömmlichen Verfahren (Kontrollfragen- und Tatwissenverfahren) die Möglichkeit der Verwendung ereigniskorrelierter Hirnpotentiale (EKP) zur Täuschungsaufdeckung. Exemplarisch wird in der vorliegenden Untersuchung die Leistungsfähigkeit der P300-Komponente, die im Oddball-Paradigma auf seltene, bedeutsame Reize hin auftritt, zur Täuschungsaufdeckung erprobt.

Zur Anwendung kommt ein Drei-Kategorien-Oddball-Paradigma. Drei Kategorien von Stimuli spielen dabei eine Rolle: Neutrale, tatspezifische und tatunspezifische Stimuli. Tatspezifische Items stellen Informationen dar, von denen man annimmt, dass sie nur im Zusammenhang mit der inkriminierten Tat erworben werden konnten. Tatunspezifische Items stellen die zweite Kategorie von bedeutsamen Items dar. Sie werden vor der Ableitung eines Elektroenzephalogramms (EEG) auswendig gelernt. Während ein EEG abgeleitet wird, erscheinen neutrale, tatspezifische und tatunspezifische Items auf einem Bildschirm. Der Proband soll auf neutrale und tatspezifische Items beispielsweise eine Antworttaste betätigen, auf die zuvor gelernten, tatunspezifischen Items soll er dagegen mit dem Drücken einer anderen Taste reagieren. Neutrale Items sind neue Items, die weder aufgaben- noch tatbezogen sind.

Tatunspezifische Reize sollten unabhängig vom Täterstatus eine große P300-Amplitude, neutrale Items eine P300 mit geringer Amplitude hervorrufen. Es wurde angenommen, dass die P300 bei Probanden mit Tatwissen bei tatspezifischen Reizen eine größere Amplitude aufweist als bei neutralen Reizen. Die maximale Amplitude der P300, die bei einem Probanden ohne Tatwissen auf die tatspezifischen Reize hin auftritt, soll erwartungsgemäß kleiner oder gleich derjenigen bei neutralen Reizen sein. Die Verläufe der P300-Komponente bezüglich tatspezifischer Reize sollen für einen Täter dem Verlauf der Kurve bezüglich tatunspezifischer Reize ähneln, während die P300 für den Nichttäter im Verlauf erwartungsgemäß eher der bezüglich neutraler Reize gleicht. Da es in der Praxis schwierig ist, Tatinformationen zu generieren, die dem Nichttäter nicht bekannt sind, besaßen die Nichttäter in der vorliegenden Untersuchung relevante Tatinformationen, ohne dass jedoch der Tatvorwurf auf sie zutraf. Auf diese Weise sollte die Annahme überprüft werden, dass Nichttäter auch dann anhand der P300 korrekt klassifiziert werden können, wenn sie über Tatwissen verfügen.

Dreiundvierzig männliche Probanden im Alter zwischen 20 und 43 Jahren nahmen an der Untersuchung teil. Im Rahmen des so genannten „Bonner Szenarios“, das die Durchführung einer Hybridstudie in Anlehnung an eine Sexualstraftat ermöglicht, wurde 23 Probanden suggeriert, bezüglich eines Films, der ein Sexualdelikt zeigt, eine schlechte Beobachtungsleistung gezeigt zu haben. Eine schlechte Beobachtungsleistung implizierte, dass sich der entsprechende Proband in der Weise inadäquat mit den Filminhalten beschäftigt haben sollte, dass er mehr auf sexuelle Aspekte geachtet habe, anstatt sich an die vereinbarten Beobachtungsregeln zu halten. Bezüglich eines Films mit einer Mordthematik wurde diesen Probanden eine gute Beobachtungsleistung zurückgemeldet. Dies hieß gleichzeitig, dass sie die Beobachtungsregeln eingehalten hatten. Weiteren 20 Probanden wurde rückgemeldet, sich inadäquat mit der Mordthematik auseinandergesetzt zu haben. Diesen Probanden wurde bezüglich des Films mit der Sexualstraftatthematik eine gute Beobachtungsleistung bescheinigt. Die Rückmeldung erfolgte anhand so genannter „Feedbackbögen“.

Ein EEG wurde von Fz, Cz und Pz abgeleitet. Zur Artefaktkontrolle wurden zusätzlich die Augenbewegungen registriert. Tatunspezifische und tatspezifische Items wurden im Verhältnis 1 zu 4 zu den neutralen Items dargeboten. Die Darbietungszeit pro Item betrug 300 ms, die Intervalle zwischen den Items betragen 3000 ms.

Informationen, die per Instruktion während der Filme beobachtet werden sollten, stellten die tatunspezifischen Reize dar. Neue Items wurden als neutrale Reize verwendet. Aspekte der Filme, die mit Gewalt oder sexuellen Reizen zusammenhingen, stellten die tatspezifischen Informationen dar. Der Vorwurf bestand nämlich darin, dass die Probanden sich während der Filmdarbietung an den sexuellen oder gewaltbezogenen Aspekten erfreut hätten, statt die Beobachtungsaufgabe zu absolvieren.

Die Daten wurden sowohl gruppenstatistisch als auch einzelfallbezogen ausgewertet. Um der Frage nach den P300-Unterschieden bezüglich der Täterschaftsbedingung nachzugehen wurde eine 2 (Täterschaft) x 2 (Deliktthema) x 3 (Itemtyp) faktorielle Varianzanalyse mit Messwiederholung auf dem ersten und dritten Faktor durchgeführt. Als abhängige Variable wurden die maximalen P300-Amplituden verwendet. Zur Klassifikation im Einzelfall wurden zwei Auswertungsstrategien hinzugezogen, die einen Bootstrap-Algorithmus verwendeten, mit dessen Hilfe entweder Differenzen zwischen Amplitudenmaxima oder Korrelationsbeziehungen

zwischen den Kurven bezüglich unterschiedlicher Itemtypen verglichen wurden. Um der Annahme nachzugehen, dass es keine deliktbezogenen Unterschiede hinsichtlich der P300 gibt, wurde eine 2 (Täterschaft) x 2 (Deliktthema) x 3 (Itemtyp) faktorielle Varianzanalyse mit Messwiederholung auf dem zweiten und dritten Faktor berechnet. Die maximale P300-Amplitude diente als abhängige Variable.

Die gruppenstatistische Auswertung zeigt, dass diejenigen Probanden, die Täter bezüglich der Vergewaltigungsthematik und Nichttäter bezüglich der Mordthematik waren, die erwarteten Unterschiede zwischen Täter- und Nichttäter-Bedingung hinsichtlich der P300-Amplituden aufweisen. Anhand der Einzelfallauswertung konnten einige Probanden korrekt ihren Bedingungen zugeordnet werden. Insgesamt zeigt jedoch die Einzelauswertung sowohl mit Hilfe des korrelativen Verfahrens, als auch mit Hilfe der Amplitudenvergleiche, dass die Klassifizierung der Probanden nicht besser als zufällig erfolgt. Es ist festzustellen, dass falsch-positive Klassifikationen nicht überwiegen. Die hohe Fehlerrate ließ sich also nicht dadurch erklären, dass Nichttäter über tatspezifische Informationen verfügten und aufgrund dieses Wissens als Täter klassifiziert wurden. Es zeigten sich Gruppenunterschiede zwischen den Deliktthemen, die darauf zurückzuführen waren, dass die P300-Amplituden bezüglich der Sexualstraf-tatsthematik jeweils größer waren. Täter zeigten außerdem dann, wenn sie als Täter bezüglich der Vergewaltigungsthematik galten, im Vergleich zur Nichttäterbedingung größere positive Differenzen zwischen tatspezifischen und neutralen Items als solche Täter, die als Täter bezüglich der Mordthematik galten. Schließt man diejenigen Probanden aus, die sich durch den Vorhalt nicht beeindruckt zeigen und wählt als Kriterium für die Täterschaftsbedingungen die von den Probanden erinnerten Leistungen, fällt die Probandenklassifikation deutlich besser aus, als bei Betrachtung der Gesamtstichprobe.

Insgesamt fallen die Klassifikationen anhand der P300 weniger korrekt aus als in Studien, die bisher zur Täuschungsaufdeckung mit Hilfe der P300-Komponente durchgeführt wurden. Weitere Untersuchungen werden benötigt, um beispielsweise mehr über die Anforderungen an die Itemformulierung oder die Bedeutung von Rechtfertigungstendenzen bei den Probanden für die Klassifikation anhand der P300 zu erfahren.

10. Literatur

- Abrams, S. (1989). *The complete polygraph handbook*. Massachusetts: Lexington Books.
- Allen, J. J. B. & Iacono, W. G. (1997). A comparison of methods for the analysis of event-related potentials in deception detection. *Psychophysiology*, *34*, 234-240.
- Allen, J. J. B., Iacono, W. G. & Danielson, K. D. (1992). The identification of concealed memories using the event-related potential and implicit behavioral measures: A methodology of prediction in the face of individual differences. *Psychophysiology*, *29*, 504-522.
- Amelung, K. (1982). Anmerkungen zum Beschluss des BVERFG v. 18.8.81 – 2 BVR 166/81. *Neue Zeitschrift für Strafrecht*, *1*, 38-40.
- American Polygraph Association (1997). The validity and reliability of polygraph testing. *Polygraph*, *26*, 215-239.
- Aronson, E. (1994). *Sozialpsychologie: Menschliches Verhalten und gesellschaftlicher Einfluss*. Heidelberg: Spektrum Akademischer Verlag.
- Augustinus, A. (1979). *Die Lüge und Gegen die Lüge*. Würzburg: Echter.
- Backster, C. (1963). Polygraph professionalization through technique standardization. *Law and Order*, *11*, 63-64.
- Baldwin, K. B. & Kutas, M. (1997). An ERP analysis of implicit structured sequence learning. *Psychophysiology*, *34*, 74-86.
- Balloun, K. D. & Holmes, D. S. (1979). Effects of repeated examinations on the ability to detect guilt with a polygraphic examination: A laboratory experiment with a real crime. *Journal of Applied Psychology*, *64*, 316-322.
- Barland, G. H. & Raskin, D. C. (1973). Detection of Deception. In W. F. Prokasy & D. C. Raskin (Eds.), *Electrodermal Activity in Psychological Research* (pp. 417-477). New York: Academic Press.
- Barland, G. H. & Raskin, D. C. (1975). An evaluation of field techniques in detection of deception. *Psychophysiology*, *12*, 320-330.
- Begleiter, H. & Platz, A. (1969). Cortical evoked potentials to semantic stimuli. *Psychophysiology*, *6*, 91-100.
- Berg, P. (1999). *Konstanz Format. EEG and psychophysiology data analysis programs*. Unveröffentlichtes Manuskript.

