



Martin Lory

Dr. sc. techn. Dipl. El.-Ing. ETH,
DAS Applied Statistics

Experte für Schusswaffen und Brände
Leitender Wissenschaftler Forensisches Institut Zürich

Bulletin d'information de la
Informationsbulletin der
Bulletin d'informazione della
Newsletter of the

et/und/e/and

Chambre Suisse des experts judiciaires techniques et scientifiques
Schweizerischen Kammer technischer und wissenschaftlicher Gerichtsexperten
Camera svizzera degli esperti giudiziari tecnici e scientifici
Swiss chamber of Technical and Scientific Forensic Experts

Swiss Experts Certification SA (SEC)
Certification de personnes selon ISO 17024
Personenzertifizierung nach ISO 17024
Certificazione delle persone secondo la norma ISO 17024
Certification of persons according to ISO 17024

BEFUND, HYPOTHESE, DIAGNOSE – STANDARTISIERTE BEWERTUNGS- METHOTIK

Die evaluative Befundbewertung ermöglicht es, technische Befunde nicht nur zu beschreiben, sondern deren Aussagekraft unter klar formulierten Hypothesen transparent und ausgewogen zu bewerten. Der Artikel zeigt, weshalb eine standardisierte Bewertungsmethodik hilft, vorhandene Unsicherheit verständlicher zu kommunizieren.

Die technischen Zusammenhänge von Unfällen und Schäden zu klären, ist anspruchsvoll. Deshalb sollten in Zivil- und Strafverfahren Sachverständige¹ beigezogen werden, wenn zur Feststellung des Sachverhalts oder zur Beurteilung verschiedener Varianten des Sachverhalts spezielle Kenntnisse erforderlich sind. Die abschliessende Herausforderung zeigt sich erst beim Verfassen des Berichts oder Gutachtens: Wie lassen sich die technischen Feststellungen wissenschaftlich korrekt, nachvollziehbar und transparent darstellen? Wie sollen vorhandene Unsicherheiten beschrieben werden? Reicht eine Formulierung wie «ist zu vernachlässigen» und sind Ausdrücke wie «dürfte eher» oder «kann nicht gänzlich ausgeschlossen werden» angemessen? Für die Leserin oder den Leser des Berichts wird die Situation dadurch nicht einfacher. Selbst wenn die technischen Zusammenhänge verständlich beschrieben und anschaulich bebildert sind, bleiben häufig Unsicherheiten in der sprachlichen Bewertung: Was bedeutet etwa «eher wahrscheinlich», «nicht auszuschliessen» oder «vernachlässigbar» konkret?

¹ Siehe Art. 175, Art. 181 und Art 183 ff ZPO (SR 272, Zivilprozessordnung) und Art. 182 StPO (SR 312.0, Strafprozessordnung).

EXPERTENSUCHE

Die Mitglieder der Schweizerischen Kammer technischer und wissenschaftlicher Gerichtsexperten und die zertifizierten Expertinnen und Experten finden Sie mittels Stichwortsuche im Internet:

RECHERCHE D'EXPERTS

Vous pouvez trouver les membres de la Chambre suisse des experts judiciaires techniques et scientifiques ainsi que les experts certifiés à l'aide de mots clés aux adresses internet suivantes:

RICERCA ESPERTI

Può avvenire con l'inserimento di parole chiavi nel sito internet:

SEARCH FOR EXPERTS

Experts for a particular task can be found on the internet with the aid of keywords:

www.swiss-experts.ch
www.experts-certification.ch
www.international-experts.ch

Zieglerstrasse 29
CH-3007 Bern
T +41 31 838 68 68
office@swiss-experts.ch

IMPRESSUM

Redaktion: Schweizerische Kammer technischer und wissenschaftlicher Gerichtsexperten und Swiss Experts Certification SA.
Sekretariat: Zieglerstrasse 29, CH-3007 Bern, T 031 838 68 72. Empfänger: Zertifizierte Expertinnen und Experten, Mitglieder der Schweizerischen Kammer technischer und wissenschaftlicher Gerichtsexperten, Gericht, Versicherungen und andere interessierte Kreise.



