



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2022

Ein Vorschlag zur methodischen Aktualisierung der Beweiswürdigung in aussagenpsychologischen Gutachten. Kriminalistik

Haas, Henriette

Abstract: Die Aktualisierung der aussagenpsychologischen Begutachtung betrifft die einseitige sog. Unwahrhypothese und das trügerische Widerlegenwollen derselbigen. Zeugenaussagen sollten besser von einer neutralen Ausgangsposition her mit Bayes'scher Logik analysiert und in beide Beweisrichtungen gewichtet werden. Die Zwischenschritte der Methode, das Nennen der Anknüpfungstatsachen, das Aufstellen von Hypothesen und die Auswertungsinstrumente haben sich hingegen langjährig bewährt.

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-222394>

Journal Article

Published Version

Originally published at:

Haas, Henriette (2022). Ein Vorschlag zur methodischen Aktualisierung der Beweiswürdigung in aussagenpsychologischen Gutachten. *Kriminalistik*. *Kriminalistik*, 76(10):567-573.

Redaktion: Schweizerische Kriminalprävention, Chantal Billaud, Tel. 0041 31 511 00 09, info@skppsc.ch

Ein Vorschlag zur methodischen Aktualisierung der Beweiswürdigung in aussagenpsychologischen Gutachten

Von Henriette Haas

Der Expertenstreit zwischen Vertreter:innen der Rechtspsychologie und denen der Psychopathologie, der hier einführend dargelegt wird, verweist auf ein grundsätzliches Problem in der gerichtlichen Beweiswürdigung bei einem Schuldspruch. Es ist dasjenige zwischen der Beweislast für die Schuld (in Abwesenheit einer relevanten Beweislast für die Unschuld) einerseits und dem Schätzen einer Irrtumswahrscheinlichkeit andererseits. Die aktuelle kriminalistische Sichtweise, die hier eingeführt und vertreten wird, kennt die Bestimmung der Beweislast in beide Richtungen (Anklage und Verteidigung), die zudem geeignete Erwägungen zur Kombination verschiedener Beweismittel ermöglicht. Veraltet, weil nicht durchführbar, ist hingegen der Ansatz, der meinte, man könne (und müsse) bei einem Schuldspruch das Irrtumsrisiko abschätzen und minimal halten. Hier geht es um eine Einführung in die neue Denkweise der sog. New Evidence Scholarship. Sie bietet eine dialektische Aufhebung der psychologischen Kontroverse, ohne dass die mehrheitlich bewährten Methoden der Aussagenpsychologie über Bord geworfen werden müssten. Aus Platzgründen kann diese Einführung in die Thematik jedoch weder eine Anleitung für jeden Einzelfall liefern und noch die erforderliche Weiterbildung in eine neue wissenschaftliche Methode ersetzen.



Prof. Dr. Henriette Haas, Psychologisches Institut der Universität Zürich

1. Expertendisput zwischen Rechtspsychologie und Psychopathologie

1.1. Die Errungenschaft der Aussagenpsychologie

Im Grundsatz herrscht Methodenfreiheit in der Begutachtung. Für die Aussagen-

psychologie sind jedoch vom deutschen und schweizerischen Bundesgericht Standards etabliert (BGH 1 StR 618/98 vom 30.7.1999, BGE 6B_572/2008). Sachverständige dürfen zwar davon abweichen, müssen das aber begründen (z. B. BGE 6B_304/2015 E 2.4). Historisch war das 1998 regulierte Verfahren ein Fortschritt. Es fordert eine Transparenz in der Deklaration der verwendeten Hypothesen, Instrumente und in der Beantwortung der Fragen des Gerichts ein, die zuvor fehlte. Dieser methodische Standard darf nicht leichtfertig über Bord geworfen werden. Vorher wurde nicht selten mit „klinischer Erfahrung“ und gelehrt tönenden Phrasen argumentiert. Dabei war keinerlei Verbindlichkeit und Überprüfbarkeit vorhanden.

1.2. Fachleute im Dilemma

Nach Art. 139 CH-StPO setzen die Strafbehörden „zur Wahrheitsfindung alle nach dem Stand von Wissenschaft und Erfahrung geeigneten Beweismittel ein, die rechtlich zulässig sind.“ Doch was gilt als Stand der Wissenschaft, wenn eine Kontroverse herrscht wie derzeit zwischen Expert:innen der Aussagenpsychologie und jenen der Psychopathologie, speziell der Traumaforschung und der Psychotherapie? Psychotraumatologie hat sich in den letzten Jahrzehnten zu einem international anerkannten Fachgebiet mit namhaften und einflussreichen Fachleuten entwickelt. Ihre Ergebnisse flossen in die Inter-

national Classification of Diseases ICD ein. Sachverständige können sich weder über die deutsche Aussagenpsychologie noch über den Stand der psychopathologischen Forschung hinwegsetzen.

Im Zentrum der Kontroverse stehen m.E. drei Postulate der Aussagenpsychologie. Erstens die Prämisse, dass die Unwahrrhypothese der Zeugenaussage am Anfang jeder Begutachtung stehe. Zweitens die Denkweise, man könne aufgrund von Analysen des Beweismaterials quasi deduktiv einen Irrtum „mit an Sicherheit grenzender Wahrscheinlichkeit ausschliessen“. Drittens eine Abgrenzungsfrage: Gibt es Mindestanforderungen an Detailreichtum und Aussagenkonsistenz, damit eine Inhaltsanalyse zur Erlebnisbasiertheit – und damit eine Begutachtung – überhaupt möglich ist?