- Berrien, F. K. (1942). Ocular stability in deception. *Journal of Applied Psychology*, 26, 55-63.
- Bersh, P. J. (1969). A validation study of polygraph examiner judgement. *Journal of Applied Psychology*, 53, 399-403.
- Birbaumer, N. & Schmidt, R. F. (1999). *Biologische Psychologie*. Heidelberg: Springer.
- Boaz, T. L., Perry, N. W., Raney, G., Fischler, I. S. & Shuman, D. (1991). Detection of guilty knowledge with event-related potentials. *Journal of Applied Psychology*, 76, 788-795.
- Bradley, M. T. & Ainsworth, D. (1984). Alcohol and the Psychophysiological Detection of Deception. *Psychophysiology*, 21, 63-71.
- Bradley, M. T. & Black, M. E. (1998). A control question test oriented towards students. *Perceptual and Motor Skills*, 87, 691-700.
- Bradley, M. T. & Cullen, C. (1993). Polygraph lie detection on real events in a laboratory setting. *Perceptual and Motor Skills*, 76, 1051-1058.
- Bradley, M. T. & Janisse, M. P. (1981). Accuracy demonstrations, threat and the detection of deception: Cardiovascular, electrodermal and pupillary measures. *Psychophysiology*, 18, 307-315.
- Bradley, M. T., MacLaren, V. V. & Carle, S. B. (1996). Deception and nondeception in guilty knowledge and guilty actions polygraph tests. *Journal of Applied Psychology*, 81, 153-160.
- Bradley, M. T. & Rettinger, J. (1992). Awareness of crime-relevant information and the guilty knowledge test. *Journal of Applied Psychology*, 77, 55-59.
- Bradley, M. T. & Warfield, J.F. (1984). Innocence, information and the guilty knowledge test in the detection of deception. *Psychophysiology*, 21, 683-689.
- Bröder, A. (1998). Deception can be acceptable. *American Psychologist*, 53, 805-806.
- Brunia, C. H. M. & Haagh, S. A. V. M. (1986). Preparation for action: Slow potentials and EMG. In H. Heuer & C. Fromm (Eds.). *Generation and Modulation of Action Patterns* (pp.28-40). Berlin: Springer.
- Bundesgerichtshof (1954). Urteil vom 16. Februar 1954 – 1 StR 578/53. *Entscheidungen des Bundesgerichtshofes in Strafsachen (BGHSt)*, 5, 332-338.
- Bundesgerichtshof (1999). Urteil vom 17. Dezember 1998 – 1 StR 156/98. *Praxis der Rechtspsychologie (Sonderheft)*, 9, 205-220.

- Bundesverfassungsgericht (1981). Beschluss vom 18. August 1981 – 2 BvR 166/81. *Neue Zeitschrift für Strafrecht*, 1, 446-447.
- Capps, M. H., Knill, B. L. & Evans, R. K. (1993). Effectiveness of the symptomatic questions. *Polygraph*, 22, 285-298.
- Christensen, L. (1988). Deception in psychological research: When is its use justified? *Personality and Social Psychology Bulletin*, 14, 664-675.
- Coles, M. G. H. & Rugg, M. D. (1995). Event-related brain potentials: an introduction. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of mind. Event-related brain potentials and cognition* (pp. 1-26). New York: Oxford University Press.
- Corcoran, J. F., Lewis, M. D. & Garver, R. B. (1978). Biofeedback-conditioned galvanic skin response and hypnotic suppression of arousal: A pilot study of their relation to deception. *Journal of Forensic Sciences*, 23, 155-162.
- Craick, F. I. M. & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671-684.
- Curio, I. & Scholz, O. B. (1991). Glaubhaftigkeitsbeurteilung von kurzen Zeugenaussagen mittels behavioraler und psychophysiologischer Parameter. *Zeitschrift für experimentelle und angewandte Psychologie*, 38, 188-200.
- Cutrow, R. J., Parks, A., Lucas, N. & Thomas, K. (1972). The objective use of multiple physiological indices in the detection of deception. *Psychophysiology*, 9, 578-588.
- Dawson, M. E. (1980). Physiological detection of deception: Measurement of responses to questions and answers during countermeasure maneuvers. *Psychophysiology*, 17, 8-17.
- Deegener, G. (1995). *Sexueller Mißbrauch: Die Täter*. Weinheim: Psychologie Verlags Union.
- Dickinson, A. & Dearing, M. F. (1979). Appetitive-aversive interactions and inhibitory processes. In A. Dickinson & R. A. Boakes (Eds.), *Mechanisms of learning and motivation* (pp. 203-231). Hillsdale: Erlbaum.
- Donchin, E., Ritter, W. & McCallum, C. (1978). Cognitive Psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Tueting & S. H. Koslow

- (Eds.), *Event-related brain potentials in man* (pp. 349-411). New York: Academic Press.
- Duncan-Johnson, C. C. & Donchin, E. (1977). On quantifying surprise: The variation of event-related potentials with subjective probability. *Psychophysiology*, *14*, 456-467.
- Ebbinghaus, H. (1985). *Über das Gedächtnis*. Leipzig: Duncker & Humblot.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, *7*, 1-26.
- Ekman, P. (1985). *Telling lies. Clues to deceit in the marketplace, politics, and marriage*. New York: W. W. Norton & Company, Inc.
- Elaad, E. (1990). Detection of guilty knowledge in real-life criminal investigations. *Journal of Applied Psychology*, *75*, 521-529.
- Elaad, E. (1994). The accuracy of human decisions and objective measurements in psychophysiological detection of knowledge. *The Journal of Psychology*, *128*, 267-280.
- Elaad, E. (1998). The challenge of the concealed knowledge polygraph test. *Expert Evidence*, *6*, 161-187.
- Elaad, E. & Ben-Shakhar, G. (1990). Effects of mental countermeasures on psychophysiological detection in the GKT. *International Journal of Psychophysiology*, *11*, 99-108.
- Elaad, E., Ginton, A. & Jungman, N. (1992). Detection measure in real-life criminal guilty knowledge tests. *Journal of Applied Psychology*, *77*, 757-767.
- Ellwanger, J., Rosenfeld, J. P., Hannkin, B. L. & Sweet, J. J. (1999). P300 as an index of recognition in a standard and difficult match-to-sample test: A model of amnesia in normal adults. *Clinical Neuropsychologist*, *13*, 100-108.
- Ellwanger, J., Rosenfeld, J. P., Sweet, J. J. & Bhatt, M. (1996). Detecting simulated amnesia for autobiographical and recently learned information using the P300 event-related potential. *International Journal of Psychophysiology*, *23*, 9-23.
- Endres, J., Scholz, O. B. & Summa, D. (1997). Aussagesuggestibilität bei Kindern – Vorstellung eines neuen diagnostischen Verfahrens und erste Ergebnisse. In M. Stadler, T. Fabian, L. Greuel (Hrsg.), *Psychologie der Zeugenaussage* (S. 189-204). München: PVU.

- Fabiani, M., Gratton, G. & Coles, M. G. H. (2000). Event-related brain potentials. In J. T. Caccioppo, L. G. Tassinary & G. G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 53-84). New York: Cambridge University Press.
- Farwell, L. A. & Donchin, E. (1991). The truth will out: Interrogative Polygraphy („Lie Detection“) with event-related brain potentials. *Psychophysiology*, 28, 531-547.
- Faul, F. & Erdfelder, E. (1992). *Gpower: Apriori, post-hoc and compromise power analysis for MS-DOS* (computer program). Friedrich-Wilhelms-Universität Bonn.
- Fisher, R. P. & Geiselman, R. E. (1992). *Memory-enhancing techniques for investigative interviewing*. Springfield: Charles C. Thomas.
- Frank, M. G., Ekman, P. & Friesen, W. V. (1997). Behavioural markers and recognizability of the smile of enjoyment. In P. Ekman & E. L. Rosenberg (Eds.), *What the face reveals. Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)* (pp. 217-238). New York: Oxford University Press.
- Friedman, D., Simson, R., Ritter, W. & Rapin, I. (1975). The late positive component (P300) and information processing in sentences. *Electroencephalography and Clinical Neurophysiology*, 38, 255-262.
- Furedy, J. J., Davis, C. & Gurevich, M. (1988). Differentiation of deception as a psychological process: A psychophysiological approach. *Psychophysiology*, 25, 683-688.
- Giesen, M. & Rollison, M. A. (1980). Guilty knowledge versus innocent associations: Effects of trait anxiety and stimulus context on skin conductance. *Journal of Research in Personality*, 14, 1-11.
- Ginton, A., Daie, N., Elaad, E., Ben-Shakhar, G. (1982). A Method for evaluating the use of the polygraph in a real-life situation. *Journal of Applied Psychology*, 67, 131-137
- Gödert, H. W., Rill, H.-G. & Vossel, G. (2001). Psychophysiological differentiation of deception: The effects of electrodermal lability and mode of responding on skin conductance and heart rate. *International Journal of Psychophysiology*, 40, 61-75.
- Gray, J. A. (1982). *The neuropsychology of anxiety*. Oxford: Clarendon Press.

- Green, D. M. & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Gudjonsson, G. H. (1983). Suggestibility, intelligence, memory recall and personality: An experimental study. *British Journal of Psychiatry*, 142, 35-37.
- Gudjonsson, G. H. (1984). A new scale of interrogative suggestibility. *Journal of Personality and Individual Differences*, 5, 303-314.
- Gudjonsson, G. H. (1993). *The psychology of interrogations, confessions and testimony*. Chichester: Wiley.
- Gudjonsson, G. H. & Clark, N. K. (1986). Suggestibility in police interrogation: A social psychological model. *Social Behaviour*, 1, 83-104.
- Gustafson, L. & Orne, M. T. (1963). Effects of heightened motivation on the detection of deception. *Journal of Applied Psychology*, 47, 408-411.
- Gustafson, L. & Orne, M. T. (1964). The effects of task and method of stimulus presentation on the detection of deception. *Journal of Applied Psychology*, 48, 383-387.
- Gustafson, L. & Orne, M. T. (1965). The effects of verbal responses on the laboratory detection of deception. *Psychophysiology*, 2, 10-13.
- Häcker, H. & Stapf, K. H. (1994). *Dorsch. Psychologisches Wörterbuch*. Göttingen: Hans Huber.
- Heckel, R. V., Brokaw, J. R., Salzberg, H. C. & Wiggins, S. L. (1962). Polygraphic variations in reactivity between delusional, non-delusional, and control groups in a "crime" situation. *Journal of Criminal Law, Criminology and Police Science*, 53, 380-383.
- Heinrich, Christina U. (1995). *Ausdruck und Eindruck. Die Zuschreibung von Glaubwürdigkeit anhand visueller und vokaler Informationen*. Dissertation, Universität Bielefeld.
- Hira, S. & Matsuda, T. (1998). Contingent negative variation (CNV) in detection of deception task using a serial presentation of pictures, *The Japanese Journal of Psychology*, 69, 149-155.
- Hopper, R. & Bell, R. A. (1984). Broadening the deception construct. *Quarterly Journal of Speech*, 70, 288-302.

