

Diskussionen

Muss DNA-Evidenz schwer verständlich sein?

Der Ausweg aus einem Kommunikationsproblem

von Stefan Krauss und Ralph Hertwig

1. Wie »sicher« ist der genetische Fingerabdruck?

»Auf der Basis des DNA-Profiles der am Tatort an der Leiche gefundenen Blutspur erstellte der DNA-Sachverständige ein Gutachten. Dort gab er zu Protokoll, dass das DNA-Profil der DNA-Probe des Angeklagten mit dem DNA-Profil der am Tatort gefundenen Spur übereinstimmt. Zusätzlich führte er aus, dass in einer Stadt von der Größe Berlins ungefähr eine Million Männer als mögliche Täter in Frage kommen. Im Falle der Schuld des Angeklagten zeigt die verwendete DNA-Analyse mit an Sicherheit grenzender Wahrscheinlichkeit eine Übereinstimmung des DNA-Profiles des Angeklagten mit dem in der Blutprobe an der Leiche sichergestellten genetischen Fingerabdruck. Die Wahrscheinlichkeit, dass der genetische Fingerabdruck eines Unschuldigen *rein zufällig* mit dem DNA-Profil übereinstimmt, das in der Blutprobe an der Leiche gefunden wurde, beträgt 0,0001 %.«

Stellen Sie sich für einen Moment vor, Sie müssten die Aussagekraft dieser DNA-Übereinstimmung bewerten: Mit welcher Wahrscheinlichkeit ist der Angeklagte der Verursacher der am Tatort sichergestellten DNA-Spur?

In den Medien wird DNA-Evidenz häufig als ein äußerst aussagekräftiges Indiz dargestellt: Wird anhand einer DNA-Analyse eine Übereinstimmung des genetischen Profils eines Tatverdächtigen mit einer am Tatort hinterlassenen Blutspur gefunden, muss der Verdächtige auch schuldig sein. *Hubert Poche* von der Deutschen Gesellschaft für Rechtsmedizin wurde bezüglich einer solchen DNA-Übereinstimmung mit den Worten zitiert: »Da kommt keiner raus« (*Frankfurter Rundschau* 1998, Nr. 85, S. 28). Dem Präsidenten derselben Gesellschaft, *Bernd Brinkmann*, wurde die Aussage zugeschrieben, wenn sich der Täter in einer Gruppe von Getesteten befindet, bestünde »eine hundertprozentige Sicherheit, ihn zu identifizieren« (*Frankfurter Allgemeine Zeitung* 1998, Nr. 86, S. 13). Beide Aussagen fielen im Zusammenhang mit den gentechnischen Untersuchungen, die im Zuge der Ermittlungen über den Mord an der 11-jährigen Christina N. durchgeführt wurden. Die Polizei führte damals auf der Suche nach Christinas Mörder im Raum Cloppenburg eine bis dahin einmalige Genmassenuntersuchung von 18.000 potentiellen Tatverdächtigen durch.

Ist dieses Zutrauen in die Aussagekraft von DNA-Evidenz angemessen? Es gibt mehrere Gründe, warum DNA-Evidenz, wie jede andere Evidenz auch, kritisch hinterfragt werden sollte. Einer der Gründe hängt mit einem in der psychologischen Urteilsforschung häufig beobachteten Phänomen, der »Basisraten-Vernachlässigung« (*base rate neglect*), zusammen. Im Folgenden werden wir darstellen, wie die psychologische Forschung in den letzten Jahrzehnten gezeigt hat, dass Laien sowie Experten verschiedenster Profession statistische Informationen nur unzureichend verstehen. Ziel dieses Artikels ist es, einen Vorschlag zu machen, wie dieses unzureichende Verständnis von unsicherer Information vor Gericht – z. B. für die statistischen Informationen eines DNA-Gutachtens – überwunden werden könnte.