1.3. Zur Unwahrrhypothese und dem Wahrscheinlichkeitstrugschluss

Das schweizerische Bundesgericht (z. B. BGE 6B_572/2008 E 2.2.1) stützt sich auf deutsche Fachliteratur und damit auf den Leitentscheid des deutschen Bundesgerichtshofs in Karlsruhe (BGH 1 StR 618/98, E 12 a):¹

„Das methodische Grundprinzip besteht darin, einen zu überprüfenden Sachverhalt (hier: Glaubhaftigkeit der spezifischen Aussage) so lange zu negieren, bis diese Negation mit den gesammelten Fakten nicht mehr vereinbar ist. Der Sachverständige nimmt daher bei der Begutachtung zunächst an, die Aussage sei unwahr (sog. Nullhypothese). Zur Prüfung dieser Annahme hat er weitere Hypothesen zu bilden. Ergibt seine Prüfstrategie, daß die Unwahrrhypothese mit den erhobenen Fakten nicht mehr in Übereinstimmung stehen kann, so wird sie verworfen, und es gilt dann die Alternativhypothese, daß es sich um eine wahre Aussage handelt.“

Dies entspreche dem Popperschen Falsifikationsansatz, so Köhnken und Gallwitz (S. 19, Fn. 23). Daraus hat Volbert (2005, S. 171) die Leitfrage für aussagenpsychologische Gutachten präzisiert: *„ob die aussagende Person unter Berücksichtigung der konkreten Umstände, der intellektuellen Leistungsfähigkeit und der Motivlage die zu beurteilende Aussage auch ohne realen Erlebnishintergrund machen könnte.“* (BGer 6B_760/2010 E 2.3).

Die zu immer widerlegende Default-Hypothese ist die Erfindung (Fantasie, Lüge). Weitere Hypothesen zu Suggestion, Autosuggestion, Aggravation, Transfer von Er-

innerungen, oder psychopathologischen Phänomenen ergeben sich aus der Aktenlage. Nebst den bekannten Realkennzeichen hat die Aussagenpsychologie mehrere regelmässig angewendete Analyseheuristiken zur Behandlung der verschiedenen Hypothesen geschaffen (vgl. 4.1, Greuel et al. 1998, Arntzen 2011).

Die Idee, man könne die „*Alternativhypothese, daß es sich um eine wahre Aussage*“ handle, deduktiv aus der Analyse von Beweismaterial ableiten, ist m.E. überholt. Carl Hempel nannte es den induktiven Fehlschluss (vgl. Haas 2020, S. 696; Haas 2017, S. 118).

1.4. Gibt es Schwellenwerte für die Qualität der Vernehmungprotokolle?

Das Kernstück der Begutachtung besteht in der Anwendung der 19 Realkennzeichen (RK). Sie erfassen Ähnlichkeiten mit erwiesenermassen erlebnisbasierten Aussagen in Abgrenzung zu Erfundenen. Ohne Kriterienanalyse lässt sich keine Erlebnisbasiertheit feststellen. Die aussagenpsychologische Schulmeinung stellt hohe Ansprüche an die Aussagenqualität. Sie verlangt für die Evaluation eine wortreiche lebhaft Schilderung als spontane Eigenleistung und eine Aussagen-Konstanz als Mindestanforderungen für eine Begutachtung. Zudem setzt sie eine psychologisch korrekte Vernehmungs- und Protokollierung voraus. Dazu gibt es engere und weitere Auslegungen. Bei den Vorgaben handelt es sich oft um Aufrufe zur Verbesserung der Vernehmungstechnik (z. B. Ludewig, Tavor & Baumer 2011, S. 1426). Später werden sie aber als Veto gegen eine Begutachtung (miss)-verstanden.

Aufgrund einer summarischen Sichtung von Protokollen werden u. U. Gutachtenaufträge abgelehnt, weil die innere Konsistenz (RK 1) oder die Reichhaltigkeit (RK 3) nicht gegeben seien. Solche Plausibilitätsüberlegungen sind anfällig für Irrtümer, weil man sich dabei auf saliente (d. h. ins Auge springende) Merkmale abstützt. Das Wesen von schwierig aufzuklärenden Justizfällen (für die überhaupt erst Sachverständige gerufen werden), ist aber, dass der Schein trügen kann.

1.5. Kritik aus der Traumalogie und der Psychotherapie

Traumaforscher:innen und Therapeut:innen monieren, die Realkennzeichen seien fast nur an psychisch gesunden Proban-

den erprobt worden, wodurch ein falscher Massstab angelegt werde (v. Hinckeldey & Fischer 2002, S. 170–186; Fegert, Gerke & Rassenhofer 2018, S. 531). Daber und Pietrowsky (2017, S. 271) schildern die typische Beweislage, wenn Menschen mit einer Borderline-Persönlichkeits-Störung (BPS) als Geschädigte eines Delikts auftreten: *„Liegt eine BPS vor, so sieht der BGH eine aussagenpsychologische Begutachtung als erforderlich an [...]. Nach den Ergebnissen der vorliegenden Studie sind die Aussagen zum Delikt von Zeugen mit einer BPS besonders häufig mit relevanten Konstanz- und Homogenitätsmängeln versehen. Die Nullhypothese kann in solchen Fällen nicht zurückgewiesen und der Erlebnisbezug deshalb nicht substantiiert werden, mit der Folge, dass – je nach Verfahrensstand – Ermittlungsverfahren eingestellt, Hauptverhandlungen nicht eröffnet oder Angeklagte freigesprochen werden. Dadurch kann das oftmals therapeutisch gesetzte Ziel der Verarbeitung einer Gewalttat konterkariert und der Stabilisierung der psychischen Verfassung des Borderline-Patienten entgegengewirkt werden.“* Ähnliches gilt für Menschen in einer Depression (Volbert & Lau 2013 in Volbert, Schemmel & Tamm 2019, S. 117) und für solche mit einer intellektuellen Beeinträchtigung (Niehaus, Krüger & Caviezel-Schmitz 2013).

Zur Erneuerung und Ergänzung wissenschaftlicher Verfahren muss man keineswegs alles Bewährte *„in Bausch und Bogen“* in Frage stellen, meinen Fegert, Gerke und Rassenhofer (2018, S. 533) – dem schliesse ich mich nur an.

