

Zweitveröffentlichung



Fegert, Jörg M.; Gerke, Jelena; Kliemann, Andrea; u. a.

Der Bundesgerichtshof (BGH) ging bei der Bewertung der Treffsicherheit der Glaubhaftigkeitsbegutachtung von falschen Prämissen aus

Datum der Zweitveröffentlichung: 07.04.2026

Verlagsversion (Version of Record), Zeitschriftenartikel

Persistenter Identifikator: urn:nbn:de:bvb:473-irb-114556x

Erstveröffentlichung

Fegert, Jörg M.; Gerke, Jelena; Kliemann, Andrea; u. a. (2025): Der Bundesgerichtshof (BGH) ging bei der Bewertung der Treffsicherheit der Glaubhaftigkeitsbegutachtung von falschen Prämissen aus, in: Forensische Psychiatrie, Psychologie, Kriminologie, Berlin ; Heidelberg: Springer, Jg. 19, Nr. 3, S. 216–227, doi: 10.1007/s11757-025-00887-2.

Rechtehinweis

Dieses Werk ist durch das Urheberrecht und/oder die Angabe einer Lizenz geschützt. Es steht Ihnen frei, dieses Werk auf jede Art und Weise zu nutzen, die durch die für Sie geltende Gesetzgebung zum Urheberrecht und/oder durch die Lizenz erlaubt ist. Für andere Verwendungszwecke müssen Sie die Erlaubnis der Rechteinhaberinnen und Rechteinhaber einholen.

Für dieses Dokument gilt eine Creative-Commons-Lizenz.



Die Lizenzinformationen sind online verfügbar:

<https://creativecommons.org/licenses/by/4.0/legalcode>



Der Bundesgerichtshof (BGH) ging bei der Bewertung der Treffsicherheit der Glaubhaftigkeitsbegutachtung von falschen Prämissen aus

Wir nutzen die Chance einer Replik auf die Kritik von Renate Volbert, Nathalie Brackmann, Elsa Gewehr, Luise Greuel, Anja Kannegießer, Andreas Mokros, Michaela Pfundmair, Kristina Suchotzki und Jonas Schemmel

Jörg M. Fegert^{1,2} · Jelena Gerke¹ · Andrea Kliemann^{2,3} · Martin Pusch⁴ · Stephan Rixen⁵ · Cedric Sachser^{1,6}

Eingegangen: 9. Mai 2025 / Angenommen: 20. Mai 2025 / Online publiziert: 10. Juli 2025
© The Author(s) 2025, korrigierte Publikation 2025

Zusammenfassung

Der Beitrag antwortet auf die Replik von Volbert et al. (2025) – im Folgenden Autor*innen bzw. Kritiker*innen genannt – zu unserer Expertise zur Methode der forensischen Glaubhaftigkeitsbegutachtung im Auftrag der Unabhängigen Beauftragten für Fragen des sexuellen Kindesmissbrauchs. Im Zentrum steht eine methodenkritische Auseinandersetzung mit dem Konzept der „Nullhypothese“, das insbesondere durch das BGH-Urteil vom 30.07.1999 geprägt wurde. Dieses Urteil hat den damaligen Stand der aussagepsychologischen Diskussion in nichttragfähiger Weise rezipiert. Diese Fehlrezeption stand einer differenzierten Debatte über Grund und Grenzen der sog. Nullhypothese in gerichtlichen Verfahren lange im Wege. Die mit der Expertise angestoßene Debatte trägt nun zu einer wissenschaftlichen Auseinandersetzung bei. Ausgangslage der Expertise war, dass Betroffene sexualisierter Gewalt in Kindheit und Jugend das Vorgehen auf Basis der Nullhypothese als Misstrauen empfinden. Auch die Kritiker*innen stimmen zu, dass der Begriff falsche Erwartungen wecken und belastend sein kann. Die juristisch laienhafte Anwendung der Nullhypothese in der instanzgerichtlichen Rechtsprechung sehen wir kritisch und es liegt eine Aufgabe dieses fehlleitenden und kontraproduktiven Begriffs nahe.

Übereinstimmung zwischen den Kritiker*innen und uns besteht darin, dass es sich bei der Begutachtung um ein qualitatives, einzelfallbezogenes Verfahren handelt. Die Autor*innen kritisieren, dass wichtige Forschungserkenntnisse in unserer Expertise fehlten. Diese werden jedoch nun hier aufgegriffen, da aktuelle Metaanalysen teilweise zum Zeitpunkt der Erarbeitung der Expertise noch nicht veröffentlicht waren. Zusammenfassend sprechen wir uns für eine methodisch präzisere, empirisch fundierte und kontextsensitiv angepasste Begutachtungspraxis aus. Dabei sehen wir viele Aussagen in der Replik von Volbert et al. als Chance, da bei näherer Betrachtung wichtige Aspekte übereinstimmend kritisch bewertet werden.

Schlüsselwörter Nullhypothese · Sexualisierte Gewalt · Rechtskontexte · Strafrecht · Soziales Entschädigungsrecht · Kindeswohlgefährdung

✉ Prof. Dr. Jörg M. Fegert
joerg.fegert@uniklinik-ulm.de

¹ Klinik für Kinder- und Jugendpsychiatrie, Psychosomatik und Psychotherapie, Universitätsklinikum Ulm, Ulm, Deutschland

² Kompetenzzentrum Kinderschutz in der Medizin Baden-Württemberg, Ulm, Deutschland

³ Fakultät Soziale Arbeit, Universität Vechta, Vechta, Deutschland

⁴ Rechtsanwälte Westpfahl Spilker Wastl, München, Deutschland

⁵ Institut für Staatsrecht, Forschungsstelle für das Recht des Gesundheitswesens, Universität zu Köln, Köln, Deutschland

⁶ Institut für Psychologie, Lehrstuhl Klinische Kinder- und Jugendlichenpsychologie Otto-Friedrich-Universität Bamberg, Bamberg, Deutschland

The evaluation by the German Federal Court of Justice on the accuracy of the credibility assessment is based on false premises

We use the chance to respond to the criticism of Renate Volbert, Nathalie Brackmann, Elsa Gewehr, Luise Greuel, Anja Kannegeßer, Andreas Mokros, Michaela Pfundmair, Kristina Suchotzki and Jonas Schemmel

Abstract

This article responds to the reply by Volbert et al. (2025),—hereinafter referred to as the authors or critics—to our expert report on forensic credibility assessment, commissioned by the Independent Commissioner for Child Sexual Abuse Issues. Central to the discussion is a methodological critique of the so-called null hypothesis, particularly shaped and endorsed by the German Federal Court of Justice ruling of 30 July 1999. This ruling adopted the then-prevailing psychological discourse on witness credibility in a way no longer considered tenable. This misreception, legitimized by the highest court, hindered a nuanced debate on the rationale and limits of the null hypothesis in judicial proceedings for years. The debate initiated by our expert report contributes to a more scientific discussion. It began with the observation that survivors of childhood and adolescent sexual violence perceive the null hypothesis as a sign of fundamental mistrust. The critics acknowledge that the concept can raise false expectations and burden those affected. We view the unreflected use in lower court rulings critically and advocate for abandoning this misleading term.

There is consensus between the critics and us that credibility assessment is a qualitative, case-specific procedure. The authors criticize that important research findings were missing from our report. We address this here, noting that some relevant meta-analyses were unpublished at the time. In conclusion, we advocate for a methodologically more precise, empirically grounded, and context-sensitive assessment practice. Many of the points raised by Volbert et al. provide opportunities for constructive reflection and shared critical evaluation.

Keywords Null hypothesis · Sexualized violence · Legal contexts · Criminal law · Social compensation law · Endangerment of children's welfare

Kritik an der Nullhypothese, wissenschaftstheoretische Missverständnisse in Bezug auf das methodische Vorgehen

Zu Recht beschreiben die Autor*innen, dass sich unsere prinzipielle Kritik gegen das Grundsatzurteil des BGH (Bundesgerichtshof) zu wissenschaftlichen Anforderungen an aussagepsychologische Begutachtungen vom 30.07.1999 (BGHSt 45.164) wendet. Genau damit sollte sich unsere Expertise im Auftrag der Unabhängigen Beauftragten für Fragen des sexuellen Kindesmissbrauchs (UBSKM) auseinandersetzen, da dieses Vorgehen auf der Grundlage der „Nullhypothese“ von Betroffenen häufig als generelles Misstrauen erlebt wird. Diese negative Konnotation wird auch im vorliegenden Papier (Volbert et al. 2025, S. 16) insbesondere mit Bezug auf Steller und Volbert (2000) eingeräumt. Um Missverständnissen vorzubeugen, habe der BGH in einem Urteil vom 30.05.2000 schon darauf hingewiesen, dass es sich bei den von Sachverständigen zu generierenden Untersuchungshypothesen um rein gedankliche Prüfschritte handelt (Lösel und Bender 2001, S. 45 ff.). Darüber hinaus schreiben die Autor*innen, dass der Begriff der „Nullhypothese“ auch „in der aussagepsychologischen Literatur wegen der mangelnden Passung zwischen dem tatsächlichen qualitativen Vorgehen und dem aus der Prüfstatistik stammenden Begriff teilweise deutlich kritisiert worden sei (z. B. Greuel 2009)“. Die Autor*innen

halten diesen Begriff für die Erläuterung gegenüber Proband*innen und für die systematische Anwendung des hypothesengeleiteten Ansatzes für verzichtbar und verweisen dabei auf Steller (2020) und Greuel (2022). Wir deuten es als positiven Effekt unserer Expertise, dass von führenden Vertreter*innen der Glaubhaftigkeitsbegutachtung eingeräumt wird, der Begriff wecke falsche Erwartungen und könne für Betroffene eine Belastung darstellen, weshalb er für „die Erläuterung gegenüber Proband*innen und für die systematische Anwendung des hypothesengeleiteten Ansatzes der Glaubhaftigkeitsbegutachtung“ als verzichtbar erscheine. Warum dann aber nicht generell darauf verzichten? Wir wollen ausdrücklich festhalten, dass uns mit den Autor*innen das Anliegen verbindet, zur Klärung einer missverständlichen Redeweise beizutragen und sich der sachlichen Probleme, die diese abzulehnende Redeweise verbirgt, zu vergewissern.

