

Autor: Daniel Effer-Uhe
Dokumenttyp: Aufsatz
Literaturnachweis: GVRZ 2025, 100-108 **JURIS**

Quelle: 
Verlag Dr. Otto Schmidt, Köln

Fundstelle: GVRZ 2025, 100-108
Zitiervorschlag: Effer-Uhe, GVRZ 2025, 100-108

Stochastik und Beweiswürdigung Zugleich Rezension zu Mark Schweizer, Beweiswürdigung und Beweismaß - Rationalität und Intuition, Mohr Siebeck, Tübingen 2015 (Habil., St. Gallen 2014)

Prof. Dr. Daniel Effer-Uhe^{*}

I. Rationale Überzeugungsbildung vor Gericht?

[1]

Mark Schweizer, der bereits mit seiner Dissertation „Kognitive Täuschungen vor Gericht - eine empirische Studie“¹ ein ertragreiches Werk aus dem Grenzbereich von Rechtswissenschaft und Psychologie vorgelegt hatte, hat sich auch mit seiner 2015 erschienenen Habilitationsschrift in den interdisziplinären Bereich begeben, wenn auch der Schwerpunkt diesmal eher auf theoretischen und insbesondere stochastischen Fragen und weniger auf empirischer Forschung liegt. Er hat sich zum Ziel gesetzt, einerseits darzulegen, wie Gerichte rational zu einer Überzeugung von der Wahrheit oder Unwahrheit bestrittener Tatsachenbehauptungen gelangen können, andererseits zu beschreiben, wie die Entscheidungsfindung tatsächlich erfolgt, und in diesem Rahmen auch einen begründeten Vorschlag für ein angemessenes Beweismaß im Zivilprozess zu machen (S. 2 f.). Gerade seine Überlegungen zu einer rationalen Überzeugungsbildung sind aber über den Zivilprozess hinausgehend auch für andere Prozessrechtssparten von Interesse.²

II. Wahrheit und Wahrscheinlichkeit

[2]

Die Arbeit gliedert sich in fünf Hauptteile und eine thesenartige Zusammenfassung als sechsten Teil. Im ersten Teil, überschrieben mit „Grundlagen“, befasst *Schweizer* sich mit den Grundlagen der Überzeugungsbildung. Der Teil beginnt mit Überlegungen zum für die gerichtliche Tatsachenfeststellung angemessenen Wahrheitsbegriff, als den er - in argumentativer Auseinandersetzung mit Kohärenztheorien und der Konsensustheorie der Wahrheit - den Wahrheitsbegriff der Korrespondenztheorie der Wahrheit identifiziert (S. 30 ff.), wonach eine Aussage dann wahr ist, wenn sie mit der Wirklichkeit übereinstimmt. Anhand der Anastasia-Formel des BGH³ legt *Schweizer* dar, dass nach der Rspr. die persönliche Überzeugtheit nicht unumstößlich sein muss, sondern nur ein für das praktische Leben brauchbarer Grad an Gewissheit zu fordern ist: Erforderlich für das Regelbeweismaß ist also nicht eine bestimmte objektive Wahrscheinlichkeit der Übereinstimmung der Behauptung mit der Wirklichkeit,

sondern eine bestimmte subjektive Überzeugtheit des Gerichts (S. 35). Damit sind allerdings Wahrscheinlichkeitserwägungen keineswegs aus dem Prozess der Überzeugungsbildung ausgeschlossen: Die subjektive Überzeugung des Gerichts ist an Denkgesetze und Erfahrungssätze gebunden, das erforderliche Beweismaß bestimmt den Grad der persönlichen Überzeugung, der erforderlich ist, damit eine beweisbedürftige Tatsache als erwiesen angesehen werden darf. Die Lehre von der formellen Wahrheit hält *Schweizer* mit überzeugender Begründung für durch die Abschaffung der gesetzlichen Beweistheorie überholt (S. 43).⁴

[3]