2. Einige Überlegungen zur Kontroverse

2.1. Irrtümer in therapeutischen Kreisen

Manche Meinungen kommen aufgrund fehlender Berufserfahrung mit divergierenden Aussagen von Streitparteien und aus juristischer Unkenntnis zustande. Wer als Therapeut:in den Patient:innen bedingungslos glaubt, verkennt, dass die Aussagen der Gegenpartei u. U. ebenso überzeugend tönen. Eine Traumatisierung besagt nichts über die Erinnerungsfähigkeit der Person, über ihre Suggestibilität oder ihre Ehrlichkeit zu einem bestimmten Sachverhalt. Das Vorliegen einer posttraumatischen Belastungsstörung ist kein *„Wahrheitsbeweis“* für die Schilderungen der Patient:innen. Auch andere in sich

stimmige und in casu zutreffende Erklärungen für Aussagenarmut, wie kulturelle Hemmungen oder Schüchternheit, können nicht als „Beweis“ für die Erlebnisbasiertheit von Aussagen gewertet werden. Sie erklären nur etwas über die Person. Den vom Gericht zu beurteilenden Lebenssachverhalt erhellen sie nicht.

2.2. Zur Handhabung der Falsifikations-Prämisse

Die Falsifikationsprämisse des BGH geht von der Abwesenheit jeglichen Beweismaterials mit Ausnahme der zu begutachteten Aussagen (meist des mutmasslichen Opfers) aus. Oft gibt es sehr wohl Einlassungen vom Beschuldigten und Drittpersonen oder rechtsmedizinische Befunde (z. B. BGer 6B_304/2015 E 2.4), nur wäre es aussichtslos, mit ihnen alleine ein Verfahren führen zu wollen. Vernachlässigbar sind schwächere Beweise aber keineswegs, wie man später sehen wird.

2.3. Zur Einschätzung von Reichhaltigkeit und Unstrukturiertheit

Wenn von unabdingbarem „quantitativem Detailreichtum“ die Rede ist (RK 3), damit eine Begutachtung in Frage käme, stützt sich das auf Experimente, in denen die Realkennzeichen anhand von erfundenen und tatsächlichen Erlebnisberichten statistisch validiert wurden. Der Detailreichtum gilt als das stabilste Einzel-Kriterium. Das unstrukturierte Erzählen (RK 2) ist gehört ebenfalls zu den besseren Kriterien (Vrij 2008, S. 233f). Indessen sind das nur statistische Durchschnitte. In den Laborstichproben war ein erheblicher Prozentsatz der erlebten und der erfundenen Versuchspersonen-Berichte anders gelagert (vgl. Vrij 2008, S. 233f). Die Ermessensunterschiede beim Bestimmen von Detailreichtum waren zudem beträchtlich. Sporer (1997, S. 389, Table 6) fand nur eine Übereinstimmung (sog. Inter-Rater-Reliabilität) von 63.8% zwischen zwei Experten über $N = 80$ zu beurteilende Berichte hinweg. Bei der Unstrukturiertheit war die Übereinstimmung mit 80% wesentlich besser.

Aus dem Wegfall der Möglichkeit, einzelne Kriterien auszuwerten, kann man nicht folgern, dass andere Realkennzeichen auch nicht mehr ausgewertet werden könnten. Dies entspräche einem schematischen Vorgehen, welches der BGH (1 StR 618/98, E 23) kritisiert. Hinter einer kategorischen Forderung nach Aussagenreichtum steht ein Denkfehler. Die Rele-

vanz eines Beweisstückes ist nicht von der Quantität des Materials abhängig, sondern von einem möglichst einmaligen und eindeutigen Informationsgehalt. Eine Faser kann u. U. das entscheidende Beweismittel sein, hingegen liefert ein ganzer Teppich am Tatort keine Information. Dasselbe Phänomen kann in Aussagen vorkommen. Wenn eine Aussage ein einziges für die Erlebnishypothese hochgradig tatrelevantes und spezifisches Detail enthält, das unmöglich erfunden oder suggeriert worden sein kann, oder das anderweitig verankert werden kann, dann gewinnt sie in hohem Grad an Glaubhaftigkeit.

2.4. Zur Handhabung der Konsistenzprüfung

Die Konsistenzprüfung der Aussage (RK 1) ist ebenfalls enorm wichtig (BGH 1 StR 618/98 E 18aa). Leider wird sie oft zu oberflächlich durchgeführt. Ob eine Aussage, die in sich widersprüchlich zu sein scheint oder die im Gegensatz zu anderen Tatsachen zu stehen scheint, es dann tatsächlich ist, kann man nur im Abgleich mit dem gesamten ermittelten Beweismaterial feststellen. Dazu ist intellektuelle Vertiefung nötig; ein einmaliges „Sichten“ genügt keinesfalls.

Experimentell konnten wir zeigen, dass es Konstellationen gibt, in denen selbst gut ausgebildete, intelligente Personen eine falsche Auslegung des Sachverhalts lieferten, die im Gegensatz zu sichtbaren Tatsachen stand (Haas, Djordjevic & van Ackere 2019). Dies passiert dann, wenn sich (unerkannt) Spuren verschiedener Ereignisse überlagern und zu einem schwer verständlichen, unüblichen Bild führen. Ungereimte und tatsachenwidrige Aussagen in guten Treuen kommen also vor, denn Inkonsistenzen und Lücken werden schlecht wahrgenommen und noch schlechter abgespeichert. Wer angelogen wurde, erzählt u. U. das Hin- und Her von Gehörtem und Erlebtem telquel. Leute ohne wissenschaftliche Ausbildung können Inkonsistenzen tw. kaum reflektieren. Andere merken, dass etwas nicht stimmt, aber es fehlt ihnen das Vokabular, um es zu benennen oder es kommt ihnen ausser „Lüge“ kein besseres Wort in den Sinn und dann sind sie blockiert.