Ferner wollen wir aber deutlich machen, dass es unser Ziel war, auch eine zentrale Personengruppe, nämlich die Jurist*innen, zu erreichen, wenn wir darauf hinweisen, dass der individuellen Überprüfung der „Nullhypothese“ im Rahmen der lege artis durchgeführten Glaubhaftigkeitsbegutachtung nicht aufgrund mathematischer Grundgesetzmäßigkeiten im juristischen Sinne ein Indizcharakter beizumessen ist. Wir bedauern es, dass dieser Teil der Kritik in der vorliegenden Replik nicht ausführlicher diskutiert wird. Es geht letztlich darum, dass von den Disziplinen und Pro-

fessionen, die mit der Glaubhaftigkeitsbegutachtung befasst sind, eine wissenschaftliche Fachdebatte, die sich durch die höchstrichterlich approbierte Fehlrezeption aussagepsychologischer Methoden nicht behindern lässt, geführt werden muss. Nur an einer Stelle in der Einleitung zur Darstellung unserer methodischen Kritik auf S. 6 des vorliegenden Papiers (Volbert et al. 2025) wird unser Anliegen zu verdeutlichen, wieso die „Nullhypothese“ im strafrechtlichen Bereich (und leider auch in anderen Rechtsgebieten) zum aussagepsychologisch nichttragfähigen Dogma wurde, erwähnt, indem referiert wird: „Die „Nullhypothese“ basiere auf einer verzerrten Interpretation des Aggregationsprinzips (vgl. auch Fegert et al. 2024b).“ In dieser letztgenannten Publikation hatten wir zusammenfassend auf diese zentralen Missverständnisse in der Urteilsbegründung des BGH verwiesen und damit auch den Fokus unserer Kritik deutlich gemacht. Wir wollen dies hier mit direkten Zitaten aus dem im Titel genannten BGH-Urteil belegen.

Unstreitig ist das Postulat, dass sich ein Sachverständiger ausschließlich solcher Methoden zu bedienen habe, „die dem jeweils aktuellen wissenschaftlichen Kenntnisstand gerecht werden (Steller 1988, S. 16, 24). Die eingesetzten Test- und Untersuchungsverfahren müssen zudem durch die gebildeten Hypothesen indiziert, d.h. geeignet sein, zu deren Überprüfung beizutragen.“ Da es sich, wie von den Autor*innen in Übereinstimmung mit uns betont, bei der Glaubhaftigkeitsbegutachtung nicht um ein normiertes Testverfahren handelt, halten wir es für sinnvoll, den Begriff „hypothesengeleitetes, individuelles Untersuchungsverfahren“ zu verwenden. Der Kern unserer Kritik in Bezug auf die dogmatische Wirkung des Urteils betrifft die vom BGH unterstellte *indizielle Bedeutung* von Merkmalen, hinsichtlich derer vom BGH quantitativ argumentiert wird, obwohl es sich, wie die Autor*innen mehrfach zutreffend betonen, um ein individuelles, auf den Einzelfall bezogenes, qualitatives Vorgehen handelt. Wörtlich formuliert der BGH: „Zur Durchführung der Analyse der Aussagequalität sind auf der Basis der dargestellten Annahmen Merkmale zusammengestellt worden, denen indizielle Bedeutung für die Entscheidung zukommen kann, ob die Angaben der untersuchten Person auf tatsächlichem Erleben beruhen. Es handelt sich um aussageimmanente Qualitätsmerkmale (z. B. logische Konsistenz, quantitativer Detailreichtum, raum-zeitliche Verknüpfungen, Schilderungen ausgefallener Einzelheiten und psychischer Vorgänge, Entlastung des Beschuldigten, deliktspezifische Aussageelemente), deren Auftreten in einer Aussage als Hinweis auf die Glaubhaftigkeit der Angaben gilt [...]“ (BGH 1 StR 618/98, Urteil v. 30.07.1999, BGHSt 45, 164, S. 4, Randziffer 21).

Der zentrale Punkt unserer Kritik ist die Verwendung des Prinzips der Aggregation als Scheinlösung bei fehlender Trennschärfe der einzelnen Kriterien. Wörtliches Zitat des BGH: „Diese sog. Realkennzeichen können als

grundsätzlich empirisch überprüft angesehen werden. Zwar handelt es sich um Indikatoren mit jeweils für sich genommen nur geringer Validität, d.h. mit durchschnittlich nur wenig über dem Zufallsniveau liegender Bedeutung. Eine gutachterliche Schlussfolgerung kann aber eine beträchtlich höhere Aussagekraft und damit Indizwert für die Glaubhaftigkeit zu beurteilenden Angaben erlangen, wenn sie aus der Gesamtheit aller Indikatoren abgeleitet wird. Denn durch das Zusammenwirken der Indikatoren werden deren Fehleranteile insgesamt gesenkt. Diesem Umstand liegt das mathematisch und psychometrisch eingehend untersuchte Prinzip der Aggregation zugrunde (Gutachten Prof. Dr. Fiedler). Dementsprechend lagen die mit Realkennzeichen in Forschungsvorhaben erzielten Ergebnisse regelmäßig deutlich über dem Zufallsniveau. Allerdings bestanden dabei teilweise nicht unerhebliche Fehlerspannen. Inwieweit ihre Bedeutung bei Verwendung gegenüber Personen aus unterschiedlichen Altersgruppen differieren kann, ist völlig offen.“

Das Gutachten Fiedler und Schmid (1999) äußert sich ebenso wie das Gutachten von Steller und Volbert (1999) zur Aggregation von Einzelbeobachtungen so, dass sich systematisch verlässliche Anteile, die auf die Wahrheit einer Aussage hindeuten, durch statistische Unabhängigkeit der Fehleranteile gegenseitig verstärkten, während sich die Fehleranteile gegenseitig ausglich. Dadurch könne im Gesamtergebnis eine Schlussfolgerung zugunsten der Wahrheit entstehen, selbst wenn einzelne Beobachtungen gegen sie sprächen. Diese fehlerkorrigierende Wirkung der Aggregation greift der BGH auf, um einem auf Einzelitem-Ebene in seiner „Validität“ kaum über dem Zufall liegenden Verfahren Indizcharakter zuzusprechen. Was jedoch passiert, wie die Replik zutreffend hervorhebt: Überprüft wird eine Unwahrhypothese, es geht nicht um eine Aggregation, die Fehler ausgleicht. Kann z. B. die Suggestionshypothese nicht ausgeschlossen werden, folgt daraus lediglich, dass die Unwahrhypothese nicht verworfen werden kann, nicht jedoch, dass die Aussage unwahr ist. Dies haben wir mit dem Begriff „binäre Entscheidung“ charakterisiert. Eine die Fehleranfälligkeit relativierende, statistische Aggregation erfolgt somit nicht und damit wird auch die relative Sicherheit der Einschätzung vor Restzweifeln nicht gesteigert. Zu induktiv-statistischen Schlüssen kommt es unter diesen Voraussetzungen nicht.

Wichtig ist die Feststellung von Fiedler und Schmid, dass einseitiges Hypothesentesten extrem in die Irre leiten kann. Wörtlich schreiben sie im Originalgutachten auf S. 22: „Durch Einschränkung der Information auf wenige selektive Indikatoren, die einem bestimmten favorisierten Modell entsprechen, und Ignorieren anderer Indikatoren, die andere denkbare Modelle bestätigen könnten, werden unter Umständen erhebliche Fehler erzeugt. So zeigen unmittelbar mit Glaubwürdigkeit befasste Experimente (z. B. Zuckerman

et al. 1984), dass Aussagen eher für falsch gehalten werden, wenn Urteiler die Hypothese einer möglichen Lüge testen, während dieselben Aussagen eher für wahr gehalten werden, wenn die Hypothese einer wahren Äußerung fokussiert wird. In der psychologischen Forschung im allgemeinen (Jussim 1991; Koehler 1991) und der Forschung zum Hypothesentesten in Gesprächen und Interviews im besonderen wurde vielfach demonstriert, dass die Ergebnisse systematisch in Richtung auf die Ausgangshypothese verzerrt sind (Snyder 1984; Pyszczynski und Greenberg 1987; Tversky und Kahneman 1974; Zuckerman et al. 1995).“ „Einer von mehreren Gründen für diesen sogenannten ‚confirmation bias‘ (Snyder & Swann, 1978) bzw. ‚auto-verification effect‘ (Fiedler, Walther & Nickel, 1999) ist die einseitige, nichtrepräsentative Suche nach Indikatoren für die leitende Hypothese und die gleichzeitige Vernachlässigung von Indikatoren für alternative Hypothesen (...)“. Wir halten fest: Das Gutachten Fiedler und Schmid plädiert für ein induktiv statistisches Vorgehen, unter Einbeziehung möglichst vieler adäquat ausgewählter Indikatoren. Man hält den deduktiv-nomologischen Beweis, also das, was unter der „Nullhypothese“ verstanden wird, als für diese Fragestellung nicht geeignet. „Während ein deduktiv-nomologischer Beweis in erster Linie ein gut bestätigtes Gesetz erfordert und dann angezweifelt werden kann, wenn das globale Gesetz nicht gesichert ist, hängen induktiv-statistische Schlüsse entscheidend von der adäquaten Auswahl und der ausreichenden Anzahl der betrachteten Indikatoren bzw. Beobachtungen ab. Mit anderen Worten, das Problem bei einer induktiv-statistisch begründeten Beweisführung liegt in der Selektivität und ausreichenden Aggregation bzw. Kombination der elementaren Informationen, nicht im (unrealistisch) hoch angenommenen Bestätigungsgrad eines einzelnen Gesetzes.“