- Honts, C. R. (1994). Psychophysiological detection of deception. *Current Directions in Psychological Science*, 3, 77-82.
- Honts, C. R. (1996). Criterion development and validity of the CQT in field application. *The Journal of General Psychology*, 123, 309-324.
- Honts, C. R., Devitt, M. K., Winbush, M. & Kircher, J. C. (1996). Mental and physical countermeasures reduce the accuracy of the concealed knowledge test. *Psychophysiology*, 33, 84-92.
- Honts, C. R. & Hodes, R. L. (1983). The detection of physical countermeasures. *Polygraph*, 12, 7-17.
- Honts, C. R. & Raskin, D. C. (1988). A field study on the validity of the directed lie control question, *Journal of Police Science and Administration*, 16, 56-61.
- Honts, C. R., Raskin, D. C. & Kircher, J. C. (1987). Effects of physical countermeasures and their electromyographic detection during polygraph tests for deception. *Journal of Psychophysiology*, 1, 241-247.
- Honts, C. R., Raskin, D. C. & Kircher, J. C. (1994). Mental and physical countermeasures reduce the accuracy of polygraph tests. *Journal of Applied Psychology*, 79, 252-259.
- Honts, C. R., Raskin, D. C., Kircher, J. C. & Hodes, R. L. (1988). Effects of spontaneous countermeasures on the physiological detection of deception. *Journal of Police Science and Administration*, 16, 91-94.
- Hopper, R. & Bell, R. A. (1984). Broadening the deception construct. *Quarterly Journal of Speech*, 70, 288-302.
- Horowitz, I. S. (1999). Die Arbeit mit dem Polygraphen in Israel. In U. Willutzki, J. Salzgeber & M. Stadler (Hrsg.), *Polygraphie. Möglichkeiten und Grenzen der PAG* (S. 165-176). Köln: Bundesanzeiger-Verlag.
- Horowitz, S. W., Kircher, J. C., Honts, C. R. & Raskin, D. C. (1997). The role of comparison questions in physiological detection of deception. *Psychophysiology*, 34, 108-115.
- Horvath, F. S. (1977). The effect of selected variables on interpretation of polygraph records. *Journal of Applied Psychology*, 62, 127-136.
- Horvath, F. S. (1979). Effect of different motivational instructions on detection of deception with the psychological stress evaluator and the galvanic skin response.

- Psychophysiology*, 64, 323- 330.
- Horvath, F. S. (1994). The Value and effectiveness of the sacrifice relevant question: An empirical assessment. *Polygraph*, 23, 261-279.
- Horvath, F. S. & Reid, J. E. (1971). The reliability of polygraph examiner diagnoses of truth deception. *Journal of Criminal Law, Criminology and Police Science*, 62, 276-281.
- Iacono, W. G. (1991). Can we determine the accuracy of polygraph tests? In J. R. Jennings, P. K. Ackles & M. G. H. Coles (Eds.), *Advances in psychophysiology* (Vol. 4, pp. 201-207). London: Jessica Kingsley Publishers.
- Iacono, W. G. & Boisvenu, G. A. & Fleming, J. A. (1984). Effects of diazepam and methylphenidate on the electrodermal detection of guilty knowledge. *Journal of Applied Psychology*, 69, 289-299.
- Iacono, W. G. & Patrick, C. J. (1987). What Psychologists should know about lie detection. In I. B. Weiner & A. K. Hess, *Handbook of forensic psychology* (pp. 460-489). New York: Wiley.
- Jager, P. (1997). *Pseudoerinnerungen bei Kindern im Vorschulalter und individuelle Suggestibilitätsunterschiede*. Unveröffentlichte Diplomarbeit, Rheinische Friedrich-Wilhelms-Universität, Bonn.
- Jasper, H. H. (1958). The ten-twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, 10, 371-375.
- Johnson, R. (1986). A Triarchic Model of P300 amplitude. *Psychophysiology*, 23, 367-384.
- Johnson, R. (1993). On the neuronal generators of the P300 component of the event-related potential. *Psychophysiology*, 30, 90-97.
- Johnson, R. & Donchin, E. (1980). P300 and stimulus categorization: Two plus one is not so different from one plus one. *Psychophysiology*, 17, 167-178.
- Johnson, M. M. & Rosenfeld, J. P. (1992). Oddball-evoked P300-based method of deception detection in the laboratory II: Utilization of non-selective activation of relevant knowledge. *International Journal of Psychophysiology*, 12, 289-306.
- Kassin, S. M., Ellsworth, P. C. & Smith, V. L. (1989). The "general acceptance" of psychological research on eyewitness testimony: A survey of the experts. *American Psychologist*, 44, 1089-1098.

- Kassin, S. M. & Kiechel, K. (1996). The social psychology of false confessions: Compliance, internalisation and confabulation. *Psychological Science*, 7, 125-128.
- Kimmel, A. J. (1998). *Ethical issues in behavioral research: A survey*. Cambridge, MA: Blackwell.
- Kircher, J. C, Horowitz, S. W. & Raskin, D. C. (1988). Meta-analysis of mock crime studies of the control question polygraph technique. *Law and Human Behavior*, 12, 79-90.
- Kleiner, M. (2002). Physiological detection of deception in psychological perspectives: A theoretical proposal. In M. Kleiner (Ed.), *Handbook of polygraph testing* (pp. 127-182). London: Academic Press.
- Köhnken, G. (1990). *Glaubwürdigkeit*. München: Psychologie Verlags Union.
- Köhnken, G. & Wegener, H. (1985). Zum Stellenwert des Experiments in der Forensischen Aussagepsychologie. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 32, 104-119.
- Kugelmass, S. & Lieblich, I (1966). Effects of realistic stress and procedural interference in experimental lie detection. *Journal of Applied Psychology*, 50, 211-216.
- Kutas, M. & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203-207.
- Loveless, N. E. & Sanford, A. J. (1974). Slow potential correlates of preparatory set. *Biological Psychology*, 1, 303-314.
- Lykken, D. T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology*, 43, 385-388.
- Lykken, D. T. (1960). The validity of the guilty knowledge technique: the effects of faking. *Journal of Applied Psychology*, 44, 258-262.
- Lienert, G. A. (1994). *Testaufbau und Testanalyse* (5. Aufl.). Weinheim: Beltz.
- MacLaren, V. & Taukulis, H. (2000). Forensic identification with event related potentials. *Polygraph*, 29, 330-343.
- Matsuda, T., Hira, S., Nakata, M. & Kakigi, S. (1990). The effect of one's own name upon event related potentials. *Japanese Journal of Physiological Psychology and Psychophysiology*, 8, 9-18.

- Matte, J. A. (1996). *Forensic psychophysiology using the polygraph*. New York: J. A. M. Publications.
- Miller, A. R. (1999). *P300 amplitude and topography in pseudomemory phenomena*. Unpublished Doctoral Dissertation, Northwestern University, Evanston.
- Mossman, D. & Somoza, E. (1991). ROC curves, test accuracy, and the description of diagnostic tests. *Journal of Neuropsychiatry*, 3, 330-333.
- Musch, J. & Rossnagel, C (1999). *Pupillenbasierte Lügendetektion im Tatwissentest*. Vortrag, gehalten auf dem 20. Kongress für angewandte Psychologie der Bundes Deutscher Psychologen. Berlin 7. - 10. Oktober 1999.
- Neville, H. J., Kutas, M., Chesney, G. & Schmidt, A. L. (1986). Event-related brain potentials during initial encoding and recognition memory of congruous and incongruous words. *Journal of Memory and Language*, 25, 75-92.
- Nietzsche, F. W. (1980). *Die Geburt der Tragödie. Unzeitgemäße Betrachtungen I-IV. Nachgelassene Schriften 1870-1873*. München: Deutscher Taschenbuch Verlag/Walter de Gruyter.
- Oberlandesgericht Bamberg (1995). Beschluss vom 14. März 1995 – 7 WF 122/94. *Neue Juristische Wochenschrift*, 48, 1684-1685.
- Öhman, A., Hamm, A. & Hugdahl, K. (2000). Cognition and the autonomic nervous system. In J. T. Caccioppo, L. G. Tassinary & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (pp. 533-575). Cambridge: University Press.
- Offe, H. (2001). Fallbericht: Polygraphentest im Familienverfahren. *Praxis der Rechtspsychologie*, 11, 55-57.
- Offe, H. & Offe, S. (2001). Das BGH – Urteil zum Polygraphen: Eine Herausforderung an die Psychologie. *Praxis der Rechtspsychologie*, 11, 5-15.
- Office of Technology Assessment (OTA) (1987). *Scientific Validity of Polygraph Testing: A Research Review and Evaluation*. Washington: Eden Press.
- Orne, M. T., Thackray, R. I. & Paskewitz, D. A. (1972). On the detection of deception. A model for the study of physiological effects of psychological stimuli. In N. S. Greenfield & R. A. Sternbach (Eds.), *Handbook of psychophysiology* (pp. 743-785). New York: Holt, Reinhart and Winston, Inc.