BEFUND, HYPOTHESE, DIAGNOSE – STANDARDISIERTE BEWERTUNGSMETHODIK

Viele Fachleute tun sich mit diesen sprachlichen und methodischen Unsicherheiten schwer. Bei Formulierungen wie «mit an Sicherheit grenzende Wahrscheinlichkeit» und «beweiskräftig» bleibt ein Unbehagen: Sie können eine Sicherheit suggerieren, die durch die tatsächlich verfügbaren Feststellungen und das eigene Fachwissen nicht immer gegeben ist. Ein methodischer Ausweg ergibt sich über die Bayes'sche Denkweise, die in der forensischen Wissenschaft zunehmend an Bedeutung gewonnen hat. Unsicherheit lässt sich damit transparent, logisch konsistent und ausgewogen kommunizieren. Der Bayes'sche Ansatz hat meine Arbeit erleichtert.

Im Forensischen Institut Zürich ist die Befundbewertung bei vergleichenden (evaluativen) Untersuchungen seit 2020 standardmässig vorgesehen. Nicht darunter fallen reine Beschreibungen, Resultate von Messungen einschliesslich Messunsicherheit, Rekonstruktionen sowie die Bestimmung von Klassen oder Kategorien.

Der theoretische Ursprung dieser Befundbewertung liegt in der Bayes'schen Wahrscheinlichkeitstheorie. Der Pastor Thomas Bayes formulierte die Grundidee; Pierre-Simon Laplace entwickelte sie im späten 18. und frühen 19. Jahrhundert mathematisch weiter und machte sie für allgemeine Anwendungen nutzbar. Die mathematische Herleitung der Befundbewertung führt zu einer Formel, die für das Verständnis zentral ist.

Legende: $H_{1/2}$ sind die Hypothesen, E die Erkenntnis oder das Ergebnis, $P(x)$ die Wahrscheinlichkeit oder die Plausi-

bilität von x , das Symbol « $e|B$ » zeigt an, dass etwas unter der Bedingung von...ist.

Ein solcher Aktualisierungsprozess nach Bayes findet auch im Alltag statt: Wir haben eine vorläufige Überzeugung, erhalten neue Informationen und passen unsere Einschätzung an – zumindest idealerweise. Ein alltägliches Beispiel sind neue Informationen im Vorfeld von Abstimmungen. Häufig gewichten wir Informationen besonders stark, wenn sie unsere bisherige Überzeugung bestätigen.

Genau darin liegt die Gefahr des *Confirmation Bias*, der im fachlichen Kontext bewusst kontrolliert werden muss. Bei der fachlichen Wahrheitsfindung geht es darum, objektive Gegebenheiten – also spurekundliche Befunde (Feststellungen, Resultate, Beobachtungen) – zu erheben, die für die Hypothesen relevant sind und die eine Hypothese gegenüber einer anderen in einem bestimmten Mass stützen. Gemäss der Formel ist jedoch nicht die Wahrscheinlichkeit der Hypothesen zu bestimmen, sondern die Wahrscheinlichkeit der Befunde unter der einen und unter der anderen Hypothese. Da häufig keine passenden statistischen Daten vorliegen, müssen diese Wahrscheinlichkeiten fachlich begründet eingeschätzt werden. In diesem Zusammenhang sprechen wir von «Plausibilitäten».

Das Vorgehen lässt sich in mehrere Schritte gliedern:

- **Befundbewertung ja/nein:** Zunächst ist zu prüfen, ob eine Befundbewertung erforderlich ist. Oft erfordern Messungen, Feststellungen und deren sachliche Beschreibung keine Interpretation.