In der Urteilsbegründung hat der I. Strafsenat des BGH (1999) ein Amalgam aus beiden Zugängen gemacht und die Behauptung aufgestellt, das mathematisch und psychometrisch eingehend untersuchte Prinzip der Aggregation führe dazu, dass eine gutachterliche Schlussfolgerung im Rahmen der Anwendung der Realkennzeichen bei der „Hypothesenüberprüfung“ eine beträchtlich höhere Aussagekraft und damit Indizwert für die Glaubhaftigkeit der zu beurteilenden Angaben erlange. Obwohl Fiedler und Schmid betonen, dass die Anwendung dieser Gesetzmäßigkeit der induktiv-statistischen Methode „im forensischen Zusammenhang nicht untersucht“ wurde, wird vom BGH, als folgte er einem quasimathematischen Naturgesetz, angenommen, die Aggregation bewirke bei Verwendung möglichst vieler gut ausgewählter Kriterien gleichsam von selbst eine gutachterliche Aussage mit Indizcharakter. Diese Argumentation entspricht einem induktiven Schließen, also einer Zusammensicht möglichst vieler Merkmale, um evtl. falsche Einzelbeobachtungen relativieren zu können. Festgesetzt hat sich aber in den Köpfen vieler juristischer Anwender*innen die

ser Rechtsprechung die deduktive Überprüfung der „Nullhypothese“ bzw. Unwahrhypothese als einziger möglicher wissenschaftlicher Zugang. Es wird also ein wissenschaftlich zwingender, nur so und nicht anders möglicher Zugang zur Realität der (Aussage-)Wahrheit postuliert, obgleich die hinter der Fassade der deduktiven Alternativlosigkeit tatsächlich angewandte induktive Methode genau dies nicht leistet.

Verstärkt wird diese Problematik, wenn Volbert et al. (2025) zur Rechtfertigung ihres Ansatzes den juristischen „Trumpf“ der Unschuldsvermutung mit der Begründung ziehen, dass die „Nullhypothese“ mit der Unschuldsvermutung ganz besonders gut korrespondiere, weil durch dieses Vorgehen die Fehlklassifikation einer tatsächlich nicht-zutreffenden Belastungsaussage als zutreffend vermieden werden solle (Volbert et al. 2025, S. 20). Das erscheint einerseits, wie oben bereits angedeutet, naheliegend. Andererseits sollte genau unterschieden werden. Die „Nullhypothese“ ist weder in der verfassungsgerichtlichen und strafgerichtlichen Rechtsprechung noch in der Literatur als die einzig zulässige Methode, die der Unschuldsvermutung ausschließlich genügen könne, anerkannt. Als konstituierendes Prinzip jeder am Schuldgrundsatz orientierten Strafrechtspflege schützt sie vor einer Bestrafung ohne vollen Nachweis strafrechtlicher Schuld und bindet diesen Nachweis an die Durchführung eines ordnungsgemäßen fairen Verfahrens (Esser, in: Löwe-Rosenberg, StPO, Bd. 11, 26. Auf. (2012), EMRK Art. 6/IPbPR Art. 14 Randnr. 445). Nach einer vom BVerfG gebräuchlichen Formulierung soll sie den Beschuldigten schützen vor Nachteilen, die einem Schuldspruch oder einer Strafe gleichkommen, denen aber kein rechtsstaatliches prozessordnungsgemäßes Verfahren zugrunde liegt (Lindemann, in: Hilgendorf/Kudlich/Valerius, Handbuch des Strafrechts, 2020, IV. Randnr. 49). Dementsprechend wird es als mit dem Schutzzweck der Unschuldsvermutung als Entscheidungsregel für das Strafverfahren vereinbar angesehen, wenn die Richter, wie in § 261 StPO vorgesehen, bei der abschließenden Würdigung der im gesetzlichen Verfahren erhobenen Beweise frei sind, sofern ihr Schuldspruch auf einer subjektiven Überzeugung von der Schuld des Angeklagten beruht, die von objektiv, rational nachvollziehbaren Erwägungen getragen wird, auch wenn andere weiterhin zweifeln würden (vgl. Esser, in: Löwe-Rosenberg, StPO, Bd. 11, 26. Auf. (2012), EMRK Art. 6/IPbPR Art. 14 Randnr. 494). Weder Unschuldsvermutung noch der mit ihr im Zusammenhang stehende „In-dubio-pro-reo“-Grundsatz können danach auf der Ebene einzelner Beweismittel zur Geltung gebracht werden. Vielmehr würde es dem gleichermaßen fundamentalen Prinzip der freien richterlichen Beweiswürdigung zuwiderlaufen, wenn dem Tatrichter unter Verweis auf die „Unschuldsvermutung“ und die „Nullhypothese“ im Sinne einer Beweisregel Vorgaben dazu gemacht werden,

unter welchen einschränkenden Voraussetzungen die Aussage eines (Opfer-)Zeugen von ihm als glaubhaft anzusehen ist. Das BGH-Urteil von 1999, dessen Vorgaben (wie wir in unserer Expertise dargelegt haben) in der Rechtsprechung nicht nur von den Strafgerichten befolgt werden, mit seiner Anmutung, einen wissenschaftlich zwingenden und alternativlosen Zugang zur (Aussage-)Wahrheit zu eröffnen, hat diesen fehlsamen Eindruck, es handele sich quasi um eine ausnahmslos greifende Beweisregel, befördert. Diese höchstrichterlich induzierte Fehlentwicklung wird bis in die jüngere Vergangenheit verstärkt, indem z. B. auf Karl Popper und den „Positivismusstreit“ verwiesen wird (zur ausführlichen wissenschaftstheoretischen Diskussion: Fegert et al. 2024a).

Hinter der mit höchstrichterlichem Behauptungspathos errichteten Begründungsfassade dominieren zwischen Fehl-, Nicht- und Missverständnis changierende wissenschaftstheoretische Bemerkungen, die von einer kundigen Befassung mit der Thematik weit entfernt scheinen. Wenn auch unsere Kritiker*innen nicht ausschließen wollen, „dass sich fehlerhafte, der Beurteilungslogik der Glaubhaftigkeitsbegutachtung widersprechende Formulierungen in manchen Gutachten finden“, und wenn sie zudem schreiben: „Der Begriff der Nullhypothese ist mit dem methodischen Vorgehen in der Glaubhaftigkeitsbegutachtung nicht per se verknüpft; diese Verbindung hat sich erst mit der entsprechenden Formulierung im BGH-Urteil etabliert“, liegt der Ball im Lager der Jurist*innen, da wir uns über das inhaltlich methodische Vorgehen weitgehend einig sind. Ebenso wie unsere Kritiker*innen haben wir keine Daten darüber, wie häufig solche Missverständnisse bei Jurist*innen sind, und wie stark die gerichtliche Praxis durch solche Missverständnisse und Vereinfachungen geprägt ist.

Auch wenn festzuhalten ist, dass, jedenfalls für die strafrechtliche Praxis, im Einklang mit unseren Aussagen in der Expertise der im Abstract der Replik getroffenen Aussage „Im Fazit bleibt eine lege artis durchgeführte aussagepsychologische Begutachtung aktuell die beste Methode zur Beurteilung der Glaubhaftigkeit einer Aussage; wobei eine kontinuierliche wissenschaftliche Weiterentwicklung und Qualitätssicherung stets erforderlich ist“ nicht widersprochen werden kann, halten wir es aber für an der Zeit, dass der BGH sich noch einmal grundsätzlich zum Vorgehen in der Glaubhaftigkeitsbegutachtung äußert und deutlich macht, dass – wie die Autor*innen es darlegen – es sich um eine ideographisch ausgerichtete Einzelfalldiagnostik im Sinne eines qualitativen Vorgehens handelt.

Kontinuierliche wissenschaftliche Weiterentwicklung durch induktiv statistische Studien? Wie kann Qualitätssicherung der ideographisch qualitativen Einzelanalyse gelingen?