2.5. Einigkeit der Psycho-Fachleute

Einig sind sich die Schulen aber in folgendem (Volbert, Schemmel & Tamm 2019, S. 113): „Völlig in Übereinstimmung mit

aussagepsychologischer Literatur führen die Autoren weiter aus, dass es sich bei der aussagepsychologischen Begutachtung nicht um ein statistisches Testverfahren, sondern um einzelfallorientierte Entscheidungen handelt, bei denen das Auftreten von Merkmalen je nach den im Einzelfall vorliegenden Voraussetzungen gewichtet und interpretiert werden muss; schematische Anwendungen sind daher zu kritisieren.“

Diese Vorgabe der Rechtsprechung bereitet einen Weg zur dialektischen Auflösung. Der Entscheid BGH 1 StR 618/98 lässt genügend Ermessensspielraum für den Fall von nur partiell vorhandenen Schilderungen und für psychische Störungen zu. In den Erwägungen E 21–22 und E 24 ist keine Rede davon, dass die Auswertung der RK dann gar nicht angewendet werden dürfen; in E 17 geht es um die freie Wahl der Methoden; in E 23 geht es um Einzelfälle, in denen Vorliegen nur einzelner Realkennzeichen zulässig sein kann und E 27 bb) spezifiziert, dass jegliche Auswertung die Individualität der begutachteten Person berücksichtigen muss. Zudem ist es keineswegs so, dass der BGH beim Vorliegen schwerer psychischer Störungen (die per Definitionem Irrationalität und eine Verengung des Verhaltensrepertoires beinhalten) von aussagenpsychologischen Begutachtungen absehen würde, ganz im Gegenteil, er definiert die jeweiligen Zuständigkeiten von Psychiatrie oder Psychologie für solche Fälle (BGH 1 StR 5/02 zit. in Böhm & Lau 2007, S. 52). Bezüglich der Abklärung der Zeugentüchtigkeit sind sich beide Schulen einig. Die Diagnostik ist allseits bekannt.

3. Die New Evidence Scholarship zur Beweiswürdigung

3.1. Kritik an der nicht neutralen Prämisse der Aussagepsychologie

Gemeinhin nehmen kriminalistische Sachverständige das Nichtwissen in beide Beweisrichtungen als Ausgangslage an oder überlassen ihre Bewertung dem Gericht. Sie unterstellen nicht eine Nullhypothese gegen nur eine Partei, die zu widerlegen sei. Das vollständige Widerlegen der Unschuldsumutung im Sinn eines logischen Exklusionsschlusses obliegt einzig und alleine dem Gericht und zwar auf der Basis des gesamten Beweismaterials und nur in Bezug auf den angeklagten Tatbestand und die zugehörigen Probanda der Verbrechenslehre. Es ist mir keine methodolo-

gische Vorschrift bekannt, die besagte, dass das Falsifikationsprinzip angewendet werden dürfe, ohne alle vorhandenen Daten zu berücksichtigen. Ein Ignorieren von erhobenen und sachdienlichen Daten würde Poppers Ansatz (und übrigens jedem redlichen, wissenschaftlichen Unterfangen) geradezu widersprechen. Formallogische Beweisregeln können nämlich versagen, wenn man sie unreflektiert auf empirische Lebenssachverhalte anwendet (Haas 2020, 2017).

Expert:innen der modernen Kriminalistik berücksichtigen das und bewerten nur den Erkenntnisgewinn resp. das Gewicht, der ihnen zur Begutachtung vorgelegten Beweismittel in beide Beweis-Richtungen: Anklage und Verteidigung.

3.2. Der veraltete frequentist approach mit Wahrscheinlichkeitsüberlegungen

Die „Wahrnehmung“ per Umkehrschluss wird frequentist approach genannt. Die neue evidence basierte Kriminalistik schaut den Anspruch einer umfassenden Falsifikation als undurchführbar an (Taroni, Biedermann & Bozza 2016 e32). Die Irrtumswahrscheinlichkeit (auch in der Statistik) bezieht sich immer nur auf die untersuchten Daten und nur auf die angewendete Methode. Sie kann nicht verallgemeinert werden. Sie besagt nichts über den möglichen Irrtum der Untersuchung angesichts der unbekanntes und niemals umfassend rekonstruierbaren historischen Grundwahrheit. Dies hat Russel (1912, Chap. VI) mit seiner Metapher von einem Huhn gezeigt, das die Gesetzmässigkeiten seines Universums verstehen möchte (hier ein bisschen ausgeschmückt):

Ein Huhn verfällt dem frequentist approach

Ein Huhn erhält jeden Tag Futter, Obdach und Pflege und schliesst daraus induktiv, dass der Bauer ihm wohlgesinnt sei. Tag für Tag bestätigt sich die Statistik. Aus heiterem Himmel passiert dann ein unglaublicher Verrat: das Huhn wird zum Braten. Nichts hat aus seiner Perspektive auf eine bevorstehende Ermordung hingedeutet. Seine Wahrnehmung und Berechnung waren völlig korrekt, Fisher's t-Test ergab eine „Irrtumswahrscheinlichkeit“ von $Pr < 0.01$. Nur hat das Huhn nicht bedacht, dass unbekannte (ihm nicht zugängliche) Faktoren in den Lauf der Dinge eingreifen können, so wie das eigentliche Motiv des Bauern für die Hühnerhaltung.

Illustration 1

Wegen der unknown unknowns kann nie ausgeschlossen werden, dass eines Tages ein neues Beweismittel oder eine neue Methode auftaucht, welche den Beweiswert des vorliegenden Materials grund-

sätzlich in Frage stellt. In DNA Gutachten wird man Formulierungen, wie „es könne mit an Sicherheit grenzender Wahrscheinlichkeit geschlossen werden, dass ...“ heute hoffentlich nicht mehr finden. Die Bayes'sche Logik ist eine Abwandlung der Wahrscheinlichkeitsrechnung, die einen wesentlich bescheideneren Anspruch hat, sie will nämlich nur noch den Erkenntnisgewinn beziffern und nicht mehr die „Wahrheit“ feststellen.