Im zweiten Teil, „Die Denkgesetze der Beweiswürdigung“, befasst er sich mit der Frage nach den logischen Grundlagen der Beweiswürdigung. Zunächst stellt er sich die Frage, was genau unter den Denkgesetzen, gegen die die Beweiswürdigung nicht verstoßen darf, zu verstehen ist (S. 83 ff.). Die Bedeutung der deduktiven Logik hält er nachvollziehbar für gering, da es bei der Beweiswürdigung nicht um die deduktive Ableitung notwendigerweise wahrer Folgerungen aus wahren Prämissen geht, sondern vielmehr um die Ableitung plausibler Folgerungen aus unsicheren Prämissen (S. 87 ff.). Entscheidend seien also die Regeln der induktiven Logik - und damit die Regeln der Wahrscheinlichkeitstheorie. Damit *Schweizer* zur Anwendbarkeit der Regeln der Wahrscheinlichkeitstheorie gelangen kann, muss er zunächst darlegen, dass die richterliche Überzeugung von der Wahrheit einer Tatsachenbehauptung auf einer Wahrscheinlichkeitsannahme beruht bzw. beruhen sollte. Dazu befasst er sich zunächst mit der Frage, von welchem Wahrscheinlichkeitsbegriff auszugehen ist (S. 89 ff.). Im Einzelnen geht er zunächst auf den klassischen Wahrscheinlichkeitsbegriff nach Laplace ein, der als Anhänger eines epistemischen Wahrscheinlichkeitsbegriffs aufgrund eines deterministischen Weltbilds Wahrscheinlichkeit als das Verhältnis der günstigen Fälle zu allen möglichen Fällen definiert, dafür aber voraussetzen muss, dass alle Fälle gleichmöglich sind. Diese Gleichmöglichkeit ist allerdings, wie *Schweizer* anmerkt, letztlich eine *petitio principii* - in vielen Fällen ist keineswegs selbstverständlich, dass alle Möglichkeiten gleich wahrscheinlich sind, man denke nur an eine Verteilung auf einem Kontinuum, bei dem es üblicherweise nicht zu einer Gleich-, sondern vielmehr zu einer Normalverteilung in Form einer Glockenkurve kommt. Den frequentistischen Wahrscheinlichkeitsbegriff (Wahrscheinlich

- 100 -

Effer-Uhe, GVRZ 2025, 100-108

- 101 -

keit als Grenzwert der relativen Häufigkeit eines Ereignisses bei theoretisch unendlich vielen Wiederholungen) lehnt er für die Zwecke der Beweiswürdigung ab, da er keine Aussage über die Wahrscheinlichkeit eines Einzelfalls zulässt, auf den es aber im Zivilprozess gerade ankomme. Objektive Wahrscheinlichkeitsbegriffe kämen ebenfalls als Bezugspunkt der richterlichen Überzeugung nicht in Betracht, weil Wahrscheinlichkeit nur eine Eigenschaft des Wissens eines Beobachters über die Welt ist, keine Eigenschaft der beobachteten Welt selbst. Im Ergebnis versteht *Schweizer* Wahrscheinlichkeiten als subjektive, epistemische Wahrscheinlichkeiten, also als Maß einer persönlichen Überzeugung bzw. als Überzeugungsgrad, Wahrscheinlichkeitstheorie also als Logik der Überzeugungsgrade (S. 124 ff.). Eine Wahrscheinlichkeit ist danach immer eine bedingte Wahrscheinlichkeit: eine Wahrscheinlichkeit bei gegebenem Wissen. Die Überzeugungen eines Gerichts müssen, wie *Schweizer* darlegt, den Axiomen der Wahrscheinlichkeitstheorie gehorchen, wenn sie nicht gegen grundlegende Anforderungen an eine rationale Überzeugungsbildung verstoßen sollen, was eine nachvollziehbare Begründung dafür, wie das Gericht zu seiner Überzeugung gelangt ist, ausschließt.

[4]