An mehreren Stellen in der Replik von Volbert et al. wird ausgeführt, dass uns empirische Grundlagen, auf Grundlage derer wir letztendlich beurteilen könnten, ob Anwendungsprobleme oder Fehlinterpretationen häufiger oder seltener vorkommen, aus systematischen Aktenanalysen nicht vorliegen. Der Erstautor dieser Replik möchte deshalb erwähnen, dass er sich einer solchen systematischen Aktenauswertung in einem Projekt für das Sozialministerium Mecklenburg-Vorpommern um die Jahrhundertwende gewidmet hat. Der hier relevante Teil der empirischen Untersuchung wurde von König und Fegert (2009) publiziert. Die Vollerhebung in Mecklenburg-Vorpommern verglich 124 Gutachten zur Frage der Glaubwürdigkeit von 1997 bis 1998 mit 124 Gutachten, ebenfalls eine Vollerhebung, zur Frage der Glaubhaftigkeit zwischen 2000 und 2001, also kurz nach dem BGH-Urteil. Grob zusammengefasst, waren die Gutachten formal deutlich besser geworden, was unterstreicht, dass die Rechtsprechung des BGH hier durchaus eine qualitätssteigernde Wirkung hatte: Es wurde signifikant häufiger von der „Nullhypothese“ ausgegangen, vom Konstrukt der Glaubhaftigkeit statt Glaubwürdigkeit gesprochen und seltener ein projektives Testverfahren eingesetzt. Generell kamen Gutachten nach 1990 hoch signifikant häufiger zum Ergebnis, dass Zweifel an der Glaubhaftigkeit (vorher 10%, nachher 46%) bestanden; bei kognitiven Einschränkungen des Opfers (vorher 14%, nachher 60%) war der Unterschied größer. Insofern haben wir quantitative Belege für ein besonderes Risiko im Rahmen dieses methodischen Vorgehens für Betroffene mit kognitiven Einschränkungen. Das gesamte Qualitätsprojekt zur Begutachtung in Mecklenburg-Vorpommern (Fegert et al. 2003; König et al. 2005a, b) hatte einen starken Fokus auf der Übernahme von Gutachtenergebnissen in die Urteile und diente der Klärung der Frage, ob methodische Probleme zu einer geringeren Übernahme von Gutachtenempfehlungen führten. Leider war dies generell nicht so. Das Ergebnis der Urteilsanalyse ergab, dass vor dem BGH-Urteil in den einschlägigen Fällen 47% der Angeklagten nicht verurteilt wurden, während nachher 83% der Angeklagten nicht verurteilt wurden. Der BGH selbst hat in seinem Urteil die Problematik explizit angesprochen, dass die von ihm wegen der Aggregation angenommene Treffsicherheit mit Indizcharakter bei der Verwendung von Realkennzeichen in unterschiedlichen Altersgruppen im Ergebnis differieren kann: „Diese sog. Realkennzeichen können als grundsätzlich empirisch überprüft angesehen werden. Zwar handelt es sich um Indikatoren mit jeweils für sich genommen nur geringer Validität,

d. h. mit durchschnittlich nur wenig über dem Zufallsniveau liegender Bedeutung. Eine gutachterliche Schlußfolgerung kann aber eine beträchtlich höhere Aussagekraft und damit Indizwert für die Glaubhaftigkeit zu beurteilender Angaben erlangen, wenn sie aus der Gesamtheit aller Indikatoren abgeleitet wird. Denn durch das Zusammenwirken der Indikatoren werden deren Fehleranteile insgesamt gesenkt. Diesem Umstand liegt das mathematisch und psychometrisch eingehend untersuchte Prinzip der Aggregation zugrunde (Gutachten Prof. Dr. Fiedler). Dementsprechend lagen die mit Realkennzeichen in Forschungsvorhaben erzielten Ergebnisse regelmäßig deutlich über dem Zufallsniveau. Allerdings bestanden dabei teilweise nicht unerhebliche Fehlerspannen. Inwieweit ihre Bedeutung bei Verwendung gegenüber Personen aus unterschiedlichen Altersgruppen differenzieren kann, ist völlig offen“ (BGH I StR 618/98, Urteil v. 30.07.1999, BGHSt 45, 164, S. 4, Randziffer 22). Ähnlich verhält es sich nicht nur mit Entwicklungsunterschieden aufgrund des Alters, sondern auch Entwicklungsunterschieden in Bezug auf die mentalen Fähigkeiten bei mehr oder weniger starker Intelligenzminderung.

Wichtig ist der Hinweis der Autor*innen der Replik, dass die Gutachtenanalyse von Leve et al. (2023), welche verschiedene Delikte umfasst, wobei Sexualstraftaten den überwiegenden Anteil ausmachten, und die Untersuchung von Höynck et al. (2024) mit der Befragung von 191 Personen, welche eine Vergewaltigung angezeigt hatten, Hinweise darauf geben, dass in 50 % dieser Fälle eine Aussage als glaubhaft eingestuft wurde.

Selbstverständlich können wir nicht mit aktuellen Zahlen belegen, dass aussagepsychologische Gutachten derzeit, im Gegensatz zu unseren Ergebnissen, die kurz nach dem BGH-Urteil erhoben wurden, nicht sogar möglicherweise im Ergebnis zu einem Anstieg von Verurteilungen führen, wie die Autor*innen auf S. 17 ihres Beitrags schreiben, „weil nur durch die systematische Zurückweisung von Unwahrannahmen die Unschuldsumutung überwunden werden kann (vgl. Urteil des Obergerichts Solothurn vom 23. Januar 2023, zit. in Niehaus und Krause 2023).“ Allerdings fehlen hier nach unserer Kenntnis zur Qualitätssicherung auf eine Gesamtpopulation bezogene Untersuchungen ohne stichprobenbezogene wesentliche Selektionseffekte. Diese wären dringend zu fordern.

Es freut uns zu hören, dass „innerhalb der Aussagepsychologie ein Bewusstsein für die Notwendigkeit permanenter Qualitätssicherung, um ein Vorgehen *lege artis* in der Praxis sicherzustellen“, existiert. Ansätze wie das in der Kritik skizzierte, zurzeit pilotierte Peer-Review-Verfahren für aussagepsychologische Gutachten (Belke et al. 2024) sind deshalb ausdrücklich unterstützenswert. Gleiches gilt aus der Perspektive der approbierten Heilberufe in Psychotherapieverfahren, wo wir ebenfalls eine Notwendigkeit einer permanenten Aus- und Weiterbildung sowie Qualitäts-

sicherung für unabdingbar halten, um der Verwendung von suggestiven Techniken und Methoden entgegenzuwirken.

Relevante aussagepsychologische Forschung seit dem BGH-Urteil

Den Kritiker*innen ist zuzustimmen, dass die empirische Untersuchung mit der integrierten Anwendung aller Analyseschritte in der Praxis eine große methodische Herausforderung für die Zukunft darstellt. Denn als Beispiel für die wissenschaftlichen Fortschritte in der Begutachtung zählen die Kritiker*innen wiederum nur induktiv statistische Befunde auf, die letztendlich das Zielkriterium „Wahrheit“ haben, aber sie zeigen keinen Weg auf, wie man das einzelfallbezogene Vorgehen mit der „Unwahrhypothese“ deduktiv überprüfen kann.

Richtig ist, dass wir in der Expertise festgestellt haben, dass in der rechtspsychologischen Forschung keine weitergehenden, den Stand der Wissenschaft fortentwickelten Erkenntnisse erzielt werden konnten (S. 82 der Expertise Fegert et al. 2024b). Vielleicht ist diese Aussage, die auf das Kernproblem der Testgütekriterien bzw. Trefferquoten bei der Klassifikation von Fällen abzielt, zu pauschal gewesen und konnte damit als respektlos angesehen werden. Das bedauern wir, aber der Kontext der interdisziplinären Expertise mit einer Rechtssprechungsübersicht bot nicht den Rahmen für eine weitergehende Methodendiskussion. Hier gibt uns gerade diese Antwort auf die Replik in einer wissenschaftlichen Zeitschrift die Möglichkeit, wenigstens kurz auf die international publizierten Metaanalysen und ihre Bedeutung für die deutsche Praxis einzugehen. Hier müssen wir insbesondere in Bezug auf die Studien aus dem forensischen Kontext, auf die Frage nach dem Feststellungskriterium für die „ground truth“, die zugrunde liegende tatsächliche Wahrheit als Vergleichsmaßstab, eingehen. Deutlich sei hier noch einmal betont: Es geht uns bei diesen Studien nicht um eine empirische Überprüfung des Vorgehens mit der „Unwahrhypothese“, sondern um die induktiv-statistische Überprüfung der „ground truth“ durch einzelne Realkennzeichen oder einen möglichst umfassenden Satz von Realkennzeichen. Dabei ist eine erfolgte Verurteilung oder nichterfolgte Verurteilung sicherlich kein ausreichendes externes, unabhängiges Kriterium in Bezug auf die Treffsicherheit von Merkmalen oder eines Merkmalsatzes. Angesichts der hohen Übernahmequoten gutachterlicher Empfehlungen bei Gerichtsentscheidungen wäre dies geradezu tautologisch. Die von Volbert et al. zitierten Metaanalysen gehen bis auf die Arbeit von Sporer und Masip (2024) nur marginal oder gar nicht auf die Definition des Wahrheitsmaßstabs in den analysierten Studien ein. Neben Polygraphentests in mehreren Studien beschreiben Sporer und Masip (2024) ganz unterschiedliche Anhaltspunkte für

die Wahrheit in unterschiedlichen Studien. Der Zusammenfassung in der Metaanalyse von Oberlader et al. (2021) ist deshalb zuzustimmen: „Moreover, it must be noted that the independence of ground truth criteria and statement quality is largely questionable, and study results may thus be overestimated“ (S. 403).

Charakteristisch für die genannten hier kritisierten Metaanalysen ist die zusammenfassende Darstellung der Treffsicherheit von Begutachtungen in „Effektstärken“ (Amado et al. 2015, 2016, Oberlader et al. 2016, 2021). Die erste Arbeit (Amado et al. 2015) betont, dass alle relevanten Kriterien der CBCA, also der kriterienbasierten Inhaltsanalyse, einen großen positiven Effekt bei der Unterscheidung von Aussagen zeigten, warnt aber zugleich vor einer zu rigiden Anwendung, insbesondere bei traumatisierten Personen, die Schwierigkeiten damit hätten, detaillierte Aussagen zu machen. In der zweiten Arbeit von Amado et al. (2016) mit Erwachsenen wurden nur mittelgroße Effektstärken bei Zeugenaussagen ($\delta > 0,50$) und kleine bis mittlere Effekte in nichtforensischen Kontexten festgestellt (Abb. 1). Zudem war die Treffsicherheit in Feldstudien höher als in Laborstudien, was ein Hinweis auf Kontextabhängigkeit und damit Nichtgeneralisierbarkeit der Ergebnisse darstellt. Die Studie von Oberlader et al. (2016) zeigte, dass die Anwendung mehrerer CBCA-Kriterien zu höheren Diskriminierungsraten führte. Beide Studien unterstreichen die Bedeutung des Bayes-Theorems, nachdem die Treffsicherheit einer Methodik stark abhängt von der Grundgesamtheit, in der die Daten erhoben werden. Hinweise auf einen Publication Bias ergeben sich durch höhere Effektstärken in unveröffentlichten

Studien. Neuere Studien, etwa von Sporer und Masip (2024) sowie Oberlader et al. (2021) zeigen starke Schwankungen der Effektstärken einzelner Kriterien. Mit der Analyse mittels Precision Effect Test (PET) sowie einer Precision Effect Estimate with Standard Error (PEESE) lag die Effektstärke z.B. nahe 0 bzw. nicht signifikant von 0 abweichend. Nur mit dem Propensity Score Matching (3PSM [three-parameter selection method]) zeigte sich mit einer mittleren bis großen Effektstärke von 0,86: Bei erfahrungsbasierten Aussagen hatten 73 % der Fälle einen höheren CBCA Score als erfundene Aussagen.