2. Basisraten-Vernachlässigung

Zur Erläuterung der Basisraten-Vernachlässigung kehren wir noch einmal zum Eingangsbeispiel (»Mord in Berlin«) zurück. Es gibt einen Mord, eine Blutspur am Tatort und einen der Tat verdächtigen Angeklagten. Weiterhin gibt es ein Sachverständigengutachten,

das eine Übereinstimmung des genetischen Profils des Angeklagten mit der am Tatort hinterlassenen Blutspur konstatiert und in dem folgende numerische Informationen gegeben wurden:

1. In Berlin gibt es ca. eine Million Personen, die als potentielle Spurverursacher in Frage kommen.
2. Wenn der Angeklagte tatsächlich der Spurverursacher ist, beträgt die Wahrscheinlichkeit, dass sein genetisches Profil mit Hilfe einer DNA-Analyse als übereinstimmend mit der am Tatort gefundenen Blutspur festgestellt wird, (nahezu) 100 %.
3. Die Wahrscheinlichkeit, dass der genetische Fingerabdruck eines Unschuldigen rein zufällig mit dem DNA-Profil übereinstimmt, das in der Blutprobe an der Leiche gefunden wurde, beträgt 0,0001 %.

Die erste Information gibt die so genannte »A-priori-Wahrscheinlichkeit« $p(V)$ an, wobei »p« für »probability« und »V« für das Ereignis »Angeklagter ist Verursacher« steht. Dabei ist $p(V)$ die Wahrscheinlichkeit, dass der Angeklagte vor Berücksichtigung irgendwelcher Indizien – also »a priori« – die Spur verursacht hat. Sie beträgt hier 0,0001 Prozent und entspricht hier der *Basisrate* der potentiellen Spurverursacher. Die zweite Information ist die Trefferrate des DNA-Verfahrens. Sie wird als $p(\text{DNA}^+ | V)$ geschrieben und beträgt in unserem Beispiel 100 Prozent¹. Die letzte Information schließlich ist die Falsch-Positiv-Rate des DNA-Tests. Sie lässt sich schreiben als $p(\text{DNA}^+ | \neg V)$ und beträgt hier 0,0001 Prozent².

Mit diesen Informationen lässt sich nun die Wahrscheinlichkeit, dass der Angeklagte nach Feststellung einer DNA-Übereinstimmung tatsächlich der Spurverursacher war, als $p(V | \text{DNA}^+)$ schreiben und mit der so genannten Bayes-Formel berechnen. Die Bayes-Formel erlaubt es allgemein, die Wahrscheinlichkeit einer Hypothese neu zu berechnen, wenn neue Daten, die die Hypothese betreffen, auftauchen. Interpretiert man Indizien als Daten und Einschätzungen über die Schuld eines Angeklagten als Hypothesen, wird der Wert der Bayes-Formel für Gerichtsurteile offensichtlich³. Nach dieser Formel gilt in unserem Fall:

$$\begin{aligned} p(V | \text{DNA}^+) &= \frac{p(\text{DNA}^+ | V)p(V)}{p(\text{DNA}^+ | V)p(V) + p(\text{DNA}^+ | \neg V)p(\neg V)} \\ &= \frac{100\% \cdot 0,0001\%}{100\% \cdot 0,0001\% + 0,0001\% \cdot 99,9999\%} = 50\% \end{aligned}$$

Der Angeklagte in unserem fiktiven Berliner Mordfall ist also lediglich mit 50-prozentiger Wahrscheinlichkeit auch tatsächlich der Spurverursacher. Intuitiv glauben aber die meisten von uns, dass diese Wahrscheinlichkeit viel höher ist. Wie kommt diese falsche Intui-

1 Das Symbol DNA+ steht dabei für das Ereignis »DNA-Übereinstimmung«. Ein Längsstrich zwischen zwei Ereignissen symbolisiert »bedingte« Wahrscheinlichkeiten und liest sich wie folgt: $p(A|B)$ ist die »Wahrscheinlichkeit von A unter der Bedingung, dass B bereits der Fall ist«.

2 Ein \neg vor einem Ereignis bedeutet dessen Negation. Die Rückübersetzung von $p(\text{DNA}^+ | \neg V)$ lautet demnach: Wahrscheinlichkeit einer DNA-Übereinstimmung unter der Bedingung, dass der Angeklagte *nicht* der Spurverursacher ist.