3.3. Gewichtung der Erkenntnisstärke

Bayes'sche Logik besteht in einer Gewichtung der Erkenntnis-Stärke eines Beweismittels zum Beispiel diejenige von Pengelly (Robertson & Vignaux 1995, S. 57). Die likelihood ratio (LR) ist der Multiplikator, der das Gewicht des Erkenntniszuwachses gegenüber der Ausgangslage beziffert,

$$OR_{\text{post}} = LR(B) * OR_{\text{prior}}$$

Der Glaube der Expert:innen an die Beweiskraft des Beweismittels für eine Hypothese wird mit „sehr stark“, „stark“, „gut“, „mässig“ oder „schwach“ ausgedrückt. Wenn man sie beziffern kann, entsprechen diese Wertungen folgenden Zahlen:

- „Sehr stark“ entspricht einer $LR > 1000$
- „Stark“ entspricht $330 < LR < 1000$
- „gut“ entspricht $100 < LR < 330$
- „mässig“ entspricht $33 < LR < 100$
- „schwach“ entspricht $1 < LR < 33$

Wenn man nun einem Beweismittel quasi eine Note gibt, ist das eine bescheidenere Würdigung als es der frühere Denkmodus war. Früher meinte man, das Gegenteil „ausschliessen“ zu können und versuchte, den Zweifel zu bemessen. Was heute zurückhaltender mit einer $LR > 1000$ benotet wird, wäre früher fälschlich mit der Phrase einer „vernachlässigbaren Irrtumswahrscheinlichkeit von $Pr < 0.001$, kleiner

Das Huhn lernt New Evidence Scholarship und rechnet mit Bayes

Hätte Russels Huhn den Erkenntnisgewinn aus seiner Statistik mit der Likelihood ratio statt mit Fisher's t-test gerechnet, hätte es dem begründeten Glauben an die Fürsorglichkeit seines Bauern eine Stärke „gut“ gegeben ($LR \approx 100$).

Die Rechnung ist also ganz ähnlich. Der Unterschied besteht darin, dass das Huhn sich jetzt darüber im Klaren ist, dass ein unbekannter Faktor irgendwann auftauchen könnte und sich dann alles ändert.

Illustration 2

wenn ein Beweismittel neu hinzukommt. Der Einfachheit halber nehmen wir an, die Ausgangslage schuldig zu unschuldig sei 1:1. Also die $OR_{\text{prior}} = 1$, d. h. die Chancen sind am Anfang für beide Beweisrichtungen gleich verteilt. Zur Umrechnung von Wahrscheinlichkeiten in Chancen (OR), siehe Endnote. Die LR entspricht dem Er-

als 1 Promille“ überinterpretiert worden (resp. z. B. bei $LR > 33$: „kleiner als 10%“).

Gemäss den internationalen Standards muss das Beweismittel auch unter der Verteidigungshypothese mit einer Gewichtung versehen werden. Die andere Beweisrichtung wird jeweils mit der Division gerechnet. Biedermann und Vuille (2016, S. 590): „it is now widely argued in both scientific and legal literature that the scientist's findings, when considered at an evaluative stage (Section Clarifying the nature of the report), should be structured along three questions ... : first, 'what is the probability of the evidence given the prosecution's case and the case circumstances', second, 'what is the probability of the evidence given the defense's case and the case circumstances', and third, 'under which proposition are the findings more probable, under the first or the second proposition?'“ Die beidseitige Evaluation steht im Einklang mit Art. 6 Abs. 2 CH-StPO und der Beweismittelwürdigung, weil: „Indizien oft nicht von vornherein einschlägig sind, weil sie nicht ausschliesslich

auf ein bestimmtes Szenario hindeuten. Es gilt, die Indizien daraufhin zu überprüfen, ob sie ausschliesslich für eine Hypothese sprechen, oder ob sie ambivalent sind, weil sie je nach Kontext unterschiedlich verstanden werden können“ (BGE 6B_804/2017).

Die ersten zwei Fragen von Biedermann und Vuille sollten m.E. für psychologische Gutachten übernommen werden. Zur Integration der beiden Gewichte besteht ein Ermessensspielraum. Wenn fachspezifische Kenntnisse unabdingbar sind, soll sie von Expert:innen vollbracht werden, sonst eher vom Gericht (Marquis, et al. 2016, S. 366).

richtung nicht „mit einer grossen Wahrscheinlichkeit für diese Hypothese“ verwechselt werden darf, sondern dass die Gewichte anderer Beweiselemente darin sachkundig integriert werden müssen. Dies entspricht m.E. dem Wissensstand in den forensic sciences und der Medizin besser als die frequentist-Prämisse von vor zwanzig Jahren.

Für die Aussagenpsychologie heisst das konkret, dass man wie bisher prüft, ob es in den Akten Hinweise z. B. auf Suggestion hat. Nehmen wir hier an, es habe suggestive Fragen in der Vernehmung gegeben. Als Ausgangslage von 1:1 wird dann erhoben, was dafür und was dagegen

vorhanden, wenn elaborierte Suggestivfragen unkritisch bestätigt werden in Abwesenheit einer eigenständigen Einlassung zum Thema. Je nach dem muss noch die Suggestibilität als Charaktermerkmal getestet werden. Beide Tendenzen, die Resistenz und die Willfähigkeit müssen gegeneinander abgewogen werden. Sachverständige würden dann der Suggestionshypothese ein Gewicht zusprechen, z. B. mässiges Gewicht gegen Suggestion $33 > LR < 100$ (pro Anklage) oder z. B. starkes Gewicht für suggestive Wirkung $LR > 330$ (pro Verteidigung). Ein Ergebnis kann sein, dass ein Teil des Sachverhalts unbeeinflusst von Suggestivfragen war, aber ein anderer Teil war es doch, wenn er aus irgendwelchen Gründen nicht klar erinnert wird. Die Ergebnisse werden dann mit den anderen Befunden der Aussagenpsychologie integriert, wodurch am Schluss eine Note zugunsten der Anklagehypothese und eine Note zugunsten der Verteidigungshypothese resultierte. Die Staatsanwaltschaft und das Gericht können daraufhin die Gewichte der anderen Beweismittel für und gegen die Anklage eruieren und sie mit den Faktoren der Aussagenpsychologie multiplizieren oder dividieren, je nachdem in welche Richtung sie zeigen (Illustrationen 3 & 4).