In randomisierten klinischen Studien werden Effektstärken verwendet, um die Wirksamkeit einer Therapiemethode gegenüber einem Placebo oder gegenüber einem konkurrierenden Ansatz einzuschätzen. Da ein positiver therapeutischer Effekt nicht unbedingt auf die Wirkung der Therapiemethode zurückzuführen ist, sondern auch placebobedingt oder in unserem Beispiel zufallsbedingt sein kann, bezeichnet die Kenngröße der Number Needed to Treat (NNT) ein wichtiges Kriterium für die Relevanz beim klinischen Einsatz in der Praxis. Auf unser Beispiel übertragen könnten man von einer „Number Needed to Evaluate“ sprechen, also der Zahl, die ausdrückt, wie viele Patient*innen mit der Glaubhaftigkeitsanalyse untersucht werden müssen, bis eine*r nach den hier referierten Ergebnissen tatsächlich richtig eingeordnet wurde.

Wir gehen davon aus, dass Jurist*innen sich – ähnlich wie bei den Missverständnissen im BGH-Urteil – testdiagnostische Methodenfragen der Treffsicherheit, z.B. die Darstellung der Effektstärken, nicht unmittelbar erschließen, und dass Kategorien wie „moderate Effekte“ oder „großer Effekt“ nicht deutlich werden lassen, was dies für den Einzelfall bedeutet. Deshalb stellen wir die gefundenen Effektstärken, nach dem Vorgehen von Kraemer und Kupfer (2006) entsprechend ihrem instruktiven Review, anhand von Receiver-Operator-Charakteristiken, sog. ROC Kurven, dar. Der Maßstab für den Effekt ist die Area under the Curve (AUC). Geht man, wie in der einen Untersuchung von Amado et al. (2016), in Bezug auf Erwachsene von einer Effektstärke von 0,6 aus, ist das schraffierte Areal zwischen der Zufallslinie und der ROC-Kurve die zusätzliche Fläche unter der Kurve, die durch die angewandte Methode besser eingeschätzt wird als der Zufall. Ein idealer Test, wie z.B. der kombinierte Aidstest, hätte eine AUC von 0,99 oder 1, würde also fast alle Fälle richtig klassifizieren.

Abb. 2 zeigt, dass die AUC selbst bei einer relativ hohen Effektstärke von 0,9 bei 0,738 liegt, also mehr als ein Viertel der Fälle falsch klassifiziert werden. Errechnet man auf dieser Basis die Number Needed to Evaluate, müsste man 2,103 Fälle begutachten, damit ein untersuchter Fall tatsächlich aufgrund der Methode der kriterienbasierten Aussageanalyse richtig klassifiziert würde.

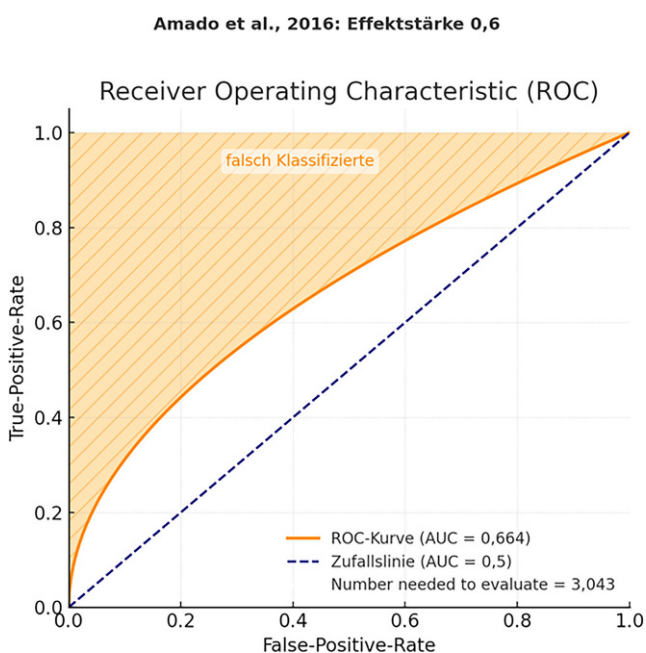


Abb. 1 Amado et al., 2016: Effektstärke 0,6 (0,56), AUC-Wert 0,664, NNE 3,043

Oberlader et al., 2021: Effektstärke 0,9

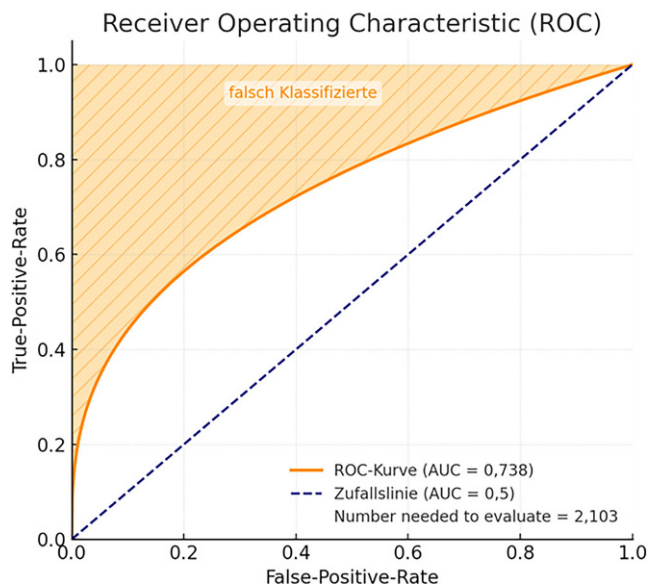


Abb. 2 Oberlader et al. (2021): Effektstärke 0,9 (0,86), AUC-Wert 0,738, NNE 2,103. Grafik erstellt mit Chat GPT. Für die Einsetzung der Tabellenwerte aus Kraemer und Kupfer (2006), wurden die Effektstärken aufgerundet

Uns fehlt hier der Platz, eine solche Analyse, bezogen auf alle Metaanalysen, darzustellen. Der Stand der Wissenschaft war zum Zeitpunkt des BGH-Urteils der, dass Einzelkriterien erhebliche Schwankungen in ihrer Treffsicherheit aufweisen, und dass die Anwendung des kompletten Kriteriensatzes zu besseren Ergebnissen führt. Dieser damalige Stand der Wissenschaft ist durch den aktuellen systematischen Review von Sporer und Masip (2024) bestätigt worden. Insofern halten wir an der kritisierten Feststellung in der Expertise fest, dass „keine weitergehenden, den Stand der Wissenschaft fortentwickelnden Erkenntnisse erzielt werden konnten“ (S. 82 der Expertise). Insbesondere gibt es keine spezifischen Untersuchungen zum im deutschsprachigen Raum als zentral angesehenen Vorgehen mit der ideographischen Überprüfung der „Unwahrhypothese“. Folgt man den Aussagen im ursprünglichen Gutachten von Fiedler und Schmid (1999), wäre hier allein schon einmal im Ansatz interessant gewesen zu überprüfen, zu welchen unterschiedlichen Ergebnissen man gelangt, wenn man die „Unwahrhypothese“ oder die „Wahrhypothese“ überprüft.

Im interdisziplinären Austausch mit der Justiz wird es darum gehen, Maßzahlen wie Effektstärken durch entsprechende Darstellungen, z. B. ROC-Kurven, und Maßzahlen wie die Number Needed to Evaluate so verständlich auszudrücken, dass tatsächlich die weiter bestehenden Herausforderungen verstanden werden und nicht Charakterisierungen wie signifikante moderate bis große Effekte zu weite-

ren Missverständnissen in Bezug auf die Treffsicherheit im Einzelfall führen.

Ein großes Problem dieser Studien sprechen auch die Kritiker*innen hier an: „Wie in den genannten Überblicksarbeiten in der Regel auch reflektiert wird, sollte die Passung zwischen experimentellen Validierungsstudien und praktischer Anwendung noch weiter verbessert werden. Eine Herausforderung für künftige Forschung besteht insbesondere darin, einerseits die tatsächlich qualitative, einfallorientierte Begutachtungssystematik stärker in den Blick zu nehmen und andererseits Möglichkeiten und Grenzen einer stärkeren Standardisierung der Glaubhaftigkeitsbegutachtung zu diskutieren.“ Auch hier können wir, wie oben dargestellt, zustimmen, denn genau dieser Forschungskonflikt zwischen einem induktiv-statistischen Vorgehen bei Aussagen in Studien über Treffsicherheit und der tatsächlich stattfindenden deduktiv-qualitativen einfallorientierten Begutachtung, die von der „Unwahrhypothese“ ausgeht, wird unseres Wissens bisher nirgendwo untersucht. Insofern gab es hier auch keine methodischen Fortschritte, die das vom BGH gewünschte Vorgehen mit der „Nullhypothese“ stärken könnten. Allerdings drängt sich als weitere Implikation auf, dass, wie es der Satz von Bayes unterstreicht, bestimmte Aussagen, z. B. über Effektstärken, immer nur für die Grundgesamtheiten gelten, in denen sie erhoben wurden, dass man also verstärkt Kenntnisse über unterschiedliche relative Wahrscheinlichkeiten, z. B. bei der Anwendung in unterschiedlichen Altersstufen im Kindesalter oder bei Intelligenzeinschränkungen, braucht. Eine Möglichkeit, wie man solche Bayes-Ansätze relativer Wahrscheinlichkeit auch in die fallbezogene ideographisch ausgerichtete Einzelfallbegutachtung einbringen kann, hat Haas (2022) vorgestellt.