3 Die Bayes-Formel ist nicht nur eine unter vielen Formeln, sondern sie spielt eine Schlüsselrolle bei der Integration verschiedener Wahrscheinlichkeitsinformationen. Der ehemalige Vorsitzende des 6. Zivilsenats des OLG Stuttgart, *Rolf Bender*, plädierte in seinem Buch »Tatsachenfeststellung vor Gericht« für regelmäßige »Bayesianische Kontrollrechnungen« vor Gericht (*Rüßmann* 1990).

tion zu Stande? Dies wird gewöhnlich durch das Phänomen der Basisraten-Vernachlässigung erklärt. Die Wahrscheinlichkeit, dass das DNA-Profil eines Unschuldigen mit dem an der Leiche gefundenen Profil übereinstimmt, ist so gering (0,0001 %), dass man geneigt ist, die Möglichkeit einer zufälligen Übereinstimmung zu vernachlässigen. Je größer aber das Kollektiv der potentiellen Tatverdächtigen (hier: eine Million) ist, desto häufiger wird es vorkommen, dass neben dem Profil des Täters auch noch die Profile von tatsächlich Unschuldigen mit der am Tatort gefundenen Spur übereinstimmen. Mit anderen Worten: Die Treffer-Rate und die Falsch-Positiv-Rate alleine sagen wenig über die juristisch relevante Wahrscheinlichkeit $p(V | \text{DNA}')$ aus. Um diese zu erhalten, muss auf jeden Fall auch die »Basisrate« der potentiell Tatverdächtigen berücksichtigt werden.

In der Bayes-Formel wird die Basisrate durch den Term $p(V)$ berücksichtigt. Die Vernachlässigung der Basisrate bei intuitiven Schätzungen wurde in der psychologischen Literatur in einer Fülle von Aufgaben dokumentiert, und es konnte gezeigt werden, dass der Basisraten-Vernachlässigung Laien wie Experten verschiedenster Profession zum Opfer fallen (siehe z.B. *Bar-Hillel* 1990).

3. Können Menschen statistisch denken?

Die Basisraten-Vernachlässigung ist ein Beispiel für eine so genannte »kognitive Illusion«. Unter kognitiven Illusionen versteht man – in Analogie zu den stabilen visuellen Täuschungen – systematische Diskrepanzen zwischen menschlichen Urteilen und den Gesetzen der Wahrscheinlichkeitstheorie sowie denen der Logik. Seit den siebziger Jahren hat die psychologische Urteils- und Entscheidungsforschung eine Reihe solcher Diskrepanzen identifiziert (siehe z.B. *Kahneman, Slovic & Tversky* 1982). Diese Forschung erlangte deshalb so viel Aufmerksamkeit, weil diese Diskrepanzen schwerwiegende Folgen haben können. In der Tat macht es einen großen Unterschied, ob die Wahrscheinlichkeit, dass ein Angeklagter Verursacher einer am Tatort hinterlassenen Spur ist, auf 50 % oder auf fast 100 % geschätzt wird.

Die Botschaft, dass das statistische Denken der Menschen fehlerhaft ist, hat in den letzten Jahren auch Eingang in die juristische Fachliteratur gefunden. *Cass Sunstein*, Professor an der Universität Chicago, präsentiert in Beiträgen in *The University of Chicago Law Review* (1997) und in *Stanford Law Review* (*Jolls, Sunstein & Thaler* 1998) seinen Fachkollegen die experimentell-psychologische Forschung zu »kognitiven Illusionen« und verbindet damit folgende Befürchtung:

»People make judgments about probability on the basis of heuristic devices, response perhaps to high costs of inquiry and decision, that work in many cases but that tend to lead to systematic errors. [This] also has implications for the jury system – suggesting that juries are likely to make many mistakes in terms of probability assessments and that correction of those mistakes is a large task of the legal system.« (*Sunstein* 1997, S. 1187)

In Deutschland machte *Steinke* (1994) darauf aufmerksam, dass die richterliche Intuition oftmals ein »irreführendes Hilfsmittel« sein kann. Er schreibt: Es »steht völlig außer Frage, daß Menschen, also auch Richter, angesichts von Ungewißheit ziemlich irrationale Schlußfolgerungen ziehen. Vor allem die Einschätzung von Wahrscheinlichkeit durch eine bestimmte Person, und darauf kommt es in Gerichtsverfahren an, kann von vielen unterschiedlichen Faktoren, auch psychologischen, beeinflusst sein«.

Petra Halder-Sinn diskutiert die Gefahr von Urteilsfehlern in Sachverständigengutachten, die vor Gericht Verwendung finden. In ihren Augen »besteht keine allzu große Hoffnung, Urteilsfehler durch rasch erlernbare und leicht vermeidbare Vermeidungsstrategien in den Griff zu bekommen« (*Halder-Sinn* 1993, S. 49). Sie bezieht sich dabei auf einschlä-

gige Literatur zum Thema »Strategien zur Vermeidung von Urteilsfehlern« (Brehmer 1986; Nisbett & Wilson 1977).