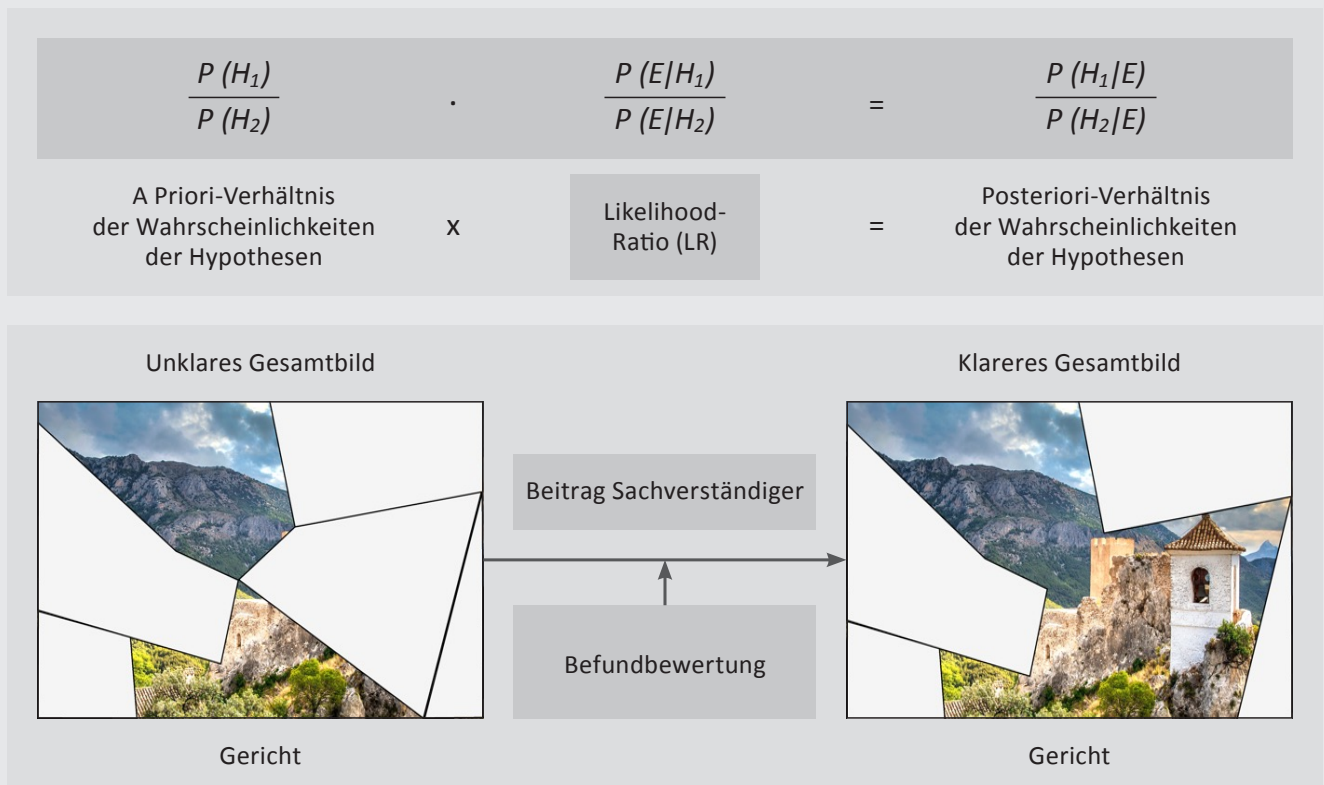


Abb. 1: Mathematische Formel der Befundbewertung und Bedeutung ihrer Terme

BEFUND, HYPOTHESE, DIAGNOSE – STANDARDISIERTE BEWERTUNGSMETHODIK

- **Befunde erarbeiten:** Die Grundlage jeder Befundbewertung sind gut dokumentierte und unverfälschte Befunde. Dazu gehören die Suche, Dokumentation und Sicherung von Spuren und Gegenständen sowie die genaue Dokumentation des angetroffenen Zustands, der durchgeführten Arbeiten, der Untersuchungen und der weiteren Abklärungen und allenfalls ein parteiöffentlicher Augenschein². Die Befunde sollen relevant sein und helfen, die gestellten Fragen zu beantworten. Wichtig ist, schon bei der Spurensuche möglichst objektiv vorzugehen und in diesem frühen, kritischen Stadium mehrere Arbeitshypothesen zu verfolgen.
- **Materialkritik:** Das sichergestellte Material sowie Fotos und Dokumente werden daraufhin geprüft, ob sie aussagekräftig und nicht kompromittiert sind. Zu klären ist insbesondere, ob sie durch unbekannte Personen, nachträgliche Veränderungen oder andere Einwirkungen in ihrer Aussagekraft beeinträchtigt wurden.
- **Hypothesenerstellung:** Auf Grundlage der Fragestellung – gegebenenfalls nach Rücksprache mit dem Auftraggeber – werden mindestens zwei sich gegenseitig ausschliessende Hypothesen formuliert. Die explizite Formulierung der Hypothesen ist Voraussetzung für eine transparente Befundbewertung. Ohne sie bleibt unklar, unter welchen Annahmen die Befunde beurteilt werden.
- **Bewertung der Befunde:** Entscheidend ist die Frage: Welche Befunde sind zu erwarten, unter der Bedingung, dass die jeweilige Hypothese zuträfe, und wie verhalten sich die tatsächlich festgestellten Befunde dazu? Stimmen sie gut, teilweise oder kaum mit den Erwartungen überein? Um die Plausibilität der Befunde unter der jeweiligen Hypothese zu bestimmen – also Zähler und Nenner des Likelihood Ratios – ist ein deduktiver Gedankenvorgang nötig. Vereinfacht gesagt handelt es sich um den Schluss von der angenommenen Ursache zur erwarteten Wirkung: von den Hypothesen zu den Befunden – zeitlich vorwärts gedacht.
- **Synthese mit Schluss-Satz:** In der Synthese werden die beiden Plausibilitäten ins Verhältnis gesetzt. Dieses Verhältnis wird als Likelihood Ratio (LR) bezeichnet. Es gibt an, wie viel plausibler die Befunde unter der einen Hypothese sind als unter der anderen. Der entscheidende Punkt ist: Bewertet wird nicht die Wahrscheinlichkeit

sicher ja	LR = ∞	Ausserhalb Skala 100% Sicherheit: Zuordnung