Ein Überzeugungsgrad lässt sich, wie *Schweizer* überzeugend darlegt, immer als Wert auf einer Skala von 0 bis 1 (oder von 0 % bis 100 %) ausdrücken (S. 129 f.),⁵ wobei *Schweizer* einräumt, dass dieser Wert durchaus schwer bestimmbar sein kann. Erst die numerische Beschreibung der Wahrscheinlichkeitsüberzeugung ermögliche eine Überprüfung, ob die Annahmen des Gerichts den Axiomen der Wahrscheinlichkeitstheorie genügen. Als Hauptvorteil eines numerischen Ausdrucks von Wahrscheinlichkeiten im Rahmen der Beweiswürdigung identifiziert er völlig richtig, dass er zu einem rationaleren Ergebnis führt als eine intuitive Würdigung. Er veranschaulicht das am Vergleich mit der Schätzung der Grundfläche einer Wohnung, die man einerseits in einem einzigen Schritt durch die Schätzung der Gesamtfläche vornehmen kann, andererseits gestaffelt durch eine Einschätzung der Längen und Breiten der einzelnen Zimmer und eine darauf aufbauende Berechnung der Grundflächen der einzelnen Zimmer. Die zweite Methode werde, so *Schweizer*, regelmäßig zu besseren Ergebnissen führen, auch wenn die Längen und Breiten der einzelnen Zimmer immer noch geschätzt werden müssen, weil kein Metermaß vorhanden ist.

III. Bayes-Theorem und Anfangswahrscheinlichkeit

[5]

Wenn man Überzeugungsgrade, wie sie zur Erfüllung der Anastasia-Formel erforderlich sind, als mathematische Wahrscheinlichkeiten interpretiert (und dazu gibt es keine sinnvolle Alternative, auch wenn das keineswegs heißt, dass man bei der Bewertung einer bestimmten Wahrscheinlichkeit praktisch sinnvoll zwischen 83,5 % und 83,6 % unterscheiden kann), dann ist die Frage, wie sich die Wahrscheinlichkeitsüberzeugung im Lichte neuer Indizien ändern sollte, mithilfe des Bayes-Theorems zu beantworten (S. 132 ff.).⁶ Dass man nicht intersubjektiv zwingend angeben kann, mit welcher Wahrscheinlichkeit ein Zeuge die Wahrheit sagt, sondern die Wahrscheinlichkeit subjektiv geprägt ist (Richter 1 wird möglicherweise einen Überzeugungsgrad von 95 % erreichen, während Richter 2 nur auf 85 % kommt), ist allerdings kein Argument gegen eine Nutzung derartiger mathematischer Methoden, sondern vielmehr nur Ausdruck der freien richterlichen Beweiswürdigung - es geht nur um die Kohärenz der verschiedenen Überzeugungen, nicht um die Richtigkeit der Überzeugung als solche (S. 154 f.).

[6]

Eine Schwierigkeit bei der Anwendung des Bayes-Theorems liegt darin, dass man eine Anfangswahrscheinlichkeit (oder Ex-ante-Wahrscheinlichkeit) angeben muss, die in die Formel einfließt, und unter Berücksichtigung des Beweiswerts neuer Indizien eine Endwahrscheinlichkeit (oder Ex-post-Wahrscheinlichkeit) begründet. Diese Anfangswahrscheinlichkeit festzulegen, ist mit erheblichen Schwierigkeiten verbunden, und in vielen Fällen kann sie nur sehr grob geschätzt werden. Um ein Beispiel aus meiner Vorlesung „Psychologie für Juristen“ zu verwenden: Wenn ein Mord auf einem Kreuzfahrtschiff verübt wird, auf dem sich 2.000 Personen befinden, die als potentielle Täter in Betracht kommen, beträgt die Anfangswahrscheinlichkeit ohne weitere Hinweise, wer von diesen Personen die Tat begangen hat, für jeden $\frac{1}{2000}$. Wenn der Mord aber z.B. in Köln stattgefunden hat, gibt es keine zwingende Möglichkeit, zu einer bestimmten Anfangswahrscheinlichkeit zu kommen: Käme der Täter aus Köln (einer Stadt mit etwa 1 Mio. Einwohnern), würde die Anfangswahrscheinlichkeit etwa $\frac{1}{1 \text{ Mio.}}$ betragen; er könnte aber auch aus anderen Teilen NRW kommen, so dass etwa 18 Mio. Menschen in Betracht

kämen (Anfangswahrscheinlichkeit $\frac{1}{18 \text{ Mio.}}$), aus ganz Deutschland ($\frac{1}{82 \text{ Mio.}}$) oder auch aus dem Ausland.