Wir teilen die pessimistische Einschätzung von Halder-Sinn nicht. Die Ansicht, dass Menschen *grundsätzlich* nicht statistisch denken können, greift unserer Meinung nach zu kurz, denn sie lässt außer Acht, dass kognitive Strategien an ein bestimmtes Repräsentationsformat von Information angepasst sind. Die Frage, ob Menschen statistisch denken können oder nicht, kann man nicht beantworten, ohne zu berücksichtigen, wie die gegebenen statistischen Informationen repräsentiert sind. Im Folgenden erläutern wir diese Argumentation im Detail.

4. Verschiedene Repräsentationsformate statistischer Informationen

Ein Taschenrechner hat einen Algorithmus für Multiplikation. Dieser verarbeitet Zahlen – aber Zahlen können in verschiedenen Formaten repräsentiert werden, z.B. als arabische, römische oder binäre Zahlen. Die Algorithmen üblicher Taschenrechner sind auf das arabische System eingestellt. Würde man in einen solchen Taschenrechner binäre Zahlen eintippen, bekäme man unsinnige Ergebnisse. Aus den unsinnigen Ergebnissen würde aber niemand den Schluss ziehen, dass der Taschenrechner keinen Algorithmus für Multiplikation hat. Er hat einen, aber eben einen, der auf arabische Zahlen eingestellt ist. Wie die Algorithmen eines Rechners sind auch die kognitiven Strategien von Menschen auf bestimmte Formate eingestellt. Welches Format von statistischer Information erlaubt es Menschen, statistisch zu denken?

In Lehrbüchern der Statistik wird statistische Information in Form von Wahrscheinlichkeiten und Prozenten präsentiert. Ist das ein Repräsentationsformat, auf das unsere kognitiven Strategien eingestellt sind? Gigerenzer & Hoffrage (1995) argumentieren, dass Menschen im Laufe der Evolution Strategien für den Umgang mit einer unsicheren Welt ausgebildet haben, und dass diese Strategien nicht auf Wahrscheinlichkeiten und Prozentwerte eingestellt sind, da diese erst wenige Jahrhunderte existieren. Sie argumentieren, dass kognitive Strategien an natürliche Häufigkeiten angepasst sind, wie Menschen sie – in einer Zeit ohne Bücher und Statistik – beobachten konnten.

Stellen Sie sich zum Beispiel einen Richter in einer Gesellschaft vor, die keine Bücher und keine Statistiken kennt. Er hat vor einiger Zeit eine Art Lügendetektortest entwickelt, dessen Logik der moderner polygraphenbasierter Lügendetektion ähnelt. Die Psychologie seines Tests besteht in der Annahme, dass Menschen, die lügen, lediglich ein Symptom zeigen, nämlich Erröten. Natürlich ist er am so genannten diagnostischen Wert seines Lügentests interessiert, d. h. er möchte gerne wissen, mit welcher Wahrscheinlichkeit er vom Erröten einer Person auf das Vorliegen einer Lüge schließen kann. Über die Zuverlässigkeit seines »Lügendetektortests« hat er bis jetzt die folgenden Erfahrungen gesammelt: Er untersuchte bereits 100 Personen mit diesem Verfahren, von denen 10 logen. Von diesen 10 erröteten 9 tatsächlich. Allerdings gab es auch unter den 90 Personen, die nicht logen, 27, die dennoch erröteten. Nun untersucht er eine neue Person: Sie errötet gleichfalls. Wie hoch ist nun die Wahrscheinlichkeit, dass sie tatsächlich gelogen hat?