äusserst stark für	1'000'000 < LR	11 Stufen für die Befundbewertung. Alle Unsicherheiten sind enthalten. → LR ✓
sehr stark für	10'000 < LR < 1'000'000	
stark für	100 < LR < 10'000	
mässig stark für	10 < LR < 10	
leicht für	1 < LR > 1/10	
gleichermassen dafür/dagegen	LR = ca. 1	
leicht dagegen	1 > LR > 1/10	
mässig stark gegen	1/10 > LR > 1/100	
stark gegen	1/100 > LR > 1/10'000	
sehr stark gegen	1/10'000 > LR > 1/1'000'000	
äusserst stark gegen	1/1'000'000 > LR	
sicher nein	LR = 0	Ausserhalb Skala 100% Sicherheit: Ausschluss

2 Siehe Art. 168 und Art 181 ff ZPO (SR 272, Zivilprozessordnung) und Art. 193 StPO (SR 312.0, Strafprozessordnung).

Abb. 2: Skala, angelehnt an die Europäische Richtlinie zur Befundbewertung (ENFSI Guideline for evaluative reporting in forensic science, 2010, https://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf)

lichkeit der Hypothesen, sondern die Wahrscheinlichkeiten der Befunde unter den jeweiligen Hypothesen.

- **Verwendung des Resultats durch den Adressaten:** Das LR ist nicht die Entscheidung selbst. Es beschreibt lediglich die Unterstützung der Hypothesen anhand der erhaltenen Befunde im Verhältnis zweier Hypothesen. Auftraggeber, Gericht oder andere Entscheidungsträger müssen diese Gewichtung der Befunde mit weiteren Beweismitteln, rechtlichen Beweisanforderungen und der vorgängigen Einschätzung des Falles zusammenführen.

Befundbewertung: In vielen Fällen erscheinen die Befunde unter einer naheliegenden Ausgangshypothese zunächst gut erklärbar. Um neutral und ausgewogen zu bleiben, ist es deshalb entscheidend, eine tragfähige Alternativhypothese zu formulieren und die Befunde auch unter dieser ernsthaft zu prüfen. Eine Alternativhypothese verliert an Plausibilität, wenn sie mit den objektiven Befunden kaum vereinbar bleibt. Daraus ergibt sich ein hohes LR, das auf einer oberen Stufe auf der verbalen Skala anzusiedeln ist (siehe Abbildung 2) und eine starke Gewichtung der Befunde ausdrückt – zugunsten der Ausgangshypothese im Vergleich zur Alternativhypothese. In einzelnen Bereichen, insbesondere bei Materialspuren, stehen Datenbanken zur Verfügung, aus denen sich Häufigkeiten resp. Seltenheiten ableiten lassen. Sind diese Daten für den konkreten Fall repräsentativ, können daraus Schätzungen eines numerischen LR erfolgen. Der Vergleich wird bezüglich Quelle der Spur durchgeführt («source level»). Bei Fragestellungen auf Aktivitätsebene («activity level») ist das LR häufig weniger ausgeprägt.