- O'Toole, D., Yuille, J. C., Patrick, C.-J. & Iacono, W. G. (1994). Alcohol and the physiological detection of deception: Arousal and memory influences. *Psychophysiology*, *31*, 253-263.
- Patrick, C. J. & Iacono, W. G. (1989). Psychopathy, threat, and polygraph test accuracy. *Journal of Applied Psychology*, *74*, 347-355.
- Patrick, C. J. & Iacono, W. G. (1991). Validity of the control question polygraph test: The problem of sampling bias. *Journal of Applied Psychology*, *76*, 229-238.
- Polich, J. (1990). Probability and inter-stimulus interval effects on the P300 from auditory stimuli. *International Journal of Psychophysiology*, *10*, 163-170.
- Polich, J. & Kok, A. (1995). Cognitive and biological determinants of P300: An integrative review. *Biological-Psychology*, *41*, 103-146.
- Raskin, D. C. (1989). Polygraph techniques for the detection of deception. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 247-296). New York: Springer.
- Raskin, D. C. & Hare, R. D. (1978). Psychopathy and detection of deception in a prison population. *Psychophysiology*, *15*, 126-136.
- Raskin, D. C. & Honts, C. R. (2002). The Comparison Question Test. In M. Kleiner (Ed.), *Handbook of polygraph testing* (pp. 1-48). London: Academic Press.
- Raskin, D. C., Honts, C. R. & Kircher, J. C. (1997). The scientific status of research on polygraph techniques: The case for polygraph tests. In D. Faigman, D. H. Kaye, K. J. Saks & J. Sanders (Eds.), *Modern scientific evidence: The law and science of expert testimony* (pp. 565-582). St. Paul, Min.: West Publishing Co.
- Raskin, D. C., Kircher, J. C., Honts, C. R. & Horowitz, S. W. (1988). *A study of the validity of polygraph examinations in criminal investigation*. Final report to the National Institute of Justice, Grant No. 85-IJ-CX-0040.
- Raskin, D. C. & Steller, M. (1989). Assessing credibility of allegations of child sexual abuse: Polygraph examinations and statement analysis. In: H. Wegener, F. Lösel & J. Haisch (Eds.), *Criminal behavior and the justice system. Psychological perspectives* (pp. 290-302). New York: Springer.
- Reid, J. E. (1947). A revised questioning technique in lie-detection tests. *Journal of Criminal Law and Criminology*, *37*, 542-547.
- Rösler, F. (1982). *Hirnelektrische Korrelate kognitiver Prozesse*. Berlin: Springer.

- Rosenfeld, J. P. (2002). Event – Related Potentials in Detection of Deception. In M. Kleiner (Ed.), *Handbook of Polygraphy* (pp. 265-286). New York: Academic Press.
- Rosenfeld, J. P., Angell, A., Johnson, M. & Quian, J. (1991). An ERP-based, control-question lie detector analogue. *Psychophysiology*, 28, 319-335.
- Rosenfeld, J. P., Cantwell, B., Nasman, V. T., Wodjac, V., Ivanov, S. & Mazzeri, L. (1988). A modified, event-related potential-based Guilty Knowledge Test. *International Journal of Neuroscience*, 42, 157-161.
- Rosenfeld, J. P., Nasman, V. T., Whalen, R., Cantwell, B. & Mazzeri, L. (1987). Late vertex positivity in event-related potentials as a guilty knowledge indicator: A new method of lie detection. *International Journal of Neuroscience*, 34, 125-129.
- Rovner, L. I., Raskin, D. C. & Kircher, J. C. (1979). Effects of information and practice on detection of deception. *Psychophysiology*, 16, 197-198.
- Sartory, R. (1999). *Vergleich zweier Forschungsmethoden zur Feststellung von Täuschungsverhalten*. Unveröffentlichte Diplomarbeit, Rheinische-Friedrich-Wilhelms-Universität, Bonn.
- Saxe, L., Dougherty, D. & Cross, T. (1985). The validity of polygraph testing. *American Psychologist*, 40, 355-366.
- Scholz, O. B. (1999): Die physiologischen Parameter in der psychophysiologischen Aussagebegutachtung (PAG). In U. Willutzki, J. Salzgeber & M. Stadler (Hrsg.), *Polygraphie. Möglichkeiten und Grenzen der PAG* (S. 45-67). Köln: Bundesanzeiger-Verlag.
- Schwabe, J. (1982). Der „Lügendetektor“ vor dem Bundesverfassungsgericht. *Neue Juristische Wochenschrift*, 8, 367-368.
- Scullin, M. H. & Ceci, S. J. (2001). A suggestibility scale for children. *Personality and Individual Differences*, 30, 843-856.
- Sokolov, E. N. (1967). Die reflektorischen Grundlagen der Wahrnehmung. In Hiebsch, H. (Hrsg.), *Ergebnisse der sowjetischen Psychologie* (S. 471-504). Berlin: VEB, Deutscher Verlag der Wissenschaften.
- Squires, K. C., Donchin, E., Herning, R. I. & McCarthy, G. (1977). On the influence of task relevance and stimulus probability on event-related potential components.

- Electroencephalography and Clinical Neurophysiology*, 42, 1-14.
- Squires, K. C., Wickens, C., Squires, N. K. & Donchin, E. (1976). The effect of stimulus sequence on the waveform of the cortical event-related potential, *Science*, 193, 1142-1146.
- Stadler, M. A. (2001). Sexuelle Handlung oder sexuelle Absicht. *Praxis der Rechtspsychologie*, 11, 48-54.
- Steller, M. (1987). *Psychophysiologische Aussagebeurteilung*. Göttingen: Hogrefe.
- Steller, M. & Dahle, K.-P. (1999). Grundlagen, Methoden und Anwendungsprobleme psychophysiologischer Aussage- bzw. Täterschaftsbeurteilung. *Praxis der Rechtspsychologie*, 9 (Sonderheft BGH-Gutachten: Psychophysiologische Aussagebeurteilung), 127-204.
- Steller, M. & Köhnken, G. (1989). Criteria – based statement analysis. In C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 217-245). New York: Springer.
- Stern, R. M., Breen, J. P., Watanabe, T. & Perry, B. S. (1981). Effects of feedback of physiological information on responses to innocent associations and guilty knowledge. *Journal of Applied Psychology*, 66, 677-681.
- Sutton, S., Braren, M., Zubin, J. & John, E. R. (1965). Evoked potential correlates of stimulus uncertainty. *Science*, 150, 1187-1188.
- Sykes, G. M. & Matza, D. (1957). Techniques of neutralisation: A theory of delinquency. *American Sociological Review*, 22, 664-670.
- Timm, H. W. (1987). Effect of biofeedback on the detection of deception. *Journal of Forensic Sciences*, 32, 736-746.
- Travis, F. & Tecce, J. J. (1998). Effects of distracting stimuli on CNV amplitude and reaction time. *International Journal of Psychophysiology*, 31, 45-50.
- Undeutsch, U. (1967). Beurteilung der Glaubhaftigkeit von Aussagen. In U. Undeutsch (Hrsg.), *Handbuch der Psychologie, Band 11: Forensische Psychologie* (S. 26-181). Göttingen: Hogrefe.
- Undeutsch, U. & Klein, G. (1999). Wissenschaftliches Gutachten zum Beweiswert physiologischer Untersuchungen. *Praxis der Rechtspsychologie (Sonderheft Psychophysiologische Aussagebeurteilung)*, 9, 45-126.
- Van Hooff, J. C., Brunia, C. H. M. und Allen, J. J. B. (1996). Event related potentials as

- indirect measures of recognition memory. *International Journal of Psychophysiology*, 21, 15-31.
- Vehrs, W. (1999). Standards einer psychophysiologischen Aussagebegutachtung (PPAB) am Institut für Forensische Psychophysiologie (IFP). In U. Willutzki, J. Salzgeber & M. Stadler (Hrsg.), *Polygraphie. Möglichkeiten und Grenzen der PAG* (S. 19-30). Köln: Bundesanzeiger-Verlag.
- Vehrs, W. (2001). Psychophysiologische Aussagebegutachtung (PPAB) mit der Vergleichsfragentechnik (VFT) in familiengerichtlichen Verfahren. *Praxis der Rechtspsychologie*, 11, 16-25.
- Verbaten, M. N. (1983). The influence of information on habituation of cortical, autonomic and behavioral components of the orienting response. In A. W. K. Gaillard & W. Ritter (Eds.), *Tutorials in ERP research: Endogenous components* (pp. 201-216). Amsterdam: Elsevier.
- Vossel, G., Rill, H.-G. & Gödert, H. W. (2000). Differentiation of deception: Individual analysis and numerical scoring of skin conductance responses. *Journal of Psychophysiology*, 14, 58.
- Waid, W. M. & Orne, M. T. (1981). Cognitive, social, and personality processes in the physiological detection of deception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 61-106). New York: Academic Press.
- Waid, W. M., Orne, E. C., Cook, M. R. & Orne, M. T. (1981). Meprobamate reduces accuracy of physiological arousal in the detection of deception. *Journal of Applied Psychology*, 66, 224-232.
- Waid, W. M., Orne, M. T. & Wilson, S. K. (1979). Effects of level of socialization on electrodermal detection of deception. *Psychophysiology*, 16, 15-22.
- Waid, W. M., Wilson, S. K. & Orne, M. T. (1981). Cross-modal physiological effects of electrodermal lability in the detection of deception. *Journal of Personality and Social Psychology*, 40, 1118-1125.
- Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C. & Winter, A. L. (1964). Contingent negative variation: An electrical sign of sensorimotor association and expectancy in the human brain, *Nature*, 203, 380-384.
- Wasserman, S. & Bockenholt, U. (1989). Bootstrapping: Applications to Psychophysiology. *Psychophysiology*, 26, 208-221.

- Wilkinson, R. T. & Morlock, H. C. (1967). Auditory evoked response and reaction time. *Electroencephalography and Clinical Neurophysiology*, 23, 50-56.
- Zhou, L., Yang, W., Liao, S. & Zou, H. (1999). Experimental study of lie detection with P300 in simulated crime. *Chinese Journal of Clinical Psychology*, 7, 31-33.
- Zhou, L., Yang, W., Liao, S. & Zou, H. (2000). Compared study of event-related potentials in lie detection between subjects for simulated crime condition and subjects for visiting condition. *Chinese Journal of Clinical Psychology*, 8, 86-88.

Anhang

- A Einverständniserklärung
- B Beobachtungsregeln
- C Feedbackbögen für „Mord“ und „Sexualstraftat“; für die Bedingungen „Täter“ (2) und „Nichttäter“ (1)
- D Itemlisten
- E Beschreibung der Bootstrapverfahren 1 und 2
- F Antworten der Probanden auf die Frage „Wie haben Sie damals auf den Vorwurf reagiert?“, die ein Jahr nach der Untersuchung gestellt wurde.

Anhang A
Einverständniserklärung

Informationsblatt zu psychophysiologischen Experimenten in der Abteilung Klinische Psychologie

Betr: Experiment zur Überprüfung der Beobachtungsleistung durch EEG

Verantwortlicher Versuchsleiter:

Proband Nr.

Mir ist mitgeteilt worden, dass das Experiment an zwei Tagen stattfinden und jeweils ca. Minuten dauern wird. Es wurde mir versichert, dass körperliche oder seelische Schädigungen als Folgen des Experiments nicht auftreten werden.