Um diese Wahrscheinlichkeit zu bestimmen, benötigt dieser Richter keine komplizierten Berechnungen mit Wahrscheinlichkeiten. Er braucht lediglich die beobachtete Anzahl der Menschen, die erröteten und die tatsächlich logen (nämlich 9), durch die Anzahl aller Menschen, die insgesamt erröteten (9 + 27) zu dividieren. Man sieht nun sofort (siehe *Abbildung 1*, rechts unten), dass der diagnostische Wert des Lügentests miserabel ist: Die Wahrscheinlichkeit, dass eine Person, die errötet, tatsächlich gelogen hat, beträgt lediglich 25 %. Auf der linken Seite von *Abbildung 1* steht dieselbe Aufgabe, wie sie in der heute

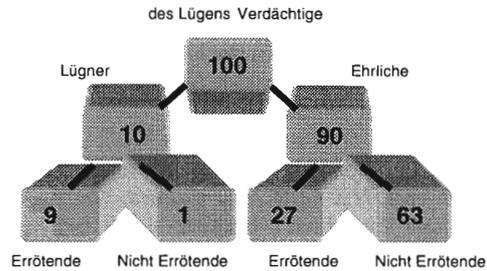
üblichen Wahrscheinlichkeits-Nomenklatur in einem Statistiklehrbuch erscheinen würde; auf der rechten Seite werden die »natürlich« beobachteten Häufigkeiten in einem Baumdiagramm visualisiert:

Abbildung 1

»Wahrscheinlichkeiten«

a priori: $p(\text{„Lügen“})$	= 10%
$p(\text{„Erröten Lügen“})$	= 90%
$p(\text{„Erröten Ehrlich“})$	= 30%

»Natürliche Häufigkeiten«



$$p(\text{„Lügen“} \mid \text{„Erröten“}) =$$

$$= \frac{90\% \cdot 10\%}{90\% \cdot 10\% + 30\% \cdot 90\%} = 25\%$$

$$\frac{9}{9 + 27} = 25\%$$

Erst wenn wir die statistische Information in das Wahrscheinlichkeitsformat übersetzen (siehe *Abbildung 1*, links), kommen Bezeichnungen wie »A-priori-Wahrscheinlichkeit« oder »bedingte Wahrscheinlichkeit« ins Spiel: Die zu den Beobachtungen äquivalenten Wahrscheinlichkeitsaussagen sind: $p(\text{„Lügen“}) = 10\%$, $p(\text{„Erröten|Lügen“}) = 90\%$ und $p(\text{„Erröten|Ehrlich“}) = 30\%$. Die Wahrscheinlichkeit, dass eine Person, die errötet, auch gelogen hat, also $p(\text{„Lügen|Erröten“})$, ergibt nun mit der Formel von Bayes ebenfalls 25%.

Interessanterweise ist erst mit den in Statistiklehrbüchern üblichen Wahrscheinlichkeiten und Prozenten die Gefahr der Basisraten-Vernachlässigung gegeben. Im Falle der natürlichen Häufigkeiten können die »Basisraten« (a priori 10 Lügner und 90 Ehrliche) gar nicht vernachlässigt werden, da sie als Ausgangspunkt für die Beobachtung des Errötens genommen wird (siehe *Abbildung 1*, Baumdiagramm): Obwohl 90% der Lügner tatsächlich erröten und nur 30% der Ehrlichen erröten, bedeutet das nicht, dass es sich bei einer errötenden Person mit großer Wahrscheinlichkeit um einen Lügner handelt. Warum nicht? Das Häufigkeitsformat liefert die Antwort: Obwohl 90%, also 9 der 10 Lügner erröten, und nur 30%, also 27 der 90 Ehrlichen erröten, gibt es insgesamt immer noch mehr Ehrliche, die erröten, als Lügner (27 gegenüber 9). Vernachlässigt man die Basisrate, überschätzt man die Güte des »Lügendetektors«. Der zentrale Punkt dieses Beispiels ist, dass kognitive Strategien für »Bayesianische« Aufgaben einfacher sind, wenn die relevanten Informationen in natürlichen Häufigkeiten statt in Wahrscheinlichkeiten oder Prozentwerten vorgegeben werden. Das liegt daran, dass im Häufigkeitsformat für eine Bayesianische Berechnung weniger Rechenschritte nötig sind (siehe *Abbildung 1*). Wir sind davon überzeugt, dass diese Einsicht helfen kann, Menschen, die von Berufs wegen mit Wahrscheinlichkeiten zu tun haben, die Arbeit zu erleichtern. Wir werden im Folgenden zeigen, dass beispielsweise Geschworene, Rechtsanwälte, Richter und Gutachter tatsächlich von diesen Überlegungen profitieren können.