Multiplikative Kombination der Beweismittel bei Zweifel an der Schuld

Die Kombination eines Gutachtens alter Schule, das alle Alternativen zur Erlebnisbasiertheit vollumfänglich „widerlegte“, kann durch schwache gegenläufige Indizien erschüttert werden. Wenn aus dem Gutachten eine $LR_1 \approx 1000$ resultierte, aber ein unabhängiges Beweismittel (z. B. Aussagen des Beschuldigten, bevor er Akteneinsicht hatte) schwache Zweifel an der Schuld weckt mit $LR_2 \approx 33$, dann wäre das Ergebnis:

$$LR_{\text{komb.}} = LR_1 / LR_2 \approx 1000 / 33 \approx 33.$$

Der „Schuldbeweis“ wäre nur noch schwach oder passabel. Diese Würdigung müsste in dubio pro reo zum Freispruch führen und zwar trotz eines schlüssigen, nachvollziehbaren und vollständigen Glaubhaftigkeits-Gutachtens, welches die Unwahrscheinlichkeit vermeintlich widerlegt hat.

Würde jemand intuitiv Beweisgewichte addieren und subtrahieren, dann käme er zu einem ganz anderen Schluss, nämlich:

$$„LR_{\text{falsch komb.}} = LR_1 - LR_2 \approx 1000 - 33 \approx 967“$$

Das wäre fast ein sehr starker „Schuldbeweis“, der aber nach heutigem state of the art falsch ist.

Illustration 3

Wichtig für die Beweiswürdigung ist, dass sich die Gewichte von unabhängigen Beweismitteln multiplizieren und nicht addieren. Ein Beweismittel vermag u. U. trotz schwachem Gewicht einen entscheidenden Einfluss auszuüben.

spricht, dass Suggestivfragen tatsächlich einen Einfluss ausgeübt haben, indem alle Antworten unter die Lupe genommen werden. Unabhängig sind Ueberhangantworten und sonstige Modifikationen, sowie die Antwort: „ich weiss es nicht“ und

Kombination der Beweismittel mit schwachen Zusatzindizien für Schuld

Umgekehrt könnte ein aussagenpsychologisches Gutachten mit Bayes zu einem nuancierten Befund kommen, der den Aussagen eine mässige Unterstützung der Anklagehypothese nach Pengelly zuspricht (und keine für die Verteidigung), in einer Grössenordnung von $LR_1 \approx 90$. Wenn weitere schwache Indizien für die Schuld sprächen, die das Gewicht von nur $LR_2 > 10$ erreichten und sie stochastisch unabhängig vom Aussagenbeweis sind, dann würde das kombinierte Gewicht nahe an einen sehr starken Beweis kommen:

$$LR_{\text{komb.}} = LR_1 * LR_2 > 900.$$

Illustration 4

Marquis et al. (2016, S. 370) empfehlen Sachverständigen darauf hinzuweisen, dass ein starkes Gewicht für eine Beweis-

„ich glaube nicht“. Die Antworten signalisieren Resistenz gegen Beeinflussung. Mehr oder weniger Willfähigkeit wäre

4. Synthese zur dialektischen Aufhebung der Kontroverse

4.1. Beibehaltung der bewährten Auswertungsinstrumente

Selbstredend sollen die bewährten von BGH 1 StR 618/98 genannten Auswertungsheuristiken als Vorbereitungsarbeiten nach wie vor zur Anwendung kommen: das Benennen der Anknüpfungs- und Befundtatsachen und das Aufstellen von Unter-Hypothesen (E 14, 17), die Kompetenz-Analyse (Zeugentüchtigkeit) (E 35 – E 36, E 49), die Auswertung der Realkennzeichen (E 22 – E 24), die Konstanzanalyse (E 26), die Motivationsanalyse (E 28, E 34), die Aussagengenesse (E 29), sowie auch der Strukturvergleich (Ludewig, Tavor & Baumer 2011, S. 1428).

4.2. Nuancierte psychologische Aussagegewichtsanalyse nach Bayes

Ich plädiere dafür, dem Gericht eine möglichst vollständige und nuancierte Darstellung des psychologischen Beweisgewichtes aller Aussagen der begutachteten Person vorzulegen, was nicht apriori die Re-

duktion auf eine ja/nein-Meinung sein kann. Anstelle der unrealistischen Falsifikation sollte ein Gutachten eine Beweisstärke für den Aussagenbeweis auf der Skala von Pengelly vorlegen. Je nach Fall können verschiedenen Beweisthemen verschiedene Noten bekommen (z. B. lange zurückliegende Vorfälle erhalten tiefere Werte als die neuen, die besser erinnert werden). Diese Noten erlauben es dem Gericht, die Volbertsche Leitfrage selber zu beantworten, je nach dem, welche Beweismittel hinzukommen und wie es diese würdigt.

Es stellt sich die Frage an die Gerichte, ob im hier relevanten Szenario die Gewichtung des Beweismaterials mit geschätzten Zahlen für die Evaluation der Kombinationen als semantische Auslegung zulässig sei. Anders als bei einigen (keineswegs allen) Gutachten in der Kriminalistik, bei denen es kombinatorisch möglich ist, eine LR zu berechnen, ergibt sich aus den Verfahren der Aussagenpsychologie keine Zahl. Indessen kommen die Zahlen in der herkömmlichen Beweiswürdigung durch die Hintertüre trotzdem ins Spiel. Ein traditionelles Gutachten, das die Volbert'sche Leitfrage mit „nein“ beantwortet, nimmt ebenfalls eine geschätzte mathematische Gewichtung vor – aber auf intransparente Art und dazu noch auf willkürliche Weise absolut. Das „nein“ kann in seiner semantischen Determiniertheit nur heissen, dass das Gutachten von einer verschwindend geringen Irrtumswahrscheinlichkeit ausgeht. Etwas anderes kann es nicht bedeuten, es sei denn die Sachverständigen wären „Gott“ und kennten „die Wahrheit“ als solche.