Die Vorteile – und gleichzeitig die Herausforderungen – einer nachvollziehbaren Befundbewertung zeigen sich erst bei der konkreten Anwendung in der Fallarbeit. Der Schluss-Satz einer Befundbewertung kann beispielsweise lauten: *Die Befunde sprechen stark dafür [Stufe], dass der dokumentierte Schaden [Befund E] durch die einmalige mechanische Einwirkung [Hypothese 1] und nicht durch normale Abnutzung [Hypothese 2] entstanden ist.*

Keine Befundbewertung: Von einer Befundbewertung ist abzusehen, wenn die Befunde für die Fragestellung nicht relevant sind oder wenn die fachliche Grundlage für eine belastbare Befundbewertung fehlt. Wurden beispielsweise die falschen Spuren gesichert oder sind kaum Details zum Ablauf des Vorfalles bekannt, kann keine Befundbewertung erfolgen.

Datengrundlage: Für die Bewertung der Befunde müssen Erfahrungen der sachverständigen Person geeignete Fachliteratur, Datenbanken, selbst erhobene Daten oder Ergebnisse fallspezifischer Versuche zur Verfügung stehen; idealerweise stützt sich die Bewertung auf eine Kombination mehrerer solcher Grundlagen. Die Auswertung der Daten muss ebenfalls einer strengen Logik folgen: Soweit nume-

rische Methoden eingesetzt werden, müssen diese validiert und ihre Auswertungen kalibriert sein. Im forensischen Kontext bedeutet dies etwa, dass Methoden gemäss ISO/IEC 17025 validiert und numerische Bewertungen im Sinne der ISO 21043 kalibriert sein müssen. Dadurch wird vermieden, dass methodisch nicht vergleichbare Daten oder Bewertungsebenen miteinander vermengt werden.

Rekonstruktion: Bei der Rekonstruktion von Ereignissen werden Befunde genutzt, um mögliche Ursachen oder Abläufe zu entwickeln, also Arbeitshypothesen zu bilden. Um einseitige Deutungen zu vermeiden, sollten mehrere Hypothesen geprüft werden. Bereits die Spurensuche wird unter mehreren Hypothesen durchgeführt. Zu Beginn einer Untersuchung ist häufig noch unklar, welche technischen Abläufe, welche Ursachen oder welche Verantwortlichkeiten überhaupt in Betracht kommen. Da vieles noch nicht geklärt ist, müssen zunächst geeignete Arbeitshypothesen zu möglichen Ursachen oder Abläufen entwickelt werden. Dieser rückwärtsgerichtete Gedankenprozess wird als Abduktion bezeichnet: Aus beobachteten Wirkungen werden mögliche Ursachen abgeleitet. Der Prozess ähnelt der Diagnose, bei der aus Symptomen verschiedene mögliche Krankheiten hergeleitet werden.

Grenzen: Sachverständige Personen treffen fachliche Beurteilungen, fällen aber keine rechtliche (Gesamt-) Entscheidungen. Ihre Aufgabe besteht darin, Befunde zu erheben, fachlich zu bewerten und die dabei bestehenden Unsicherheiten transparent darzustellen. Sachverständige Personen sollen sich auf die für ihre fachliche Aufgabe relevanten Informationen beschränken. Diese Begrenzung schützt vor kontextbedingten Verzerrungen und unterstützt die fachliche Neutralität. Sowohl bei der Befundbewertung als auch bei der Rekonstruktion arbeiten wir mit mehreren Hypothesen, um die vorhandene Unsicherheit unter verschiedenen Blickwinkeln transparent aufzeigen zu können. Fallinformationen können von entscheidender Bedeutung für die Bewertung der Befunde sein. Ändern sich diese Fallinformationen, muss die Befundbewertung erneut durchgeführt werden.