5. Lassen sich diese Erkenntnisse auf reale juristische Kontexte übertragen?

Lindsey, Hertwig & Gigerenzer (1999) testeten 127 Jurastudenten und 27 wissenschaftliche Angestellte an der juristischen Fakultät der Freien Universität Berlin (im Folgenden bezeichnen wir beide Populationen als Juristen). Jeder Teilnehmer an der Studie erhielt detaillierte Fallbeschreibungen zweier Mordfälle, die jeweils ein DNA-Gutachten beinhalteten. Die Juristen hatten eine Stunde Zeit, sich in die zwei Fälle einzulesen, und sollten danach verschiedene Fragen zu den DNA-Gutachten beantworten. Zum Beispiel die Frage, die im Eingangsbeispiel schon gestellt wurde: »Mit welcher Wahrscheinlichkeit ist der Angeklagte der Verursacher der am Tatort sichergestellten DNA-Spur?« Auf der Grundlage der verfügbaren Evidenz (im Wesentlichen lediglich des DNA-Gutachtens) wurden die Juristen anschließend gebeten, zu einem Urteilspruch zu gelangen (»schuldig« vs. »nicht schuldig«).

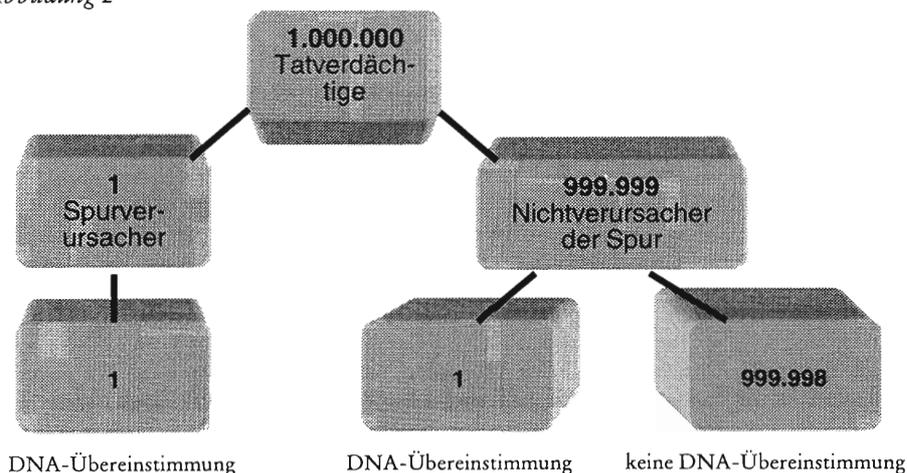
In einem der beiden Mordfälle wurden die statistischen Informationen des DNA-Gutachtens in Form von Wahrscheinlichkeiten gegeben, im anderen Fall waren die Informationen im DNA-Gutachten in Form von Häufigkeiten formuliert. Für das eingangs geschilderte Gutachten (»Berliner Mordfall«) lautet die Häufigkeitsformulierung zum Beispiel (Lindsey 1998):

»Auf der Basis des DNA-Profiles der am Tatort an der Leiche gefundenen Spur erstellte der Sachverständige ein Gutachten. Dort gab er zu Protokoll, dass in einer Stadt von der Größe Berlins ungefähr eine Million Männer als mögliche Täter in Frage kommen. Im Falle der Schuld des Angeklagten zeigt die verwendete DNA-Analyse mit an Sicherheit grenzender Wahrscheinlichkeit die Übereinstimmung des DNA-Profiles des Angeklagten mit dem in der Blutprobe an der Leiche sichergestelltem genetischen Fingerabdruck auf. Unter den weiteren 999.999 Tatverdächtigen wird wahrscheinlich einer zu finden sein, bei dem der genetische Fingerabdruck *zufällig* mit dem DNA-Profil übereinstimmt, das in der Blutprobe an der Leiche gefunden wurde. Im vorliegenden Fall stimmt das DNA-Profil der DNA-Probe des Angeklagten mit dem DNA-Profil der am Tatort gefundenen Spur überein.

Bei wie vielen der 1.000.000 Tatverdächtigen ist das der Fall? Wie groß ist also die Wahrscheinlichkeit, dass es sich bei dem Angeklagten aufgrund der DNA-Übereinstimmung um den Verursacher der Spur handelt?«

Übersetzt man diese Informationen in einen Häufigkeitsbaum, ergibt sich folgende Struktur:

Abbildung 2



In dieser Visualisierung der Häufigkeitsangaben kann man sehr gut sehen: Unter den eine Million potentiellen Tatverdächtigen gibt es insgesamt *zwei*, deren DNA-Profil mit dem an der Leiche gefundenen übereinstimmt. Demnach ist die Wahrscheinlichkeit, dass *einer* der potentiell Tatverdächtigen der Spurverursacher ist, wenn sein DNA-Profil darauf hinweist, 50 %.