[7]

Kommt man über das Problem der Anfangswahrscheinlichkeit hinweg, müssen in die Formel weiterhin bedingte Wahrscheinlichkeiten eingestellt werden: Wie wahrscheinlich ist das Auftreten der Indizatsache (z.B. eines passenden Teilfingerabdrucks bei einer zufällig ausgewählten Person)? Wie wahrscheinlich ist das Auftreten der Indizatsache, wenn die Haupttatsache (Täterschaft des Verdächtigen) gegeben ist? Manche der benötigten Wahrscheinlichkeiten können im Einzelfall sehr plausibel abgeschätzt werden: Z.B. gibt es in der kriminalistischen Literatur Überlegungen dazu, mit welcher Wahrscheinlichkeit bei einer zufällig ausgewählten Person eine bestimmte Anzahl an Übereinstimmungen mit einer daktyloskopischen Spur vorliegt.⁷ Andere Einschätzungen sind mit sehr viel größeren Unsicherheiten behaftet: Wenn ein Zeuge bekundet, der Täter habe eine tiefe Stimme gehabt, kann man nicht sinnvoll angeben, dass genau 10 % der Bevölkerung eine „tiefe“ Stimme haben - ebenso könnte man 5 % oder 15 % für plausibel halten, schon weil die Maßstäbe darüber, was als „tief“ wahrgenommen wird, auseinandergehen. Das ändert aber nichts daran, dass der Einsatz des Bayes-Theorems, also die Schätzung mit anschließender mathematischer Berechnung der Endwahrscheinlichkeit, einen nicht zu unterschätzenden Gewinn an Rationalität der Beweiswürdigung mit sich bringt. Denn während man den Beweiswert einzelner Indizien für sich genommen noch einigermaßen plausibel abschätzen kann (wenn auch oft innerhalb eines gewissen Korridors vertretbarer Einschätzungen), tun sich Menschen schwer damit, intuitiv den Gesamt

- 101 -

Effer-Uhe, GVRZ 2025, 100-108

- 102 -

beweiswert mehrerer unabhängiger Indizien abzuschätzen, hier kommt es tendenziell zu Unterschätzungen des Beweiswerts (S. 161). Außerdem besteht eine Tendenz, einfach zu quantifizierende Beweismittel wie ein DNA-Gutachten gegenüber schwer quantifizierbaren Beweismitteln wie einer Zeugenaussage überzugewichten (S. 189). Kontrolliert man seine intuitive Einschätzung, indem man eine plausible Anfangswahrscheinlichkeit zum Ausgangspunkt nimmt und dann für jedes einzelne Indiz schätzt, wie verbreitet es ist, wenn die Haupttatsache vorliegt, und wie verbreitet es ist, wenn die Haupttatsache nicht vorliegt, und die Werte rechnerisch verknüpft, kann das vor intuitiven Fehlschlüssen bewahren. Allerdings wird kaum ein Jurist die sperrige Formel des Bayes-Theorems als geeignet für die alltägliche Arbeit bezeichnen - hier stellt ein Wahrscheinlichkeitsbaum, den man auf die beiden relevanten Äste zusammenstreicht, eine für die meisten Fälle gangbare Alternative dar⁸ - wenn sie auch gegenüber den später von *Schweizer* vorgestellten Bayes-Netzen in sehr komplexen Fällen den Nachteil haben, dass sich anhand von ihnen die Kohärenz verschiedener Teilüberzeugungen nicht überprüfen lässt. Damit können sie nicht ausschließen, dass einer gerichtlichen Entscheidung widersprüchliche Überzeugungen zugrunde gelegt werden. Für die allermeisten - auch komplexen - Fälle genügt dagegen ein Wahrscheinlichkeitsbaum ohne weiteres, und einen Rationalitätsgewinn gegenüber einem rein intuitiven Vorgehen stellt er in jedem Fall dar.

IV. Der Anscheinsbeweis im Licht des Bayes-Theorems

[8]