Bei den bisherigen rein semantischen und deshalb impliziten Würdigungen ist eine Kritik am Beweisgewicht, dass die einzelnen Richter:innen den Beweismitteln zumessen, nicht möglich. Zudem besteht die Gefahr, dass diese Gewichte summiert statt multipliziert würden. Das hier vorgeschlagenen *Procedere* ist transparenter. Es ermöglicht, die Zuschreibung der Werte zu kritisieren und dazu mehrere Varianten durchzurechnen, um obere und untere Schätzwerte zu eruieren.

Nicht zu vergessen ist, dass sich das Gericht vorgängig mit der (ursächlichen) Unabhängigkeit oder Abhängigkeit der Beweismittel untereinander befassen muss. Die Formeln werden komplexer, wenn das Material stochastisch nicht unabhängig ist. Abhängig sind z. B. Aussagen von Mitbeschuldigten, die schon von der ersten Einvernahme an teilgenommen haben.

Wie Beweisgewichte unter Abhängigkeit zu schätzen wären, müssten Fachleute exemplarisch für häufige Sachverhalte berechnen, die dann als Vorlage dienen. Im Zweifelsfall sollte der geschätzte Erkenntnisgewinn nur einmal in die Rechnung einbezogen werden.

Für psychologische Sachverständige ändert sich letztlich nicht viel, für die tatsächlichen Opfer von Straftaten und für zu Unrecht Beschuldigte würde sich jedoch sehr viel zur Aufklärung ändern. Wenn sich anhand der bewährten Heuristiken zur Evaluation der Hypothesen herausstellt, dass Anhaltspunkte für eine Erlebnisbasierte dürftig sind, dann wird das psychologische Gutachten dem Beweiswert der Aussagen ein geringes Gewicht zumessen. Die Berücksichtigung der Beweisgewichte zugunsten der Verteidigungshypothese, käme z. B. beim Unterschied zwischen den objektiven Tatbestandselementen und den Subjektiven zum Tragen. Bei den Sexualdelikten muss die Erkennbarkeit der Nicht-Zustimmung zu sexuellen Handlungen für die Schuld eines Täters berücksichtigt werden. Eine Nicht-Erkennbarkeit des eigentlichen Willens des mutmasslichen Opfers kann sich u. U. in den Aussagen, im Verhalten während der Tat oder im Aussageverhalten bei der Einvernahme spiegeln oder sie kann als Teil eines erfüllten Realkennzeichens entdeckt werden.

4.3. Meldung von Unbekanntem

Die Abkehr vom Falsifikationsansatz bedeutet nicht, dass Unsicherheiten im Dunkeln bleiben sollen. Es bietet sich folgende Lösung an: Alle unbekanntes Grössen, welche Sachverständige in ihrer Arbeit antreffen, die das Beweisgewicht ihres Gutachtens beeinflussen könnten, können sie in der Konklusion unter dem Titel „Unbekannte Faktoren“ melden (vgl. Haas 2020, S. 701). Meldungen sind keine Würdigung, keine Priorisierung und keine Empfehlung. Es kann sein, dass sich unter ihnen solche befinden, die rechtlich unerheblich oder undurchführbar sind.

4.4. Sollen Richter:innen in Zukunft die „Wahrheit“ berechnen?

„Der Richter rechnet nicht“, hiess es im römischen Recht – sondern er interpretiert. Der Vorschlag, die multiplikative Bayes-Regel zu berücksichtigen, setzt keine mathematischen Kenntnisse voraus, die über Mittelschulwissen hinausgehen. Es handelt sich um ein caveat und ein Ausschöp-

fen neuer wissenschaftlicher Entwicklungen und nicht um eine Abkehr von der Hermeneutik. Letztlich kommt jede wissenschaftliche Tätigkeit an den Punkt, an dem die Wahrnehmung des Zueinanderpassens von Beweismaterial mit dem jeweiligen Element des Probandums nicht mehr weiter begründet werden kann. Man ist sich subjektiv einfach sicher. Schliesslich identifiziert man auf der Strasse mühelos und sicher alte Bekannte, obwohl eine verbale Beschreibung fadenscheinig wäre („hat blaue Augen“) oder gar ev. nicht mehr zutrifft, wenn z. B. die Haare grau geworden sind.

Es ist übrigens nicht so, dass die Bayes'sche Logik auf jedes Beweisproblem eine Antwort hätte und eine Zahl berechnen könnte, z. B. wenn Scherben perfekt zusammenpassen, ist das zwar sichtbar, aber nicht berechenbar. Oft fehlen die statistischen Parameter zur Ausgangslage und es besteht die Gefahr, dass man etwas zu Hilfe nimmt, das leicht erreichbar ist, ohne darüber nachzudenken, dass es als empirische Grundlage haltlos ist (Haas 2020, S. 697f). Das entspräche dann wieder den unbedarften Rechnungen des Russelschen Huhns.

Literatur

- Arntzen, F. (2011). *Psychologie der Zeugenaussage*. München: Beck.
- Biedermann, A. & Vuille, J. (2016). Digital evidence, 'absence' of data and ambiguous patterns of reasoning. *Digital Investigation*, 16, 86–95.
- Böhm, C. & Lau, S. (2007). Borderline-Persönlichkeitsstörung und Aussagetüchtigkeit. *Forensische Psychiatrie, Psychologie, Kriminologie* 1, 50–58.
- Daber, B. & Pietrowsky, R. (2017). Die Fähigkeit zur Wirklichkeitskontrolle und die Aussagequalität bei Zeugen mit einer Borderline-Persönlichkeitsstörung. *Forensische Psychiatrie, Psychologie, Kriminologie*, 11(3), 265–272.
- Fegert, J. M., 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 Betroffene? *Nervenheilkunde*, 37, 525–534.
- Greuel, L., Offe, S., Fabian, A., Wetzels, P., Fabian, T., Offe, H. & Stadler, M. (1998). *Glaubhaftigkeit der Zeugenaussage. Theorie und Praxis der forensisch-psychologischen Begutachtung*. Weinheim: Beltz.
- Haas, H. (2019). Obskurantismus als Gegenspieler zur kriminalistischen Aufklärung. *Kriminalistik*, 73(10), 615–622.
- Haas, H. (2020). Stringente Beweisführung und begründete Zweifel versus Schein-Beweise. *Kriminalistik*, 74(11), 699–707.
- Haas, H., Djordjevic, M. & van Ackere, A. (2019). Tarski and the intricacies of reasoning under uncertainty. In W. Rother (Hg.), *Irrtum und Erkenntnis*, Zürich: Conexus, 2, 25–51.
- Haas, H. (2017). Zur Würdigung des Aussagenbeweises. *Kriminalistik*, 71(2): 117–124.