Ich weiß, dass ich jederzeit frei bin, das Experiment abubrechen. Ich weiß auch, dass ich die mir gegebenen Informationen vertraulich behandeln muss und nicht an andere potentielle Versuchspersonen weitergeben darf.

Es wird mir völlige Anonymität zugesichert. Die Daten des Experiments werden so gespeichert und aufbewahrt, dass Rückschlüsse auf meine Person nicht möglich sind. Was die Videomitschnitte betrifft, so werden sie lediglich zu wissenschaftlichen Zwecken verwendet.

Nach Beendigung der Versuchsreihe bin ich berechtigt, den Hintergrund des Experiments, alle Details des Versuchsablaufs, die Gründe für die Durchführung des Experiments, den Verwendungszweck der Daten und auch die Ergebnisse zu erfahren. Ich weiß, dass ich diesen Anspruch dadurch geltend mache, dass ich nach den entsprechenden Informationen frage.

Mit meiner Unterschrift bestätige ich, dass ich freiwillig bereit bin, an diesem Experiment als Versuchsperson teilzunehmen.

Bonn, den

.....

Unterschrift

Anhang B

Beobachtungsregeln

Emotionale Erregung, Wahrnehmung und Erinnerung

Kassin, Ellsworth und Smith (1989) befragten Experten der „Psychologie der Augenzeugen“ zu verschiedenen Aspekten. Ein Ergebnis dieser Befragung bestand darin, dass die Mehrzahl der Experten der Meinung war, dass emotional erregende Situationen schlechter erinnert werden. Das trifft aber nur teilweise zu.

Im folgenden wird auf einige Erkenntnisse hingewiesen, die in diesem Zusammenhang von Bedeutung sind. Es lassen sich aus diesen Informationen bestimmte Verhaltensregeln ableiten, die zu einer Verbesserung der Wahrnehmungs- und Erinnerungsleistung führen. Diese Techniken sollen mit Ihnen trainiert und ihre Anwendung erprobt werden. Wenn Sie die Anweisungen befolgen, haben Sie die Chance, Ihre Beobachtungsleistung in Extremsituationen zu verbessern. Die Relevanz dieser Fähigkeit wird Ihnen sicherlich offensichtlich sein.

Von-Restorff-Effect: Tritt in einer homogenen Reihe von Items ein emotionales, herausragendes Item auf, so wird dieses besser erinnert. Gleichzeitig besteht retrograde Amnesie für das vorherige Item. Eine Gruppe von Items mit einem solchen emotionalen Item wird schlechter erinnert als eine Gruppe ohne ein solches Item. Das emotionale Ereignis an sich wird gut erinnert, aber die zeitlich nach oder vorhergehenden Informationen werden schlechter erinnert.

Übereinstimmung der emotionalen Grundstimmung: Die Erinnerungsleistung ist davon abhängig, wie gut die emotionale Grundstimmung in der Enkodierphase und der Abfragephase übereinstimmen.

Zentrale vs. periphere Information: Zentrale Aspekte einer emotional erregenden Situation werden besser erinnert als in einem neutralen Kontext. Periphere Aspekte werden in einer emotional erregenden Situation schlechter erinnert als in einer neutralen Situation. Hier ist der *Waffenfokuseffekt* (der Aufmerksamkeitsfokus des Zeugen war auf

die Waffe gerichtet, so dass er später z.B. das Gesicht des Täters nicht beschreiben kann) oder auch der sogenannte *Tunnelblick* (die gesamte Aufmerksamkeit ist z. B. auf die Gewalttat gerichtet) einzuordnen.

Sicherheit der Erinnerung: Untersuchungen zeigen, dass das Maß, in dem sich Probanden ihres Urteils bezüglich erinnelter Informationen sicher sind, nicht mit der Korrektheit dieser Informationen zusammenhängt. Ein berühmtes Beispiel sind Flashbulb-Memories, die mit großer subjektiver Sicherheit erinnert werden, aber häufig Falschinformationen enthalten

Daraus lassen sich folgende **Beobachtungsregeln** ableiten:

- Emotionale Verfassung während der Enkodierung sollte der Verfassung während des Abrufs der Informationen gleichen -> Versuchen Sie sich in emotional erregenden Situationen zu entspannen, die gleiche Verfassung lässt sich dann später leichter herstellen. Atmen Sie z.B. ein und lang und tief aus.
- Versuchen Sie, Gefühle so weit wie möglich auszuschalten, Konzentrieren Sie sich auf das Beobachten. Lassen Sie nur ihren Verstand arbeiten!
- Beschäftigen Sie sich gedanklich nicht mit den zentralen, reißerischen Aspekten der Situation (z.B. Wunden). Das lenkt ab und involviert sie emotional. Versuchen Sie, nicht empathisch zu reagieren.
- Gelingt es Ihnen nicht, sich emotional zu distanzieren, versuchen sie sich durch eine kurze impulsive Handlung abzulenken (mit dem Fuß aufstampfen)!
- Beobachtungsverhalten: In emotional erregenden Situationen neigen Menschen dazu, den Fokus auf zentrale Aspekte zu lenken. Dadurch werden periphere Details schlechter erinnert als in neutralen Situationen. -> Periphere Merkmale ganz bewusst, beachten, nicht von zentralen, reißerischen Aspekten (Blut, Verletzte, Waffen) der Situation ablenken lassen.
- Wenn eine emotional erregende Situation auftritt, wiederholen Sie zeitlich vorangegangene Details im Kopf. Elaboriertes Erinnern beugt dem Vergessen vor!
- Seien Sie sich bewusst, dass auch zeitlich nachgeschaltete Informationen durch das Auftreten der emotional erregenden Stimuli schlechter verarbeitet und erinnert

werden. Ziehen Sie daher in Gedanken einen Schlussstrich und konzentrieren Sie sich ganz auf die aktuellen Ereignisse. Lassen Sie nicht zu, dass Ihre Aufmerksamkeit abnimmt! Bleiben Sie wachsam!

- Nehmen Sie später keine Auswahl der zu erinnernden Details nach Aspekten ihrer Urteilssicherheit vor. Auch wenn Sie sich absolut nicht sicher sind, dass es so war, kann diese Information sich später als wichtig erweisen.

Der Untersuchungsleiter wird nun mit Ihnen die verschiedenen Techniken noch einmal durchsprechen. Das soll Ihnen auch Gelegenheit geben, eventuell bei Ihnen aufgetretene Fragen bezüglich der Beobachtungsregeln zu erörtern.

Anhang C Feedbackbogen (Mordszene)

1

Feedbackbogen

Wenn Sie die nachfolgenden Fragen beantworten und die Auswertungsregeln beachten, können Sie sich ein Bild darüber verschaffen, wie gut Ihre Beobachtungsleistung beschaffen war.

Film B: Mord

1. Hatte der Täter eine Warze auf der Nase oder auf der Wange?
2. Wie viele Zähne fehlten dem herausgefallenen Gebiß?
3. War die Schrift auf der Innenseite der Beifahrertür lila oder weiß?
4. Standen Strommasten an dem Weg, auf dem die geschädigte Person überwältigt wurde?
5. Wie kam es dazu, daß die Wagenscheibe verschmutzt wurde und wodurch wurde sie verschmutzt?
6. Wie oft schlug der Täter der geschädigten Person den Felsbrocken auf den Schädel (Geben Sie eine genaue Zahl an!)?
7. Welche Farbe hatte die lederne Tasche, aus der der Täter die Eisenstange nahm?
8. Welche Farbe hatte das Karo der wollenen Decke, durch das das Blut sickerte?
9. Was machte dem Täter zu schaffen, bevor er sich entschloß die vermeintliche Leiche zu entsorgen und durch welche Einstellung wird dies deutlich?
10. Baumelte vom Rückspiegel des Wagens ein Puppenkopf aus Gummi oder ein Rosenkranz?
11. Wie sah die Jacke aus, die der Täter trug?
12. In einer der Szenen wird die Aufmerksamkeit des Betrachters gezielt auf einen bestimmten Körperteil gelenkt, auf welchen?

Das Bewertungsschema können Sie sich vom VL aushändigen lassen.

Feedbackbogen

Wenn Sie die nachfolgenden Fragen beantworten und die Auswertungsregeln beachten, können Sie sich ein Bild darüber verschaffen, wie gut Ihre Beobachtungsleistung beschaffen war.

Film B: Mord

1. Wo befand sich ein auffälliges Merkmal im Gesicht des Täters und worum handelte es sich dabei?
2. Was fiel der geschädigten Person aus dem Mund?
3. Wie lautete die Schrift auf der Innenseite der Beifahrertür?
4. Was fiel am Rand der Straße, auf der die geschädigte Person überwältigt wurde, auf?
5. Welche Spur hinterließ die geschädigte Person an der Scheibe, als ihr Kopf an die selbe gedrückt wurde?
6. Welches Mordinstrument fand der Täter am Flußufer?
7. Woher nahm der Täter die Eisenstange?
8. Aus welchem Material, war die Decke, durch die das Blut sickerte, bevor die geschädigte Person endgültig starb?
9. Welchen entstellten Körperteil sah man in Großeinstellung, woraufhin der Täter annahm, die geschädigte Person sei tot?
10. Aus welchem Material war der Gegenstand der vom Rückspiegel baumelte und worum handelte es sich?
11. Was fiel an der Kleidung des Täters sonst noch auf?
12. War die Hand der geschädigten Person am Ende sauber oder schmutzig?

Das Bewertungsschema können Sie sich auf Wunsch vom VL aushändigen lassen.

Bewertungsschema

Gemessen wird Ihre Leistung daran, wie gut Sie periphere, möglicherweise zeugenrelevante Informationen beobachtet und behalten haben. Dafür erhalten Sie jeweils Pluspunkte. Für richtige Antworten zu reißerischen, gewaltbezogenen Details geben Sie sich jedoch Minuspunkte.

Geben Sie sich jedoch zunächst einen Punkt für jede vollständig und korrekt beantwortete Frage! Die Vergabe von + und – erfolgt anschließend.

Richtige Antworten:	richtig?
1. Auf der Wange	<input type="radio"/>
2. Dem Gebiss fehlte ein Zahn	<input type="radio"/>
3. Weiße Schrift	<input type="radio"/>
4. Ja	<input type="radio"/>
5. Geschädigte Person wurde dagegen gedrückt und hinterließ eine Speichelspur	<input type="radio"/>
6. Sechsmal	<input type="radio"/>
7. schwarz, braun, dunkelgrün	<input type="radio"/>
8. mehrere Karos, dunkelbraun und beige	<input type="radio"/>
9. Blick auf entstellten Männerkopf	<input type="radio"/>
10. Puppenkopf aus Gummi	<input type="radio"/>
11. Jeansjacke	<input type="radio"/>
12. Auf die schmutzige Hand.	<input type="radio"/>

Nun ändern Sie die Vorzeichen der Punkte folgendermaßen. Geben Sie sich für die Fragen 1, 3, 4, 7, 10, 11 jeweils einen +Punkt, wenn Sie sie vollständig und richtig beantwortet haben!!! Geben Sie sich für die Fragen 2, 5, 6, 8, 9, 12 jeweils einen -Punkt, wenn Sie sie vollständig und richtig beantwortet haben!!!