Dass das Bayes-Theorem nicht nur eine mathematische Spielerei ist, zeigt sich, wenn man es - wie *Schweizer* (S. 177 ff.) - auf die umstrittene Frage der Natur des Anscheinsbeweises anwendet: Wenn ein Erfahrungssatz so stark ist, dass schon allein durch den Erfahrungssatz in Verbindung mit den bekannten Indizatsachen die Schwelle des erforderlichen Beweismaßes für die Haupttatsache überschritten wird, dann sind weitere Beweiserhebungen für den Beweis der Haupttatsache entbehrlich (S. 177). Damit unterscheidet sich der Anscheinsbeweis nur graduell von sonstigen Indizienbeweisen, bei denen die Beweismaßgrenze im Hinblick auf die Haupttatsache noch nicht überschritten ist (S. 178), so dass weitere Beweismittel zu erheben sind. Gibt es dagegen Aspekte, die gegen einen „typischen“ Geschehensablauf im konkreten Einzelfall sprechen, kann das den Anscheinsbeweis schon dann erschüttern, wenn dadurch die Überzeugung unter das zu erreichende Beweismaß sinkt: Schon dann, wenn die Überzeugung leicht hinter dem erforderlichen Beweismaß zurückbleibt, sind - soweit vorhanden - weitere Beweismittel zu erheben (S. 178). Mit *Schweizer* kann man den „typischen“ Geschehensablauf beim Anscheinsbeweis also als einen „plausiblen“ Geschehensablauf interpretieren, der durch seine Plausibilität die Beweismaßgrenze bereits erreicht (S. 180): Entscheidend ist, dass „sich für Dritte nachvollziehbar begründen lässt, warum die Sachverhaltsdarstellung des Klägers gestützt auf die feststehenden Tatsachen und die Lebenserfahrung des Richters bereits ohne Abnahme weiterer Beweismittel die Überzeugung begründet, dass die behauptete Ursache die tatsächliche Ursache des beobachteten Erfolgs ist“ (S. 180). Die Bezeichnung als „Typizität“ ist insofern mindestens irreführend - in der Praxis werden längst nicht nur besonders typische Abläufe mit Hilfe des Indizienbeweises gelöst,⁹ und durch eine ausreichend generalisierende Beschreibung lässt sich eine vermeintliche Typizität auch in wenig typischen Fällen herbeischreiben.¹⁰ Fehlende Typizität bedeutet also nichts weiter, als dass die behauptete Verknüpfung von Ursache und Wirkung nicht so plausibel ist, dass sie schon für sich genommen die volle richterliche Überzeugung zu begründen vermag (S. 181). Konsequenz dieser von *Schweizer* sehr überzeugend angenommenen Interpretation des Anscheinsbeweises müsste dann allerdings auch eine Ausweitung seines Anwendungsbereiches sein: Warum man nicht auch bei individuellen Willensentschlüssen¹¹ in geeigneten Fällen eine hinreichende Plausibilität annehmen kann, ist in der Tat schwer zu begründen - und konsequent wäre es dann auch, sogar im Strafrecht eine Überschreitung des erforderlichen Beweismaßes aufgrund der angenommenen Plausibilität zu ermöglichen (was praktisch durchaus vorkommt, aber nicht offen so dargestellt und als „Anscheinsbeweis“ betrachtet wird).

[9]

Allerdings birgt diese Vorgehensweise erhebliche Probleme bei einer Häufung bestimmter, für sich genommen sehr seltener Vorkommnisse in großen Populationen. Als Beispiel mag der Verdacht von Tötungsdelikten, begangen durch Krankenpflegepersonal, dienen, der sich allein an einer „auffälligen“ Häufung von Todesfällen in der Schicht ein- und derselben Krankenschwester entzündet.¹² So ist es extrem unwahrscheinlich, dass in den Schichten eines bestimmten (!) Krankenpflegers rein zufällig 10 Todesfälle in einem Zeitraum vorkommen, in dem in Schichten ohne seine Beteiligung kein einziger Patient stirbt. Nehmen wir an, dass er in 25 % der Schichten im Dienst ist, und die Todesfälle sich zufällig über sämtliche Schichten verteilen - dann beträgt die Wahrscheinlichkeit, dass der nächste Todesfall auf seiner Station während seiner Schicht auftritt, $\frac{1}{4}$, die Wahrscheinlichkeit, dass die nächsten beiden Todesfälle bei ihm auftreten, $(\frac{1}{4})^2 = \frac{1}{16}$, und die Wahrscheinlichkeit, dass die nächsten 10 Todesfälle bei ihm auftreten, $(\frac{1}{4})^{10} = \frac{1}{1.048.576}$, also ungefähr 1:1 Mio. Das klingt erstmal nach einem überwältigenden Beweis für seine kausale Beteiligung an den Todesfällen. Geht man allerdings davon aus, dass etwa 1 Mio. Menschen in Deutschland in der Krankenpflege arbeiten, dann wird deutlich, dass es rein zufällig immer wieder vorkommen wird, dass einzelne Personen eine solche Serie von Todesfällen in ihren Schichten haben. Die Lösung kann nur in einer sehr behutsamen Schätzung