Maslow's „Law of the Hammer“, eine Methode für alle Kontexte oder unterschiedliche Vorgehensweisen in unterschiedlichen rechtlichen Kontexten

Entgegen der in der Replik vorgetragenen Auffassung, dass es nur um unterschiedliche bzw. niedrigere Beweisschwellen beim gleichen Konstrukt der Glaubhaftigkeit gehe, und dass diese Thematik seit Langem auch von Co-Autor*innen aus der Replik anerkannt und diskutiert worden sei, gehen wir von grundsätzlich unterschiedlichen Fragestellungen in unterschiedlichen Verfahrenskontexten aus. Bei der Wahl des geeigneten Vorgehens darf also nicht Maslows „Law of the Hammer“ regieren, wonach gilt, dass für diejenigen, die nur einen Hammer als Werkzeug besitzen, jedes Problem zum Nagel wird. Einig sind wir mit den Autor*innen, dass im Strafverfahren im Zweifelsfall die Priorität besteht, falsch-positive Entscheidungen, also ungerech-

fertigte Verurteilungen, zu vermeiden. In kindschaftsrechtlichen Verfahren setzt die Entziehung des Sorgerechts „eine gegenwärtige, und zwar in einem solchen Maße vorhandene Gefahr voraus, dass sich bei der weiteren Entwicklung der Dinge eine erhebliche Schädigung des geistigen oder leiblichen Wohls des Kindes mit ziemlicher Sicherheit voraussehen lässt“ (BGH 14.07.1956 IV ZB 32/56). Klinisch gesprochen handelt es sich hier also um eine Prognosefrage, bei der aktuelle Risiken in Bezug auf ihre künftige Entwicklung gewichtet werden müssen. Ein Gutachten hat sich also, neben der Frage einer Glaubhaftigkeit der Aussage, v. a. mit der Wahrscheinlichkeit und zeitlichen Nähe eines möglichen Schadenseintritts sowie mit der absehbaren Rechtsgutsbeeinträchtigung des Kindes auseinanderzusetzen. Der mögliche Schadenseintritt hängt bei Weitem nicht allein von der Aussage eines Kindes ab. Auch wenn man hier eine Beeinflussung durch einen Elternteil zugunsten eines anderen Elternteils vermutet und einen unterbleibenden Umgang mit dem anderen Elternteil ebenfalls als gegenwärtige in solchem Maß vorhandene Gefahr annimmt, dass durch diese eine erhebliche Schädigung des Kindeswohls zu befürchten sei, heißt das nicht, dass hier, wie von den Kritiker*innen nahegelegt und mit Verweis auf Dettenborn und Walter (2015), Niehaus (2018) sowie Volbert und Kuhle (2019) ausgeführt, ungerechtfertigte Beeinträchtigungen der Beziehungen zu den Eltern den Folgen anderer potenziell belastender und traumatisierender Ereignisse quasi kompetitiv gegenübergestellt werden. Es geht hier nicht um ein Entweder-oder, sondern um eine Berücksichtigung der Gesamtsituation und die Beantwortung der Frage, ob sich in der Zukunft, unter Fortbestehen der festgestellten Ausgangsbedingungen, eine erhebliche Schädigung des Kindeswohls mit ziemlicher Sicherheit voraussehen lässt. Die Annahme, dass man, etwa nach der Feststellung suggestiver Einfüsse, nicht weiter über mögliche Entwicklungsrisiken nachdenken müsse, widerspricht dem Prüfmaßstab kindschaftsrechtlicher Verfahren, nämlich der am Kindeswohl orientierten, zukunftsgerichteten Gefahreneinschätzung.

Im Sozialen Entschädigungsrecht gibt es eine Begutachtungsfrage in Bezug auf die Feststellung der Tat (in der Sprache des Sozialen Entschädigungsrechts: des schädigenden Ereignisses) und eine weitere, die Kausalitätsfrage, in Bezug auf die Tatfolgen (in der Sprache des Sozialen Entschädigungsrechts: Schädigungsfolgen). Vorliegend geht es um die Tauglichkeit der Methode der Glaubhaftigkeitsbegutachtung im Hinblick auf die Tatfeststellung. Mit der Darstellung der entsprechenden Rechtsprechung des Bundessozialgerichts, wie sie von den Kritiker*innen (S. 21) genannt wird, stimmen wir überein. Umso mehr freut uns diese Wendung: „Auch wenn es keine wissenschaftliche Basis dafür gibt, die Wahrscheinlichkeit verschiedener Hypothesen mit numerischen Angaben zu quantifizieren, erscheint

es doch in vielen Fällen möglich, die relative Wahrscheinlichkeit konkurrierender Hypothesen aus aussagepsychologischer Sicht vergleichend zu diskutieren (Lösel und Bender (2001) für ein vergleichbares Vorgehen in Asylverfahren).“ Genau das haben wir für die Begutachtung im Sozialen Entschädigungsrecht gefordert, nämlich nicht die primäre Überprüfung der Unwahrhypothese, sondern die primäre Überprüfung konkurrierender Erklärungsansätze, also eine Anpassung der Fragestellung an die Vorgaben des Gesetzgebers und der früheren Rechtsprechung des Bundessozialgerichts, wo von der „guten Wahrscheinlichkeit“ in Bezug auf anspruchsbegründende Tatsachen gesprochen wird. Das heißt aber, dass die vom BSG in seinem Urteil aus dem Jahre 2016 (Urteil vom 15.12.2016, B9 V 3/15 R) als (vorgebliche) Ansicht „der“ Aussagepsychologie wiedergegebene Ansicht, die Methode der Glaubhaftigkeitsbegutachtung könne nicht nach dem Beweismaßstab differenziert werden (Urteil vom 15.12.2016, B9 V 3/15 R), richtigerweise im Sinne der früheren Rechtsprechung, die dies für möglich hielt (BSG, Urteil vom 17.04.2013, B9 V 3/12 R), korrigiert werden muss.

In diesem Zusammenhang haben wir den Vorschlag gemacht, dass man unnötige Belastungen für Betroffene in solchen Verfahren, insbesondere wenn psychische und physische Langzeitfolgen und Erkrankungen geltend gemacht werden, dadurch reduzieren könnte, dass man qualifizierte Angehörige der Heilberufe mit solchen Begutachtungsaufgaben betraut, da dann beide Fragen, nämlich die der Tatfeststellung und die der Kausalitätsfeststellung durch eine einzige hinreichend qualifizierte Person getroffen werden könnten. Natürlich ist den Autor*innen der Replik zuzustimmen, dass eine psychotherapeutische Ausbildung weder mit theoretischen und methodischen Erkenntnissen zur Unterscheidung zwischen Wahrheit und Täuschung noch mit systematischem Erwerb von Wissen zu Scheinerinnerungen einhergeht (S. 22, in Bezug auf das Zitat von Schemmel et al. 2024). Gleichzeitig gilt umgekehrt, dass Rechtspsycholog*innen eben in der Regel nicht approbiert sind und keine Erfahrung mit der Diagnostik und Krankenbehandlung traumatisierter Patient*innen haben. Eine ideale begutachtende Person müsste demnach Kenntnisse, Fähigkeiten und Erfahrungen aus beiden Bereichen haben.

Der Gesetzgeber hat aber mit der Einführung des grundständigen Psychotherapiestudiums einen zweiten grundständig ausgebildeten Heilberuf geschaffen. Das Studium vermittelt wesentliche Kenntnisse im Bereich der Psychotherapie und führt zur Approbation. Das derzeit nichtgelöste Problem ist der Engpass der Fachweiterbildung in Erwachsenen- sowie Kinder- und Jugendlichenpsychotherapie. Diesbezüglich ist es bemerkenswert, dass der Gesetzgeber nicht nur Ausbildungszeiten in einschlägigen Kliniken und Praxen zur Anerkennung vorsieht, sondern auch Zeiten im Allgemeinkrankenhaus, im Jugendamt oder in ande-

ren Tätigkeitsfeldern. Wir haben derzeit massive Defizite in der Besetzung der Trauma-Ambulanzen bundesweit, aber insbesondere auch im Osten Deutschlands. Deshalb war von den Ländern die Konkretisierung einiger Qualifikationsvoraussetzungen immer wieder kritisch diskutiert worden. Was spräche dagegen, dass man gut qualifizierte junge Psychotherapeut*innen mit Approbation, aber ohne Fachtherapeutenqualifikation, nach entsprechenden vorhandenen Curricula auch spezifisch rechtspsychologisch ausbildet? Damit könnten z. B. in Ausbildungsambulanzen Kompetenzzentren auch für die Begutachtung, welche sowohl die Verfahren nach SGB XIV beschleunigen könnten als auch durch die Maxime der schonenden Verfahrensbewicklung durch Vermeidung von Mehrfachuntersuchungen die Last der Begutachtung für die Betroffenen reduzieren können, geschaffen werden.