Rechnen Sie nun Ihre Minus- und Ihre Pluspunkte zusammen!

Gesamtpunktzahl <0 = Schlechte Zeugenleistung **gesamt =**

Gesamtpunktzahl >0 = Gute Zeugenleistung

Bewertungsschema

Gemessen wird Ihre Leistung daran, wie gut Sie periphere, möglicherweise zeugenrelevante Informationen beobachtet und behalten haben. Dafür erhalten Sie jeweils Pluspunkte. Für richtige Antworten zu reißerischen, gewaltbezogenen Details geben Sie sich jedoch Minuspunkte.

Geben Sie sich jedoch zunächst einen Punkt für jede vollständig und korrekt beantwortete Frage! Die Vergabe von + und – erfolgt anschließend.

Richtige Antworten:	richtig?
1. Von der Person aus links, etwas unterhalb des Auges.	<input type="radio"/>
2. Gebiss	<input type="radio"/>
3. Zapnij	<input type="radio"/>
4. Zahlreiche Strommasten	<input type="radio"/>
5. Speichelspur	<input type="radio"/>
6. Felsbrocken oder Stein	<input type="radio"/>
7. Aus einer schwarzen Ledertasche	<input type="radio"/>
8. Wollene Decke	<input type="radio"/>
9. Entstellter Männerkopf	<input type="radio"/>
10. Puppenkopf aus Gummi	<input type="radio"/>
11. Jeansjacke	<input type="radio"/>
12. Schmutzig	<input type="radio"/>

Nun ändern Sie die Vorzeichen der Punkte folgendermaßen. Geben Sie sich für die Fragen 1, 3, 4, 7, 10, 11 jeweils einen +Punkt, wenn Sie sie vollständig und richtig beantwortet haben!!! Geben Sie sich für die Fragen 2, 5, 6, 8, 9, 12 jeweils einen - Punkt, wenn Sie sie vollständig und richtig beantwortet haben!!!

Rechnen Sie nun Ihre Minus- und Ihre Pluspunkte zusammen!

Gesamtpunktzahl <0 = Schlechte Zeugenleistung **gesamt =**

Gesamtpunktzahl >0 = Gute Zeugenleistung

Feedbackbogen (Vergewaltigungsszene)

1

Feedbackbogen

Wenn Sie die nachfolgenden Fragen beantworten und die Auswertungsregeln beachten, können Sie sich ein Bild darüber verschaffen, wie gut Ihre Beobachtungsleistung beschaffen war.

Film A: Sexuelle Nötigung

1. Trug einer der Täter einen schwarzen, langen Mantel?
2. Trug einer der Täter einen Ohrstecker?
3. Welche Farbe hatten die Streifen auf dem kurzen Mini der geschädigten Person?
4. Was trug die geschädigte Person Auffälliges?
5. Wo wurde die geschädigte Person überwältigt?
6. Was fiel im Kopfbereich der geschädigten Person auf?
7. Was taten die Täter nachdem sie die geschädigte Person auf dem Weg zurückgelassen hatten?
8. Wo wurde die geschädigte Person durch einen Biß verletzt?
9. Wohin trat die geschädigte Person, bevor sie fliehen konnte?
10. Wie nannte einer der Täter die geschädigte Person?
11. Den Tätern war die geschädigte Person nicht gänzlich unbekannt. Fällt Ihnen eine weitere Bezeichnung der Täter für die geschädigte Person ein, die einen Hinweis auf ihre Tätigkeit geben könnte?
12. Sind Ihnen an einem der Täter schlechte Zähne aufgefallen?

Das Bewertungsschema können Sie sich vom VL aushändigen lassen.

Feedbackbogen

Wenn Sie die nachfolgenden Fragen beantworten und die Auswertungsregeln beachten, können Sie sich ein Bild darüber verschaffen, wie gut Ihre Beobachtungsleistung beschaffen war.

Film B: Sexuelle Nötigung

1. Womit waren die Täter bekleidet?
2. Welchen Schmuck trug einer der Täter?
3. Was trug die geschädigte Person unter dem schwarzen Lackmantel?
4. Was trug die geschädigte Person um den Bauch?
5. Worüber sah man die geschädigte Person mit blanker Brust hängen?
6. Welche Haarfarbe hatte das Opfer?
7. Was taten die Täter anschließend, was bei den Ermittlungen weiterhelfen könnte?
8. Wohin biß einer der Täter die geschädigte Person nachdem er ihr Oberteil geöffnet hatte?
9. Was tat die geschädigte Person, um sich zu befreien?
10. Wie nannte einer der Täter die geschädigte Person, nachdem er sagte, als er sie am Straßenrand erblickte?
11. Wie nannten die Täter die geschädigte Person außerdem?
12. Welches auffällige Merkmal fällt Ihnen zum Aussehen eines der Täter außerdem ein?

Das Bewertungsschema können Sie sich auf Wunsch vom VL aushändigen lassen.

Bewertungsschema

Gemessen wird Ihre Leistung daran, wie gut Sie periphere, möglicherweise zeugenrelevante Informationen beobachtet und behalten haben. Dafür erhalten Sie jeweils Pluspunkte. Für richtige Antworten zu reißerischen, erotischen Details geben Sie sich jedoch Minuspunkte.

Bitte beachten, es geht hier nicht um „richtige“ Antworten, sondern darum, die Aufmerksamkeitslenkung abzuschätzen. Geben Sie sich daher einen Punkt für jede vollständig korrekt beantwortete Frage! Die Vergabe von + und – erfolgt anschließend.

Richtige Antworten:	richtig?
1. Ja	<input type="radio"/>
2. Ja.	<input type="radio"/>
3. gelb-oranger Minirock mit dünnen grünen und schwarzen Streifen	<input type="radio"/>
4. goldene Bauchkette	<input type="radio"/>
5. über dem Zaun hängend	<input type="radio"/>
6. blonde Haare	<input type="radio"/>
7. in ein Auto steigen	<input type="radio"/>
8. Biß in Brust	<input type="radio"/>
9. Tritt in Hoden	<input type="radio"/>
10. Geile Pussy!	<input type="radio"/>
11. Dumme Nutte!	<input type="radio"/>
12. Ja.	<input type="radio"/>

Nun ändern Sie die Vorzeichen der Punkte folgendermaßen. Geben Sie sich für die Fragen 1, 2, 7, 9, 11, 12 jeweils einen +Punkt, wenn Sie sie vollständig und richtig beantwortet haben!!! Geben Sie sich für die Fragen 3, 4, 5, 6, 8, 10 jeweils einen -Punkt, wenn Sie sie vollständig und richtig beantwortet haben!!!

Rechnen Sie nun Ihre Minus- und Ihre Pluspunkte zusammen!

Gesamtpunktzahl <0 = Schlechte Zeugenleistung **gesamt:**

Gesamtpunktzahl >0 = Gute Zeugenleistung

Bewertungsschema

Gemessen wird Ihre Leistung daran, wie gut Sie periphere, möglicherweise zeugenrelevante Informationen beobachtet und behalten haben. Dafür erhalten Sie jeweils Pluspunkte. Für richtige Antworten zu reißerischen, erotischen Details geben Sie sich jedoch Minuspunkte.

Bitte beachten, es geht hier nicht um „richtige“ Antworten, sondern darum, die Aufmerksamkeitslenkung abzuschätzen. Geben Sie sich daher einen Punkt für jede vollständig korrekt beantwortete Frage! Die Vergabe von + und – erfolgt anschließend.

Richtige Antworten:	richtig?
1. Dunkles Jackett, dunkler langer Mantel, dunkles Hemd, gestreiftes Hemd, dunkle Hosen, alte Turnschuhe	<input type="radio"/>
2. Ohrstecker	<input type="radio"/>
3. Kurzer Mini	<input type="radio"/>
4. Bauchkette	<input type="radio"/>
5. Über`m Zaun	<input type="radio"/>
6. Blond	<input type="radio"/>
7. Sie stiegen in ein Auto (dunkler 7 er BMW)	<input type="radio"/>
8. Biß in die Brust	<input type="radio"/>
9. Sie trat einem der Täter in die Hoden	<input type="radio"/>
10. Geile Pussy!	<input type="radio"/>
11. Dumme Nutte!	<input type="radio"/>
12. Schlechte Zähne!	<input type="radio"/>

Nun ändern Sie die Vorzeichen der Punkte folgendermaßen. Geben Sie sich für die Fragen 1, 2, 7, 9, 11, 12 jeweils einen +Punkt, wenn Sie sie vollständig und richtig beantwortet haben!!! Geben Sie sich für die Fragen 3, 4, 5, 6, 8, 10 jeweils einen - Punkt, wenn Sie sie vollständig und richtig beantwortet haben!!!

Rechnen Sie nun Ihre Minus- und Ihre Pluspunkte zusammen!