Wie kamen die Juristen in unserer Untersuchung mit den verschiedenen Informationsformaten zurecht? Waren die statistischen Informationen im DNA-Gutachten in Wahrscheinlichkeiten ausgedrückt, konnten nur etwa 8 % der Jurastudenten die gefragten Wahrscheinlichkeiten korrekt einschätzen. Dasselbe Gutachten, in dem die Informationen in Häufigkeiten gegeben wurden, führte zu knapp 40 % korrekten Urteilen⁴. Hatten die natürlichen Häufigkeiten auch einen Einfluss auf das Urteil der Juristen? Ja, denn es zeigte sich: Wurden beim DNA-Gutachten die Informationen in Form von Wahrscheinlichkeiten gegeben, waren mehr als die Hälfte (55 %) der Juristen von der Schuld des Angeklagten überzeugt. Wurden – bei ansonsten identischer Fallschilderung – im DNA-Gutachten die Informationen in Häufigkeiten dargeboten, plädierten *bei gleicher Sachlage* nur noch 33 % der Juristen auf »schuldig«. Diese Ergebnisse sind alarmierend, wenn man bedenkt, dass die Juristen mit Wahrscheinlichkeitsinformationen zum größten Teil (92 %) gar keine korrekten Schlussfolgerungen aus den statistischen Informationen des DNA-Gutachtens ziehen konnten.

Die Beurteilung von Wahrscheinlichkeiten durch Experten kann also deutlich verbessert werden, ohne dass diese notwendigerweise in statistischem Denken unterrichtet werden müssen. Durch die Darstellung der statistischen Information eines DNA-Gutachtens in Häufigkeiten übernimmt das Format der Information einen Teil der Denkarbeit, und Fehleinschätzungen können so vermieden oder doch wenigstens reduziert werden.

6. Training von statistischem Denken

Wie wir sahen, kann allein durch die Übersetzung eines DNA-Gutachtens in Häufigkeiten *ohne ein statistisches Training* das Verständnis eines Gutachtens deutlich verbessert werden (40 % an Stelle von 7 % korrekten Urteilen). Durch ein Training von statistischem Denken, welches (im Gegensatz zu Statistiklehrbüchern) das bessere Verständnis von Häufigkeitsformaten und die Visualisierung von Häufigkeiten in Baumdiagrammen ausnutzt, können sogar noch deutlichere Verbesserungen erzielt werden.

Statistisches Denken wird im Mathematikunterricht nur begrenzt gelehrt. Die Formel von Bayes, eine der zentralen Methoden zur Einschätzung von Unsicherheit, scheint aber auch im Jurastudium kaum zu existieren. Basierend auf der These, dass natürliche Informationsverarbeitung an natürliche Häufigkeiten adaptiert ist, haben *Sedlmeier & Gigerenzer* (1999) eine effektive Methode entwickelt, statistisches Denken zu lernen. Während traditionelle Methoden lehren, Wahrscheinlichkeiten in die Bayes-Formel einzusetzen, lehrt *Sedlmeiers* (1997) computergestütztes Tutorsystem, Unsicherheiten und Risiken in Form von natürlichen Häufigkeiten zu repräsentieren. In mehreren Studien haben *Sedlmeier & Gigerenzer* (1999) die Effizienz beider Verfahren verglichen: Der unmittelbare Lernerfolg war etwa doppelt so hoch beim Lernen mit Hilfe von Häufigkeiten wie beim traditionellen Lernen der Bayes-Formel. Noch bedeutsamer ist aber, dass der

4 In den Häufigkeitsversionen wurden tatsächlich nur die Wahrscheinlichkeitsangaben in Häufigkeiten übersetzt; den Juristen wurden dabei *keine* Häufigkeitsbäume vorgesetzt, sie hatten also durchaus noch mathematische Operationen durchzuführen. Wir vermuten, dass eine zusätzliche Präsentation der Bäume zu einem noch bedeutend höheren Prozentsatz richtiger Lösungen führen kann.