Fazit

Der Einwand, wir würden eine vermeintlich bewährte Methode mit empirischer Fundierung durch ein Verfahren ohne systematischen methodischen Anspruch ersetzen, geht ins Leere. Gerade der von den Autor*innen geforderte Ansatz der ideographisch ausgerichteten Einzelfalldiagnostik bedeutet, dass am Einzelfall ausgerichtete diagnostische Fertigkeiten, Erfahrungen und Instrumente eingesetzt werden. Solche Methoden und Instrumente gibt es, sogar mit gut erforschten Testgütekriterien, z. B. bei der Diagnostik einer posttraumatischen Belastungsstörung (PTBS). Die seitens der Autor*innen vorgetragene, unseres Erachtens unnötig überspitzte Argumentation „Hier Wissenschaftlichkeit, dort klinische Zeichendeuterei“ findet sich auch in anderen Schriften, die der Kritik angeblich fehlende Wissenschaftsorientierung (Niehaus und Krause 2023) vorhalten, ohne indes das eigene Wissenschaftsverständnis zu plausibilisieren und zu problematisieren. Uns geht es um eine Vergewisserung über die wissenschaftlichen Grundlagen der Glaubhaftigkeitsbegutachtung, was einen transparenten Umgang mit dem jeweils zugrunde gelegten Verständnis von Wissenschaftlichkeit voraussetzt. Deshalb begrüßen wir, dass die Autor*innen sich auf diese grundlegende Ebene der Vergewisserung einlassen.

Das Hauptmissverständnis der Kritiker*innen bildet ihre Fehlannahme, wir kritisierten die *lege artis* durchgeführte Praxis der Glaubhaftigkeitsbegutachtung. Das tun wir selbstverständlich nicht. Allerdings wenden wir uns gegen die „Nullhypothese“, die sich angestoßen von dem BGH-Urteil aus dem Jahr 1999 zu einer Quasibeweisregel mit vorgeblich strikter Aussagekraft entwickelt hat, obgleich die vom BGH rezipierten aussagepsychologischen Erkenntnisse eine solche starke Aussagekraft nicht belegen. Wir erinnern letztlich – ganz im Sinne der Kritiker*innen – an

Grund und Grenzen aussagepsychologischer Erkenntnisse, während die Kritiker*innen in der Sorge, die Aussagepsychologie gerate generell in Misskredit, die im Ansatz inkonsistente Rezeption aussagepsychologischer Erkenntnisse des BGH verteidigen, sich so aber genau die Inkonsistenz zu eigen machen, die das Ansehen der Methode der Glaubhaftigkeitsbegutachtung generell beschädigt.

Die Kritiker*innen behaupten, wir kritisieren eine Glaubhaftigkeitsbegutachtung, die es so nicht gebe, allenfalls seien gelegentliche Qualitätsmängel einzuräumen. Uns geht es aber nicht um die möglicherweise gelegentlich fehlerhaft gehandhabte Methode, die vermeintlich dem Grunde nach in Ordnung ist. Unsere Kritik richtet sich gegen die grundlegende Fehlerhaftigkeit eines im Kern vom BGH zu vertretenden Ansatzes, der die Wertungsabhängigkeit der Glaubhaftigkeitsbegutachtung als genuin richterlich-normative Aufgabe hinter einem szientizistischen Konstrukt der angeblich zwingenden Wahrheitsermittlung verbirgt, und dies nicht ohne Grund: Das Konstrukt hat aus richterlicher Sicht den Vorteil, dass die eigene richterliche Verantwortung an Sachverständige delegiert werden kann – freilich mit dem Nachteil, dass die Methode der Glaubhaftigkeitsbegutachtung nicht mit solchen eindeutigen Erkenntnissen aufwarten kann, wie sie in anderen Wissenschaftsfeldern gleichsam mit „null Zweifeln“ möglich sind. Das Konstrukt der „Nullhypothese“ insinuiert genau dies, obwohl sie es nicht leisten kann.

Wir empfinden es als wichtiges Signal, dass die Autor*innen die Tür für eine Diskussion weit öffnen, u. a. auch mit Verweis auf Lösel und Bender (2001), wonach die relative Wahrscheinlichkeit konkurrierender Hypothesen auch aus aussagepsychologischer Sicht vergleichend diskutiert werden könne. Sollte unsere Expertise zu dieser Reaktion geführt haben, dann möchten wir nicht von einer vertanen Chance sprechen, denn für die Betroffenen in den Verfahren wird dies zu einer erheblichen Veränderung führen. Es war die am Hilfe-Telefon Sexueller Missbrauch geäußerte Kritik von Betroffenen (Fegert et al. 2018), also die in der Replik zitierten Beiträge von unserer Seite, die dazu geführt hat, dass die Expertise in Auftrag gegeben wurde. Damit war ein neuer Anlass geschaffen, um mehr als ein Vierteljahrhundert nach der grundlegenden (Fehl-)Entscheidung des BGH in Strafsachen eine Diskussion über die adäquate Methode der Glaubhaftigkeitsbegutachtung anzustoßen. Gerade mit Blick auf das Recht der Sozialen Entschädigung wird es wichtig sein, die Impulse der aktuellen Debatte aufzugreifen und die Methode der Glaubhaftigkeitsbegutachtung mit Sinn für die spezifischen Beweismaßstäbe dieses Rechtsgebiets weiterzuentwickeln.

Danksagung Wir danken den Herausgeber*innen der Zeitschrift für die Gelegenheit, unmittelbar auf die Veröffentlichung „Falsche Prämissen und eine vertane Chance – Replik auf die „Expertise“ zur Methode der forensischen Glaubhaftigkeitsbegutachtung (Fegert et al. 2024a;

Volbert et al. 2025) zu reagieren. Wir freuen uns auch, in unserer Republik zahlreiche positive Anknüpfungspunkte für eine weitere Debatte – im Sinne einer besseren Verständigung sowie mit Blick auf offene Forschungs- und Anwendungsfragen – zu finden.

Funding Open Access funding enabled and organized by Projekt DEAL.

Interessenkonflikt **J.M. Fegert** (letzte 5 Jahre): Forschungsförderung von EU, BMG, BMBF, BMFSFJ, DFG, G-BA-Innovationsfonds, Länderministerien Baden-Württemberg und Saarland, Landesstiftung Baden-Württemberg, Ingrid & Frank Stiftung, Stiftung Deutsche Krebshilfe, Auxilium Stiftung, Vector Stiftung, Evangelische Landeskirche Baden-Württemberg, Porticus; Reisebeihilfen, Referentenhonorare, Veranstaltungs- und Ausbildungssponsoring von APK, Adenauer- und Ebertstiftung, Deutschlandfunk, DFG, DJI, DKSB, Infectopharm, med update, UNICEF, Fachverbänden, Universitäten sowie Bundes- und Landesministerien; Beratertätigkeit für APK e.V., Universitätsklinikum des Saarlandes/Staatskanzlei, Bundes- und Landesministerien, Servier; keine Industrie-sponserten Vortragsreihen, kein Aktienbesitz, keine Beteiligung an Pharmafirmen., **M. Pusch** (letzte 5 Jahre): Vortragshonorare/Reisekostenerstattung: Bundesverband der Psychologinnen und Psychologen (BDP) e.V. – Sektion Rechtspsychologie., **S. Rixen** ist seit 2023 Mitglied der Unabhängigen Kommission des Bundes zur Aufarbeitung sexuellen Kindesmissbrauchs. Er hat keine relevanten finanziellen oder nichtfinanziellen Interessen zu deklarieren. **C. Sachser** Forschungsförderung von DFG, BMBF, Porticus. Vortragshonorare: Universität Zürich, Versorgungsmedizinischen Fortbildungstagung. Mitgliedschaften: DGPs, DeGPT, ESTSS, ISTSS. **J. Gerke** und **A. Kliemann** geben an, dass kein Interessenkonflikt besteht.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden. Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen. Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- Amado BG, Arce R, Fariña F (2015) Undeutsch hypothesis and criteria based content analysis: a meta-analytic review. *Eur J Psychol Appl Legal Context* 7:3–12. <https://doi.org/10.1016/j.ejpal.2014.11.002>
- Amado BG, Arce R, Fariña F, Vilariño M (2016) Criteria-based content analysis (CBCA) reality criteria in adults: a meta-analytic review. *Int J Clin Health Psychol* 16:201–210. <https://doi.org/10.1016/j.ijchp.2016.01.002>
- Belke AP, Wolf-Brandstätter L, Pfundmair M, Kannegiesser A (2024) Peer-Review-Verfahren für aussagepsychologische Gutachten im Strafverfahren. <https://www.kompetenz-rpm.de/peerreview/strafrecht/index.php>. Zugegriffen: 18. Nov. 2024
- Dettenborn H, Walter E (2015) Familienrechtspsychologie, 3. Aufl. Ernst Reinhardt, München <https://doi.org/10.36198/9783838586762>