Gesamtpunktzahl <0 = Schlechte Zeugenleistung **gesamt:**

Gesamtpunktzahl >0 = Gute Zeugenleistung

Anhang D

Itemliste (Mordszene)

probe Testreiz	target Zielreiz	Irrelevante Reize
Gebiss	Warze	Falten Ekzem Rötung Brille
Spur an Scheibe	Schrift auf Türe	Kreis am Boden Riss in Mauer Sprung in Teller Fleck auf Anzug
Felsbrocken	Strommasten	Meeressand Blumenbeet Wegweiser Heuschober
Wollene Decke	Lederne Tasche	Gläserne Schuhe Marmorner Boden Hözerne Hütte Silbernes Messer
Männerkopf	Puppenkopf	Teddykopf Bubikopf Flaschenkopf Gipskopf
Schmutzige Hand	Löchriges Shirt	Stinkiger Atem Schnoddrige Nase Schmierige Hose Filzige Haare

Itemliste (Vergewaltigungsszene)

probe Testreiz	target Zielreiz	Irrelevante Reize
Kurzer Mini	Langer Mantel	enge Hotpants lange Strümpfe knappes Höschen Push-up BH
Bauchkette	Ohrstecker	Fußkette Intimschmuck Bauchpiercing Sonnenhut
über Zaun hängen	in Auto steigen	an Brücke lehnen im Wasser liegen unter Deck gehen durch Scheibe fliegen
Blonde Haare	Schlechte Zähne	Volle Wangen Schwarze Locken Brauner Rücken Kleiner Busen
Biß in Busen	Tritt in Hoden	Kniff in Hintern Hieb in Gesicht Stich in Schenkel Kick in Gesicht
Geile Pussi	Dumme Nutte	Schnieke Lady Scharfe Ische Super Schnitte Heißer Feger

Anhang E

Bootstrappingverfahren 1

Um die Annahmen 3a) $((S \sim U) \neq N$ für Täter) und 4a) $((S \sim N) \neq U$ für Nichttäter) für den Einzelfall zu überprüfen wurde als Maß für den Zusammenhang zwischen den entsprechenden Kurven die Korrelation verwendet. Auf diese Weise sollten korrelative Zusammenhänge zwischen tatspezifischen (S) und tatunspezifischen (U) mit denen zwischen tatspezifischen und neutralen Reizen (N) verglichen werden. Durch die Mittelung aller EEG-Kurven zu einer EKP-Kurve pro Bedingung gehen alle Informationen über die Verteilung dieser EEG-Stichprobe verloren. Die Mittelung ist aber notwendig, um das EKP-Signal aus dem Spontan-EEG herauszufiltern. Darüber hinaus verletzt die Verteilung der Amplitudenmaße der zu einer EKP-Kurve gemittelten EEG-Kurve die Normalverteilungsannahme. Alle parametrischen Schätzer, die von den Korrelationen abgeleitet würden, wären damit nicht valide. Das sogenannte „Bootstrapping“ (Efron, 1979) stellt einen Ausweg aus diesem Dilemma dar, da man mit Hilfe des Verfahrens eine Verteilung generieren kann.

Das Vorgehen lässt sich in vier Schritten darstellen:

1. *Mitteln der neutralen Items*: Für jeden Probanden resultierten insgesamt 432 Epochen, davon je 72 Epochen bezüglich tatspezifischer und 72 Epochen bezüglich tatunspezifischer Items, sowie 288 Epochen vom neutralen Typ. Jeweils vier Epochen des neutralen Typs wurden zu einer gemittelt, so dass von jedem Itemtyp schließlich 72 Epochen vorhanden waren.

2. *Ziehen und Mitteln der Kurven*: Aus jedem Pool (tatspezifisch, tatunspezifisch, neutral) wurden pro Iteration 72 Epochen gezogen, wobei nach jedem Ziehen die ausgewählte Epoche zurück in den Pool gegeben wurde. Die 72 Epochen wurden jeweils gemittelt, so dass pro Iteration drei gemittelte Kurven, eine für jeden Itemtyp, gewonnen wurden.

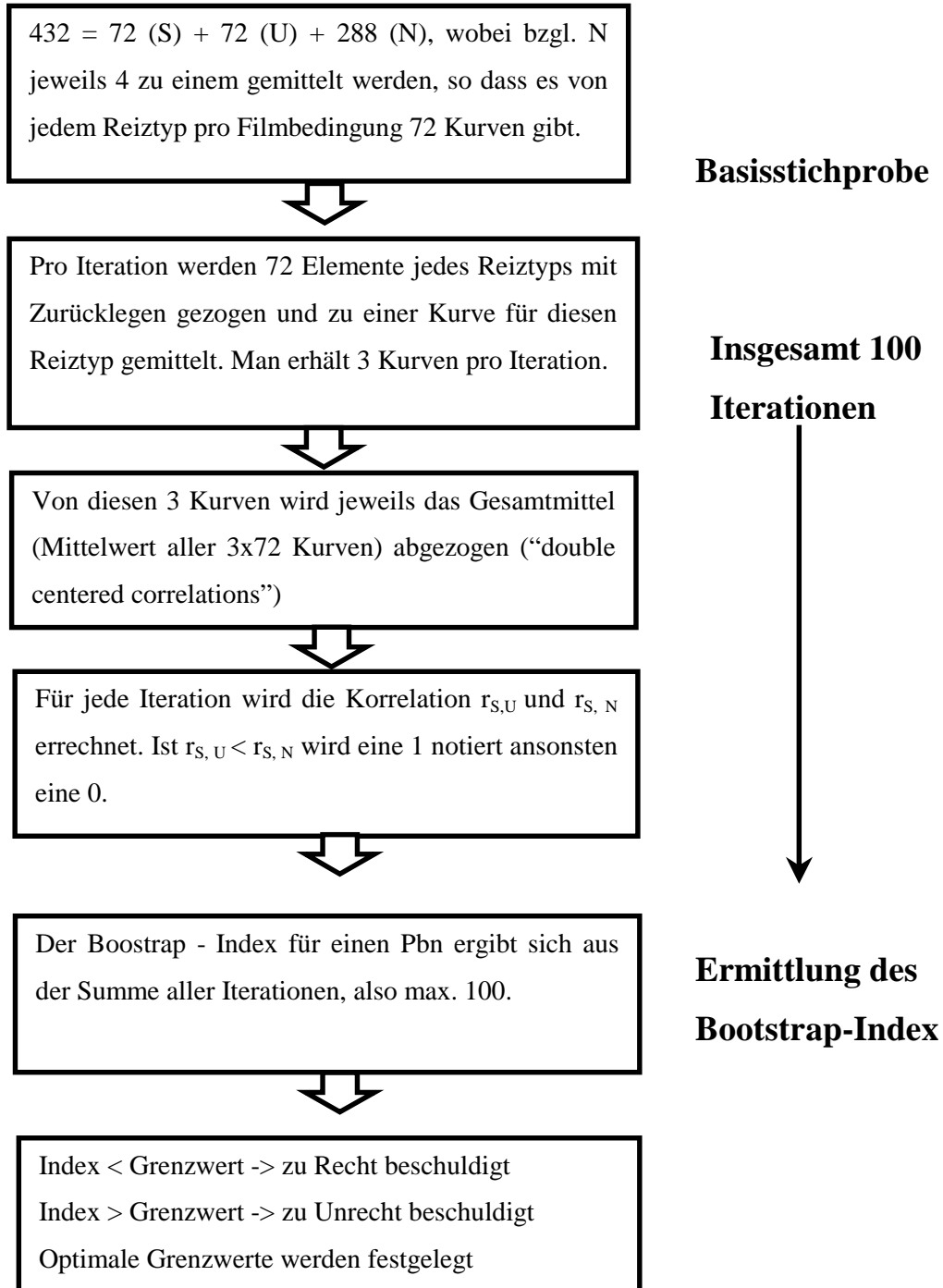
3. *Bestimmung der Korrelationen*: Es wurden jeweils die Korrelation zwischen den Kurven bezüglich S und U, sowie die Korrelation zwischen den Kurven bezüglich S und N berechnet. Korreliert wurden jeweils die zum gleichen Zeitpunkt gemessenen Amplitudenwerte. Es wurden *double-centred*¹ Korrelationen verwendet. Das heißt, dass

¹ Double-centred correlations: Korrelationen, die an zweifach um einen Mittelwert zentrierten Daten gewonnen werden. Dies geschieht zu Standardisierungszwecken.

vor der Berechnung der Kreuzkorrelationen eine über alle Bedingungen gemittelte Kurve von jeder der drei gemittelten Epochen abgezogen wurde.

4: Vergleich der Korrelationen: Indem man diesen Vorgang häufig wiederholt, generiert man eine Verteilung, die es ermöglicht, verschiedene Parameter als Schätzer für die Populationskennwerte heranzuziehen. Ein Vergleich von zwischen S und U ermittelten Korrelationen, mit denen zwischen S und N wurde verwendet, um den Bootstrap-Index zu ermitteln. Dieser Index kann Werte zwischen 0 und 100 annehmen, da die Ziehungen und Vergleiche hundertmal durchgeführt werden. Übersteigt die S x U-Korrelation die S x N-Korrelation bei einer Iteration wird eine 0 addiert, andernfalls wird eine 1 addiert. Ein niedriger Index spiegelt also einen stärkeren Zusammenhang zwischen S und U wieder, während ein hoher Index einen stärkeren Zusammenhang zwischen S und N wiedergibt. Mit anderen Worten gibt der Bootstrapindex wieder, in wie viel Prozent der Fälle die S x U-Korrelation nicht größer ist als die S x N-Korrelation.

Übersicht über die Auswertungsstrategie mit Hilfe des korrelativen Bootstrapverfahrens.