Lernerfolg beim Erlernen von Häufigkeitsrepräsentationen von Dauer ist. Bekanntlich werden mathematische Formeln sehr schnell wieder vergessen. Nicht so bei den Personen, denen beigebracht wurde, Unsicherheiten und Risiken in Form von natürlichen Häufigkeiten zu repräsentieren: Die Leistung lag fünf Wochen nach dem Training immer noch bei 90 %, wohingegen in der anderen Gruppe (die ein Training mit der Formel von Bayes absolviert hatte) nur noch 15 % der Aufgaben fünf Wochen später gelöst werden konnten. Das bedeutet, dass es gute Argumente gegen den Pessimismus von *Halder-Sinn* (1993) gibt, die befürchtet, dass »keine allzu große Hoffnung [besteht], Urteilsfehler durch rasch erlernbare und leicht vermittelbare Vermeidungsstrategien in den Griff zu bekommen«.

7. Resümee

Der Autor der »Zeitmaschine«, *H.G. Wells*, war überzeugt, dass statistisches Denken für moderne Gesellschaften genauso wichtig ist wie die Fähigkeit zu lesen und zu schreiben (siehe *Huff & Geis* 1954). Mit modernen Technologien – von HIV-Tests bis hin zu DNA-Analysen – lassen sich oftmals nur mit Unsicherheit behaftete Aussagen machen. Wir brauchen deshalb eine adäquate Psychologie, um diese Unsicherheit auch zu verstehen. Die hier skizzierten Studien und Methoden können uns, den Nutzern dieser Technologien, helfen, dabei einen großen Schritt voranzukommen.

Literatur

Bar-Hillel, M. (1990). Back to Base Rates, in: *R.M. Hogarth* (ed.), *Insights in Decision Making*. Chicago: University of Chicago Press. – *Brehmer, B.* (1986). In One Word: Not from Experience, in: *H. Arkes & K. Hammond* (eds.), *Judgment and Decision Making* (pp. 705–719). Cambridge, England: Cambridge University Press. – *Gigerenzer, G. & Hoffrage, U.* (1995). How to Improve Bayesian Reasoning Without Instruction: Frequency Formats. *Psychological Review* 102, pp. 684–704. – *Halder-Sinn, P.* (1993). Fehlerhafte Urteilsheuristiken in Sachverständigengutachten. *MschrKrim* 76, S. 44–49. – *Huff, D. & Geis, I.* (1954). *How to Lie with Statistics*. New York: Norton. – *Jolls, C., Sunstein, C.R. & Thaler, R.* (1998). A Behavioral Approach to Law and Economics. *Stanford Law Review* 50, pp. 1471–1550. – *Kahneman, D., Slovic, P. & Tversky, A.* (eds.) (1982). *Judgment under Uncertainty: Heuristics and Biases*. Cambridge, England: Cambridge University Press. – *Lindsey, S.* (1998). The Expert Between Science and Law. Paper Presented at the 1998 Schöffmann Seminar »The Expert in Modern Societies«. – *Lindsey, S., Hertwig, R. & Gigerenzer, G.* (1999). Communicating Statistical Evidence. Manuskript zur Veröffentlichung eingereicht. – *Nisbett, R.E. & Wilson, T.D.* (1977). Telling More Than We Can Know: Verbal Reports on Mental Processes. *Psychological Review* 84, pp. 231–259. – *Rießmann, H.* (1990). Entscheidungen. *Zeitschrift für Zivilprozess* 103, Heft 1, S. 62–76. – *Sedlmeier, P.* (1997). BasicBayes: A Tutor System for Simple Bayesian Inference. *Behavior Research Methods, Instruments, & Computers* 29, pp. 328–336. – *Sedlmeier, P. & Gigerenzer, G.* (1999). Teaching Bayesian Reasoning in Less than Two Hours. Manuskript zur Veröffentlichung eingereicht. – *Steinke, W.* (1994). Der Beweiswert forensischer Gutachten. *NStZ*, Heft 1, S. 17–21. – *Sunstein, C.R.* (1997). Behavioral Analysis of Law. *The University of Chicago Law Review* 64, pp. 1175–1195.

(Anshr. d. Verf.: Wiss. Mitarb. *Stefan Krauss* und *Ralph Hertwig*, Max-Planck-Institut für Bildungsforschung, Lentzeallee 94, ABC, 14195 Berlin; e-mail: Krauss@mpib-berlin.mpg.de; Hertwig@mpib-berlin.mpg.de)