- Esser R (2012) Die Strafprozeßordnung und das Gerichtsverfassungsgesetz / EMRK/IPBPR. In: Löwe-Rosenberg, Bd 11, 26. Aufl
- Fegert JM, Häßler F, Schnoor K, Rebernick E, König C, Auer U, Schläfke D (2003) Bestandsaufnahme und Qualitätssicherung der forensisch psychiatrischen Gutachtertätigkeit in Mecklenburg-Vorpommern bei Mord und Brandstiftungsdelikten. Books on Demand, Norderstedt
- Fegert JM, Gerke J, Rassenhofer M (2018) Enormes professionelles Unverständnis gegenüber Traumatisierten – Ist die Glaubhaftigkeitsbegutachtung und ihre undifferenzierte Anwendung in unterschiedlichen Rechtsbereichen eine Zumutung für von sexueller Gewalt Betroffener? *Nervenheilkunde* 37(07–08):525–534. <https://doi.org/10.1055/s-0038-1668320>
- Fegert JM, Gerke J, Kliemann A, Pusch M, Rixen S, Sachser C (2024a) Expertise Methode der forensischen Glaubhaftigkeitsbegutachtung im deutschen Sprachraum – ein interdisziplinäres Plädoyer für eine kritische Bestandsaufnahme zur Anwendung der sogenannten „Nullhypothese“ in unterschiedlichen Verfahrenskontexten. https://beauftragte-missbrauch.de/fleadmin/user_upload/Materialien/Publikationen/Experten_und_Studien/Expertise_Glaubhaftigkeitsbegutachtung.pdf. Zugegriffen: 7. Mai 2025
- Fegert JM, Sachser C, Pusch M, Kliemann A, Gerke J (2024b) Wissenschaftstheoretische Missverständnisse des BGH in Strafsachen: Überprüfung einer sogenannten „Nullhypothese“ in der Glaubhaftigkeitsbegutachtung. *Z Kinder Jug Psych* 52:342–352
- Fiedler K, Schmid J (1999) Gutachten über Methodik für Psychologische Glaubhaftigkeitsgutachten. *PdR* 9(2):5
- Fiedler K, Walka I (1993) Training lie detectors to use nonverbal cues instead of global heuristics. *Hum Commun Res* 20:199–223
- Greuel L (2009) Was ist Glaubhaftigkeitsbegutachtung (nicht)? Zum Problem der Dogmatisierung in einem wissenschaftlichen Diskurs. *Kindesmisshandl Vernachläss* 12:70–89
- Greuel L (2022) Aussagepsychologische Begutachtung im OEG-Verfahren. In: Deckers R, Köhnken G (Hrsg) Die Erhebung und Bewertung von Zeugenaussagen im Strafprozess: Juristische, aussagepsychologische und psychiatrische Aspekte, Bd. 5. Berliner Wissenschafts-Verlag, Stuttgart, S 45–63
- Haas H (2022) Ein Vorschlag zur methodischen Aktualisierung der Beweiswürdigung in aussagepsychologischen Gutachten. *Schweiz Psychol* 567:573
- Häßler F, Rebernick E, Schnoor K, Schläfke D, Fegert JM (2003) Forensische Kinder-, Jugend- und Erwachsenenpsychiatrie Aspekte der forensischen Begutachtung, 1. Aufl. Schattauer
- Hohoff U (2021) Aktuelle Fragen der aussagepsychologischen Begutachtung in Strafverfahren. *Prax Rechtspsychol* 31:151–160. <https://doi.org/10.51625/pdr20210108>
- Hohoff U (2022) Rechtliche Anforderungen an Beweiserhebung und Beweiswürdigung in Aussage-gegen-Aussage-Konstellationen. In: Deckers R, Köhnken G (Hrsg) Die Erhebung und Bewertung von Zeugenaussagen im Strafprozess: Juristische, aussagepsychologische und psychiatrische Aspekte, Bd. 5. Berliner Wissenschafts-Verlag, Stuttgart, S 11–24
- Höyneck T, Schmid P, Hurler S (2024) Die Strafverfolgung der Vergewaltigung in Niedersachsen. Kassel university press, Kassel
- Jussim L (1991) Social perception and social reality: a refutation-construction model. *Psychol Rev* 98:54–73
- Koehler DJ (1991) Explanation, imagination, and confidence in judgement. *Psychol Bull* 110:499–519
- Köhnken G (2007) Erhebung von Zeugenaussagen im Strafprozess. In: Deckers R, Köhnken G (Hrsg) Band Juristische Weiterbildung, Bd. 1. BMW, Berliner Wissenschaftsverlag, S 41
- König C, Fegert JM (2009) Zur Praxis der Glaubhaftigkeitsbegutachtung unter Einfluss des BGH-Urteils. *Interdiszip Fachz Prävention Interv DGfPI* 12(2):16–41
- König C, Schnoor K, Auer U, Rebernick E, Schläfke D, Fegert JM (2005a) Modellprojekt forensisch-psychiatrische Gutachtertätigkeit

- keit in Mecklenburg-Vorpommern bei Sexualdelikten – Bestandsanalyse. In: Schläfke D, Häßler F, Fegert JM (Hrsg) Sexualstraftaten – Forensische Begutachtung, Diagnostik und Therapie. Schattauer, S 47–56
- König C, Schnoor K, Auer U, Rebernick E, Schläfke D, Fegert JM (2005b) Modellprojekt forensisch-psychiatrische Gutachtertätigkeit in Mecklenburg-Vorpommern bei Sexualdelikten – Qualitätsanalyse unter besonderer Berücksichtigung der Prognose. In: Schläfke D, Häßler F, Fegert JM (Hrsg) Sexualstraftaten – Forensische Begutachtung, Diagnostik und Therapie. Schattauer, S 57–76
- Kraemer HC, Kupfer DJ (2006) Size of treatment effects and their importance to clinical research and practice. *Biol Psychiatry* 59(11):990–996. <https://doi.org/10.1016/j.biopsych.2005.09.014>
- Leve M, Volbert R, Schemmel J (2023) Truthful, fabricated, or based on false memory?!—A systematic analysis of 623 credibility reports in Germany. Vortrag auf der Jahrestagung der European Association of Psychology and Law (EAPL), Cluj-Napoca
- Lindemann M (2020) Prozessgrundrechte und ihre Bedeutung für das Strafverfahren. 1. Auf. In: Hilgendorf/Kudlich/Valerius, Handbuch des Strafrechts, Bd 7. Randnr. 49. <https://www.otto-schmidt.de/handbuch-des-strafrechts-9783811488076>
- Lösel F, Bender D (2001) Qualitätsstandards psychologisch-psychiatrischer Begutachtung im Asylverfahren. In: Bundesamt für Flüchtlinge (Hrsg) Traumatisierte Flüchtlinge, Bd. 7. Bundesamt für Flüchtlinge, Nürnberg
- Niehaus S (2018) Im Interesse kindlicher Opfer. *Prax Rechtspsychol* 28:99–120
- Niehaus S, Krause A (2023) Wissenschaftsorientierung in Sexualstrafverfahren in Gefahr: Fortschritte und Opferinteressen stehen auf dem Spiel. *Prax Rechtspsychol* 33:153–192
- Oberlader VA, Naefgen C, Koppehele-Gossel J, Quinten L, Banse R, Schmidt AF (2016) Validity of content-based techniques to distinguish true and fabricated statements: a meta-analysis. *Law Human Behav* 40:440–457. <https://doi.org/10.1037/lhb0000193>
- Oberlader VA, Quinten L, Banse R, Volbert R, Schmidt AF, Schönbrodt FD (2021) Validity of content-based techniques for credibility assessment—how telling is an extended meta-analysis taking research bias into account? *Appl Cogn Psychol* 35:393–410. <https://doi.org/10.1002/acp.3776>
- Pyszczynski T, Greenberg J (1987) Towards an integration of cognitive and motivational perspectives on social, inference: A biased hypothesis-testing model. In: Berkowitz L (Hrsg) *Advances in experimental social. Inference*, Bd. 20. Academic Press, New York, S 297–340
- Schemmel J, Datschewski-Verch L, Volbert R (2024) Recovered memories in psychotherapy: a survey of practicing psychotherapists in Germany. *Memory* 32:176–196. <https://doi.org/10.1080/09658211.2024.2305870>
- Schläfke D, Häßler F, Fegert JM (2005) Sexualstraftaten – forensische Begutachtung, Diagnostik und Therapie. Schattauer
- Snyder M (1984) When belief creates reality. In: Berkowitz L (Hrsg) *Advances in experimental social psychology*, Bd. 18. Academic Press, New York, S 247–305
- Sporer SL, Masip J (2024) A systematic review of the validity of criteria-based content analysis in child sexual abuse cases and other field studies. *Psychol Crime Law*. <https://doi.org/10.1080/1068316X.2024.2335971>
- Steller M (2020) Stand und Herausforderungen der Aussagepsychologie. *Forens Psychiatr Psychol Kriminol* 14:188–196. <https://doi.org/10.1007/s11757-020-00589-x>
- Steller M, Volbert R (1999) Forensisch-aussagepsychologische Begutachtung (Glaubwürdigkeitsbegutachtung). *Wissenschaftliches Gutachten für den Bundesgerichtshof*. In: PdR, S 46
- Steller M, Volbert R (2000) Anforderung an die Qualität forensisch-psychologischer Glaubhaftigkeitsbegutachtungen. *Prax Rechtspsychol* 10:102–116
- Tversky A, Kahneman D (1974) Judgement under uncertainty: heuristics and biases. *Science* 185:1124–1131
- Volbert R, Kuhle LF (2019) Sexueller Missbrauch (In: Volbert R, Huber A, Jacob A)
- Volbert R, Brackmann N, Gewehr E et al (2025) Falsche Prämissen und eine vertane Chance – Replik auf die „Expertise“ zur Methode der forensischen Glaubhaftigkeitsbegutachtung (Fegert, Gerke, Kliemann, Pusch, Rixen & Sachser, 2024). *Forens Psychiatr Psychol Kriminol*. <https://doi.org/10.1007/s11757-025-00886-3>
- Zuckerman M, Koestner R, Colella MJ, Alton AO (1984) Anchoring in the detection of deception and leakage. *J Pers Soc Psychol* 47:301–311
- Zuckerman M, Knee CK, Hodgins HS, Miyake K (1995) Hypothesis confirmation: the joint effect of positive test strategy and acquiescence response set. *J Pers Soc Psychol* 68:52–60

Hinweis des Verlags Der Verlag bleibt in Hinblick auf geografsche Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.