sowohl der Anfangswahrscheinlichkeit (falls man das Indiz „Anzahl der Todesfälle“ schon bei der Anfangswahrscheinlichkeit berücksichtigt) bzw. der Beweiskraft des Indizes liegen: Gerade weil derartige Fälle in einer ausreichend großen Population schon rein statistisch immer wieder einmal vorkommen (und dann auch mit einer gewissen Wahrscheinlichkeit auffallen und untersucht werden), kann man - im Zivilrecht wie im Strafrecht -

- 102 -

Effer-Uhe, GVRZ 2025, 100-108

- 103 -

nicht einfach von einer hinreichenden Plausibilität ausgehen, um allein aus der Häufung auf die Kausalität des Verhaltens des „Todesengels“ zu schließen. Anderenfalls unterläge man dem sog. „Zielscheibenfehler“ (Texas Sharpshooter Fallacy¹³), der dann auftritt, wenn man aus einer rein zufälligen Häufung von Ereignissen in einem kleinen Ausschnitt des Gesamtbilds auf einen kausalen Zusammenhang schließt. (Hinter der Bezeichnung „Texas Sharpshooter Fallacy“ steht die Idee eines Schützen, der rein zufällig eine große Anzahl an Schüssen auf ein Scheunentor abgibt und dann um die Stelle, an der die meisten Geschosse getroffen haben, nachträglich eine Zielscheibe malt. In einer großen Anzahl an Datenpunkten wird es immer vereinzelt Häufungen geben, die nur auf den ersten Blick „überzufällig“ erscheinen, weil man die Gesamtzahl der Datenpunkte nicht berücksichtigt.)

[10]

Eine Sonderstellung des Anscheinsbeweises gegenüber sonstigem Indizienbeweis ergibt sich aus seiner Revisibilität (S. 185). Das erklärt *Schweizer* plausibel mit der besonderen Gefahr, die davon ausgeht, dass ein Gericht allein aufgrund von Plausibilitätserwägungen, die nur auf dem - mehr oder weniger verlässlichen - Alltagswissen der Richter beruhen, Sachvortrag als erwiesen ansieht.

[11]

Für die richterliche Tätigkeit besonders wichtig ist damit also Bayes'sche Mathematik, deren Anwendbarkeit *Schweizer* praktisch anhand der Modellierung von Bayes-Netzen darlegt. (Eine ausführliche Anleitung zur Erstellung eines Bayes-Netzes gibt *Schweizer* auf Seite 112 ff. Allerdings hätte die Arbeit im Hinblick auf Bayes-Netze erheblich an Anschaulichkeit gewinnen können, wenn auch zu den Abfrageergebnissen zu *Schweizers* einfachem Beispielfall (S. 227-231) eine Grafik abgedruckt worden wäre.) Diese Bayes-Netze lassen sich mit Hilfe von zum Teil frei verfügbaren¹⁴ Computerprogrammen auch bei komplexen Sachverhalten mit etwas Übung unproblematisch und ohne großes mathematisches Vorwissen abbilden. Derartige Netze könnte ein Gericht zur Selbstkontrolle dahingehend einsetzen, ob seine Wahrscheinlichkeitsüberzeugungen tatsächlich miteinander kompatibel sind (S. 187). Der Vorteil dieser Netze liegt darin, dass die Abhängigkeit von Indizien und die versteckten Annahmen über anfängliche Überzeugungen in diesem Netz offengelegt werden müssen. Statistische Denkfehler lassen sich vermeiden, wenn man probabilistische Zusammenhänge mittels Bayes-Netzen modelliert, außerdem können Wahrscheinlichkeiten berechnet werden, für deren auch nur überschlägige Abschätzung bei einer Vielzahl von Indizien unser Arbeitsgedächtnis nicht ausreicht (S. 190). Als praxistauglich wird ein Großteil der Juristen den Umgang mit Bayes-Netzen wohl eher nicht ansehen - ernsthafte Hoffnung, dass sich Bayes-Netze flächendeckend in der Praxis durchsetzen könnten, besteht nicht. Für einen enormen Rationalitätsgewinn, der auch unter Praxisbedingungen ohne weiteres möglich ist, könnte dagegen nach Einschätzung des Rezensenten eine sehr viel einfachere Modellierung komplexer Fälle durch Wahrscheinlichkeitsbäume sorgen.¹⁵