- Köhnken, G. & Gallwitz, S. (2021). Fehlerquellen in aussagepsychologischen Gutachten. In: R. Deckers & G. Köhnken (Hg.). *Die Erhebung und Bewertung von Zeugenaussagen im Strafprozess*, Berliner Wissenschaftsverlag, 18–60.
- Ludewig, R., Tavor, D. & Baumer, S. (2011). Wie können aussagepsychologische Erkenntnisse Richtern, Staatsanwälten und Anwälten helfen?. *Aktuelle Juristische Praxis (AJP)*, 11, 1415–1435.
- Marquis, R., Biedermann, A., Cadola, L., Champod, C., Gueissaz, L., Massonnet, G., Mazzella, W., Taroni, F. & Hicks, T. (2016). Discussion on how to implement a verbal scale in a forensic laboratory: Benefits, pitfalls and suggestions to avoid misunderstandings. *Science & Justice*, 56(5), 364–370.
- Niehaus, S., Krüger, P., & Schmitz, S. C. (2013). Intellectually disabled victims of sexual abuse in the criminal justice system. *Psychology*, 4(03), 374–379.
- Robertson, B. & Vignaux, G. A. (1995). *Interpreting Evidence – Evaluating Forensic Science in the Courtroom*. Chichester: Wiley.
- Russell, B. (1912). *The Problems of Philosophy*. Oxford University Press.
- Taroni, F., Biedermann, A. & Bozza, S. (2016). Statistical hypothesis testing and common misinterpretations: Should we abandon p-value in forensic science applications?. *Forensic Science International*, 259, e32–e36
- v. Hinckeldey, S. & Fischer, G. (2002). *Psychotraumatologie der Gedächtnisleistung: Diagnostik, Begutachtung und Therapie traumatischer Erinnerungen*. München: Reinhardt UTB.
- Volbert, R. (2005). Standards der psychologischen Glaubhaftigkeitsdiagnostik. In: H.-L. Kröber &

- M. Steller (Hg.), *Psychologische Begutachtung im Strafverfahren*, Springer, 171–203.
- Volbert, R., Schemmel, J., & Tamm, A. (2019). Die aussagepsychologische Begutachtung: eine verengte Perspektive?. *Forensische Psychiatrie, Psychologie, Kriminologie*, 13(2), 108–124.
- Vrij, A. (2008). *Detecting Lies and Deceit. Pitfalls and Opportunities*. Leicester, UK: Wiley.

Endnote Formeln

Die Umrechnung von Wahrscheinlichkeiten (Pr) auf Odds ratios (OR) ist nicht besonders schwer. Was als 50% Wahrscheinlichkeit beim Prozessausgang mit Münzwurf beziffert wird ($Pr=0.5$), ist eine Ausgangslage 1:1 in Odds. Hätte man beispielsweise eine vorsätzliche Tötung einer verheirateten Frau in ihrem Heim, würde die kriminologische Statistik eine erhöhte Häufigkeit für den Partner als Täter ergeben, z. B. 75% der Fälle. Die Odds für die Schuld des Ehemanns wären dann als Ausgangslage $OR_{\text{prior}} = 75:25 = 3:1 = 3$.

$OR(a) = \frac{Pr(a)}{1 - Pr(a)}$	$Pr(a) = \frac{OR(a)}{1 + OR(a)}$
Odds aus Wahrscheinlichkeit	Wahrscheinlichkeit aus Odds

Die sog. likelihood ratio ist ein Multiplikator der Odds, der sich als Zuwachs resp. Verminderung aus der Ausgangslage ergibt.

Zusammenfassung

Die Aktualisierung der aussagenpsychologischen Begutachtung betrifft die einseitige Unwahrrhypothese und das trügerische Widerlegenwollen derselbigen. Zeugenaussagen sollten besser von einer neutralen Ausgangsposition her mit Bayes'scher Logik analysiert und in beide Beweisrichtungen gewichtet werden. Die Zwischenschritte der Methode, das Nennen der Anknüpfungstatsachen, das Aufstellen von Hypothesen und die Auswertungsinstrumente haben sich hingegen langjährig bewährt.

Kontakt

henriette.haas@psychologie.uzh.ch

Grundlagen Kriminalistik

Rechtsmedizin

Grundwissen für die Ermittlungspraxis

Von Prof. Dr. med. Dr. phil. Ingo Wirth
und Prof. Dr. Andreas Schmeling.

4., neu bearbeitete Auflage 2020
XVII, 457 Seiten. Kartoniert. € 32,-
ISBN 978-3-7832-0058-4

Auch als ebook erhältlich

Für den Kriminalisten ist es wichtig, sich über Möglichkeiten und Grenzen der Rechtsmedizin zu informieren. Der Inhalt dieses Handbuchs orientiert sich an den Anforderungen der Ermittlungspraxis und umfasst die kriminalistisch bedeutsamen Aufgabengebiete der Rechtsmedizin.

Allgemeinverständlich beschrieben werden u.a. Sterben, Tod und Leichenuntersuchung, Ursachen gewalttätiger Todesfälle, tödliche Verkehrsunfälle, Vergiftungen, Identifizierung unbekannter Toter, körperliche Untersuchung lebender Personen sowie rechtsmedizinische Spurenanalyse, insbesondere DNS-Analytik.