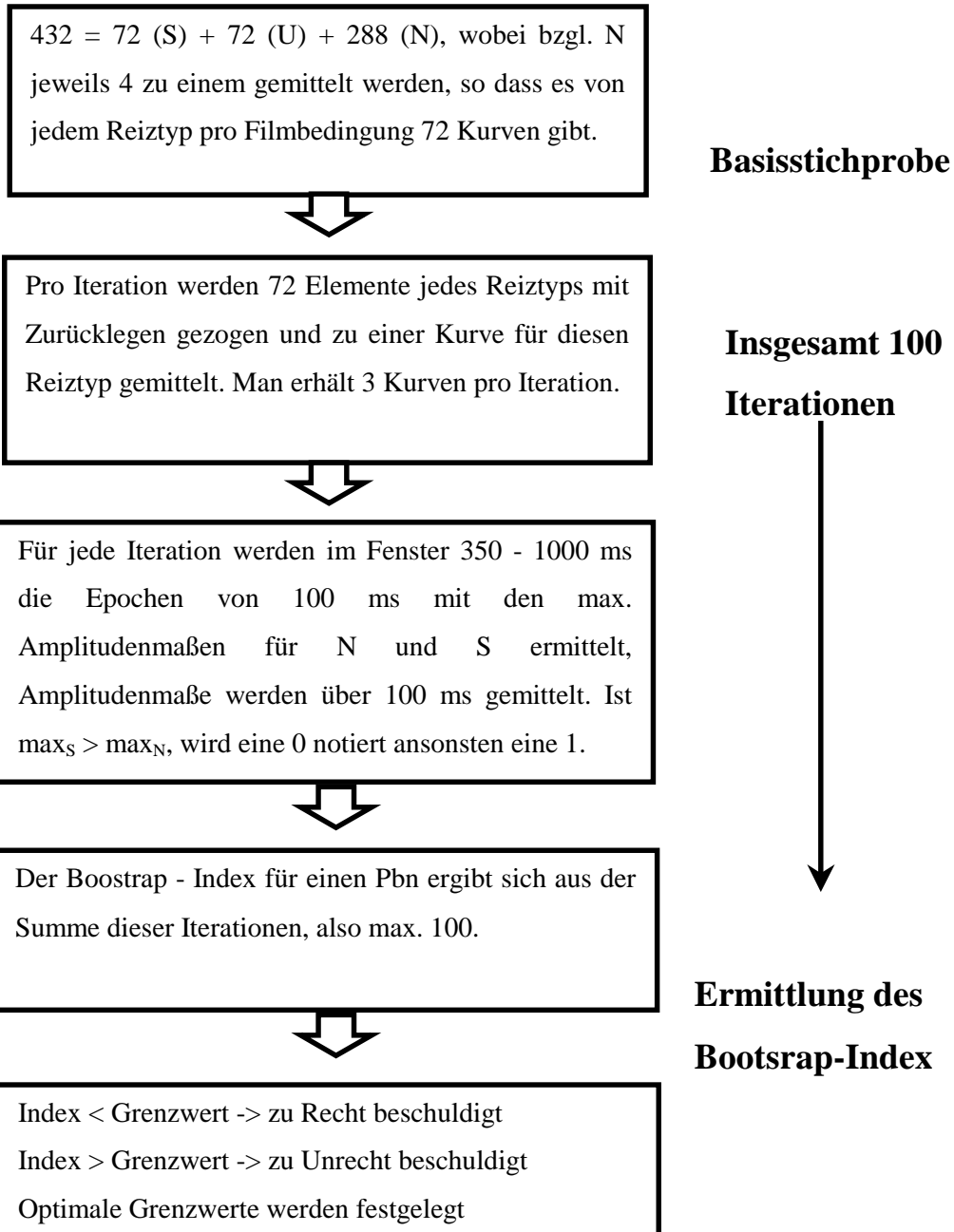
Bootstrapverfahren 2

Um die Annahmen 3b) ($S > N$ für Täter) und 4b) ($S \leq N$ für Nichttäter) für den Einzelfall zu überprüfen wurden jeweils die maximalen Amplituden zwischen tatspezifischen und neutralen Items verglichen. Führt man diese Amplitudenvergleiche für die maximalen Amplituden der EKP-Kurven durch, lassen sich wiederum keine validen parametrischen Schätzer bezüglich dieser Vergleiche für die Population ableiten, da die maximalen Amplituden der EEG-Epochen aufgrund des Rauschens des Spontan-EEGs wiederum nicht bestimmt werden können und damit auch deren Verteilung unbekannt ist. Auch hier ermöglicht es das Bootstrapverfahren, eine Verteilung zu generieren, die eine valide Schätzung auf den Populationswert zulässt. Schritt 1 und 2 des Verfahrens entsprechen dem Vorgehen des Bootstrapverfahrens 1.

3. *Amplitudenvergleich der gemittelten Kurven:* Zunächst wurde eine Baselinekorrektur vorgenommen. Das heißt, es wurde der Mittelwert der Amplitudenwerte jeder EKP-Kurve in dem Bereich 100 ms vor dem Reiz berechnet und von der jeweiligen Kurve abgezogen. Auf diese Weise wurden Unterschiede, die nicht aus der Reaktion auf den Reiz resultierten, vom Amplitudenvergleich zwischen den Items ausgeschlossen. Nach der Baselinekorrektur wurde bei jeder Iteration aus dem Zeitfenster von 350 – 1000 ms das Maximum bestimmt (Baseline-to-Peak-Methode) und in einem Zeitfenster von 100 ms um diesen Datenpunkt gemittelt. Die so entstandenen gemittelten Maxima wurden verglichen.

4. *Berechnung des Bootstrapindex pro Person:* Der Bootstrapindex ergab sich aus dem Vergleich der Kurvenmaxima bezüglich N und S über 100 Iterationen. Bei jeder Iteration, bei der das Maximum bezüglich S das bezüglich N übertraf, wurde eine 0 addiert, ansonsten eine 1. Der Bootstrapindex konnte also zwischen 0 und 100 variieren und gibt an, in wie viel Prozent der Fälle $S < N$ oder $S = N$ gilt. Das heißt, ein niedriger Index spricht für die Klassifikation eines Probanden als Täter und ein hoher für eine Klassifikation als Nichttäter.

Übersicht über das Vorgehen des Amplitudenvergleichsverfahrens



Anhang F

Erinnerte Reaktionen der Probanden auf den Vorwurf

Pb2: „Ich hab halt auf das Falsche geguckt. Bei dem Vergewaltigungsfilm hab ich das nicht gedacht. Bei dem anderen Film hab ich ja schon gesagt, dass mich das gewundert hat, dass ich da auf so Kleinigkeiten achten sollte.“ – Interviewer (I): „Was hatte ich behauptet?“ – Pb: „Weil ich so Brutalität gut finde und so.“ – I: „Hast Du angenommen, dass das eine Rolle spielt.“ – Pb: „Ja schon, wenn man objektiv gucken würde, würden einem vielleicht andere Sachen auffallen.“ – I: „Hast Du das dann angenommen, dass Du deswegen schlecht warst, weil Du so auf brutale Sachen stehst?“ – Pb: „Nö, eher dass das normal ist, dass man auf so was achtet.“

Pb6: „Ich habe gedacht „Ha, das wollen wir doch mal sehen.“ Oder so etwas in der Richtung. Ich hab mich ja von Anfang an bemüht, mir Details zu merken. Also, das ist schon. Es ist schwer zu sagen nach einem Jahr, aber wahrscheinlich – herausgefordert. Ich kann nicht ausschließen, dass ich von den Regeln abgewichen bin, aber dass ich mich dann wieder zurückgeholt habe. Natürliche Reaktion, dass ich dann einen Moment lang nicht auf Details geachtet habe, aber dann sofort wieder, quasi, „Reiß Dich zusammen!“ – Ich halte solche Reaktionen aber für normal, also nicht außergewöhnlich.“

Pb8: „Ich fand es interessant, ob das tatsächlich so ist bei mir. Ich dachte das eigentlich nicht, dass das so sein könnte, dass ich mich da... – Ich fand die Filme abstoßend. - Ich hab mir Gedanken gemacht, darüber, dass jemand von mir denken könnte, dass ich mich von .. Es wäre mir auch unangenehm gewesen, hätte es zu so einem Ergebnis geführt... Aber ich hätte das nicht geglaubt.“

Pb17: keine Erinnerung

Pb21: „Nein, ich hab das wirklich distanziert gemacht und ich glaube, dass das bei den Filmen und bei dem EEG so war – vielleicht, dass das mit dem Involviertsein bei dem Vergewaltigungsfilm weniger gewesen ist, weil die Gewaltszene abscheulicher gewesen ist, fand ich zumindest.“

Pb32: „Höchstens dass ich mir im Unterbewusstsein vielleicht die Frau genauer angeschaut habe, aber ich glaub es eigentlich nicht. Aber möglich ist das schon, oder

dass ich mich versucht habe auf zu viele Sachen zu konzentrieren und dann wichtige Dinge vergessen habe.“

Pb34: „Man will ja dann doch. Ich mein, das ist immer das Problem, wenn man so Sachen, die man denkt, einfach nicht sagt, aber wenn man das jetzt abgeleitet bekommt, dann ist das natürlich blöd. Man versucht dann so zu wirken, als ob das überhaupt keinen Einfluss hat. Aber unbewusst oder nicht unbewusst wird man durch diese Umstände dann doch beeinflusst. – Ich hab versucht mich gut zu erinnern. War bei beiden Filmen motiviert, die Aufgabe gut zu machen. – Sexualfilm, hab ich gedacht, war ja klar und so ein Mist, weil das wollte ich ja gerade nicht, dass das rauskommt, weil bei Gewalt, dass da rauskommt, dass ich Freude an Gewalt habe, da habe ich ja keine Bedenken.“

Pb38: „Beim Vergewaltigungsfilm: Ich würde mal sagen, hab ich mir wahrscheinlich ja Gedanken gemacht, weil das ja mit dem Vorwurf ein wenig in den Intimbereich geht, aber ich will hier nichts beschwören. Beim Mordfilm könnte es sein, dass man denkt, dass man zu blutrünstig ist, und dass ich mich de facto tatsächlich auch von reißerischen Details habe blenden lassen. Den Fahrradfahrer habe ich ... ist mir im Film selber nicht aufgefallen, da hab ich mehr auf den Mord geachtet. Eine gute Zeugenleistung ist natürlich nur eine, die auch Randdetails mitberücksichtigt. Insofern war das natürlich, in Anführungszeichen, nicht so ganz korrekt von mir.“

Pb39: „Das hab ich nicht so gesehen. Ich hab ja vorher gesagt bekommen, dass ich da nicht drauf zu achten hab. Und vielleicht im Gegensatz zu meinen sexuellen Phantasien und meinen Mordgelüsten, die ich so in mir habe, dann doch trotzdem auf Kleinigkeiten geachtet habe. Ich hatte den Eindruck, dass ich mir schon Mühe gegeben hatte auch die anderen Sachen wahrzunehmen. Ich war da ein bisschen überrascht davon. – differenziert nach Filmen? – unterbewusst hat man mehr an den Vergewaltigungsfilm gedacht, weil man das nicht so gerne wahrhaben will und dass einer jetzt von einem sagt, beispielsweise „typisch Mann, glotzt da auf die nackte Frau und sieht nix mehr da drumherum“, vielleicht ist das so gekommen, zumindest unterbewusst.“

Pb40: „Ja, ich hab gedacht, das war schon unangenehm wegen dem Sexfilm, aber ich schaff das schon. Eigentlich hab ich auch gedacht, ich hätte das gut gemacht

bei den Filmen, aber dann... Ich hab ja schon gesagt, dass mich das so ein bisschen gewundert hat mit den Ergebnissen nach dem Film.“

Pb41: „Ich muss sagen, dass es mich nachher gar nicht mehr so gepackt hat, diese ganze Überlegerei. Sondern ich hab halt versucht, wie bei einer Schulaufgabe es gut zu machen. Listen gelernt. Hab mich darauf konzentriert. Es kam so etwas wie Stolz auf, weil ich nicht so viele Fehler gemacht hab. – Versuch, konzentrier dich nicht auf die Gewalttat, aber man konnte es nicht ganz ausschließen. – Die Schmach versucht zu vergessen. Auf die Listen konzentriert. – Ich war vom Grundtenor beim ersten Mal schlecht. Ich hab mich schon geärgert, dass ich bei dem Mordfilm doch schlecht war, bei dem Sexfilm war ich eigentlich schlecht, da war ganz leichter Ärger da, dass ich als Zeuge versagt hatte.“