Fehler: Der Begriff des «Fehlers» verdient eine genauere Betrachtung. Nicht jede abweichende fachliche Einschätzung ist ein Fehler. Wenn transparent, methodisch korrekt und unter Offenlegung der Unsicherheiten gearbeitet wurde, kann eine andere Fachperson aufgrund anderer Erfahrungen, Daten oder Gewichtungen zu einer anderen Schlussfolgerung gelangen. Von einem Fehler sollte erst dann gesprochen werden, wenn gegen bekannte Regeln, validierte Methoden, dokumentierte Standards oder erkennbare Sorgfaltspflichten verstossen wurde. Da die tatsächlichen Abläufe oft nicht vollständig bekannt sind, lassen sich unterschiedliche fachliche Schlussfolgerungen nicht immer eindeutig als richtig oder falsch klassifizieren. Entscheidend ist, ob die Schlussfolgerung methodisch nachvollziehbar und fachlich begründet ist.

3 Siehe dazu ausführlich HOFER F./GHELFI S./LORY M./ARNOLD J., (Re-)konstruierte Wahrheiten – Auswirkungen kognitiver Verzerrungen in der Polizeiarbeit und wirksame Gegenstrategien, in: Schweizerisches Polizeiinstitut (Hrsg.), Neuchâtel, format magazine no 11-2021, S. 64–71.



BEFUND, HYPOTHESE, DIAGNOSE – STANDARDISIERTE BEWERTUNGSMETHODIK

Bias in der Fallarbeit muss so weit wie möglich erkannt, kontrolliert und reduziert werden. Deshalb führen wir am Forensischen Institut Zürich seit dem Jahr 2019 regelmässig interdisziplinäre Weiterbildungen³ zu diesem Thema durch.

Künstliche Intelligenz (KI, AI): Auch in der Gutachtenserstellung gewinnt KI zunehmend an Bedeutung. Mithilfe fallrelevanter Datengrundlagen können Machine-Learning-Verfahren berechnete Likelihood Ratios erzeugen und damit als wertvolle Zweitmeinung, also als unabhängige numerische Überprüfung der Expertenbewertung dienen. Zur Absicherung sind mehrere validierte und kalibrierte statistische Modelle einzusetzen; zudem ist die Robustheit der Resultate zu überprüfen.

In der Fallarbeit setzen wir Large-Language-Modelle (LLMs) jedoch kaum ein. Das eigenständige Verfassen von Berichten ist ein wichtiger Teil des fachlichen Denk- und Reifungsprozesses. Beim Schreiben werden Hypothesen geschärft, Befunde geordnet, Unsicherheiten erkannt und Zwischenfolgerungen überprüft. Dieser Reifungsprozess benötigt Zeit und Geduld. Wird er abgekürzt, entsteht die Gefahr von Gedankensprüngen.

LLMs können in einem geschützten Rahmen bestehende Texte auf sprachliche Klarheit, innere Logik, Vollständigkeit und Konsistenz prüfen. Die Kunst des Promptings beschränkt sich dabei nicht auf Rechtschreibung und Grammatik. KI kann auch genutzt werden, um Arbeitsmethoden, Befundbewertungen, die Plausibilität von Zahlenwerten sowie die Formulierung von Grenzen kritisch zu prüfen. Beispielsweise können interne Vorgaben und externe Normen als Prüfkriterien in die Berichtskontrolle eingebunden werden. Dadurch lässt sich systematisch kontrollieren, ob ein Bericht die vorgesehenen Anforderungen an Struktur, Transparenz, Methodik und Einbezug der Unsicherheiten erfüllt. Noch stehen wir jedoch diesbezüglich am Anfang.

Als nützlich beim Prompting hat sich ein internes Handbuch zur Erstellung von Gutachten und Berichten erwiesen. Auch der vorliegende Text wurde mit KI verbessert – allerdings erst, nachdem er vollständig selbstständig verfasst worden war, und nur mit sorgfältiger menschlicher Nachkontrolle. Entscheidend bleibt: Der Mensch muss die fachliche Verantwortung behalten – **Human in command**.