[12]

Den Einwand, dass eine formale Modellierung nichts bringe, wenn das Resultat stark abhängig von den zugrunde liegenden Wahrscheinlichkeitsannahmen ist, die zum Teil ohne empirische Basis getroffen werden, nimmt *Schweizer* vorweg und entkräftet ihn mit dem zutreffenden Hinweis, dass man ohne derartige Annahmen keine rationale Entscheidung treffen kann - und es dann besser ist, sie offenzulegen (und damit auch der [Selbst-]Kritik zu öffnen) und ihren Einfluss auf die Berechnung mathematisch zu kontrollieren (S. 248).

V. Die Psychologie der Überzeugungsbildung

[13]

Im dritten Teil geht es um die „Psychologie der Überzeugungsbildung“. Hier geht *Schweizer* auf verbreitete Fehlschlüsse ein. Zentral ist für ihn, dass man den Zusammenhang zwischen einem Indiz und einer zu beweisenden Haupttatsache nur sinnvoll einschätzen kann, wenn sowohl die Häufigkeit des gemeinsamen Auftretens von Haupttatsache und Indiz, die Häufigkeit des Indizes, wenn die Haupttatsache nicht vorliegt, die Häufigkeit der Haupttatsache ohne das Indiz und die Häufigkeit, dass die Haupttatsache nicht gegeben ist, wenn das Indiz fehlt, bekannt sind: Fehlt eine dieser Informationen oder wird die Wahrnehmung der Häufigkeiten systematisch verzerrt, komme es zu Fehlurteilen (S. 262). Das wird in der Praxis häufig nicht beachtet, was *Schweizer* an einem Beispiel veranschaulicht: So sei es in Luzern zu einem Fall eines möglichen Kindesmissbrauchs gekommen, bei dem der Beschuldigte die Tat bestritten hat. Ein Gutachter habe festgestellt, dass das Bestreiten jeder Schuld typisch für eine therapiebedürftige pädophile Neigung sei. Tatsächlich wäre aus der Erfahrung, dass ein fehlendes Geständnis für Pädophile typisch ist (wenn sie denn überhaupt zutrifft), nur dann ein Verdachtsmoment abzuleiten, wenn nicht Pädophile typischerweise pädophile Taten gestehen würden, wenn es also bei nicht pädophilen Tatverdächtigen häufiger zu Geständnissen käme als bei tatsächlich Pädophilen.¹⁶ Entscheidend ist der sog. Likelihood-Quotient (Wahrscheinlichkeit der Indiztatsache, wenn die Haupttatsache vorliegt, geteilt durch Wahrscheinlichkeit der Indiztatsache, wenn die Haupttatsache nicht vorliegt), der sich aber ohne Wissen oder Annahmen darüber, wie häufig die Indiztatsache (hier: Bestreiten) ohne die Haupttatsache (Täterschaft) vorkommt, nicht berechnen lässt. Im Beispiel des pädophilen Tatverdächtigen ist sogar davon auszugehen, dass der Nicht-Täter normalerweise erst recht nicht gestehen wird, so dass das fehlende Geständnis sogar eher ein (wenn auch sehr schwaches) Indiz gegen die Täterschaft ist.

[14]

Ein wichtiger Hinweis *Schweizers* geht dahin, dass Richter im Hinblick auf ihre Beweiswürdigung kaum Möglichkeiten haben, ihre Fähigkeiten zur intuitiven Beurteilung von Beweisen durch praktische Erfahrung zu verbessern, weil die richterliche Berufstätigkeit mangels regelmäßiger und eindeutiger Rückmeldung darüber, ob der Richter in vergangenen Fällen mit seiner Einschätzung richtig lag, eine „böartige“ Lernumgebung bilde (S. 270). Mit anderen Worten: Es gibt keinen guten Grund anzunehmen, dass ein Richter durch langjährige Berufserfahrung tatsächlich seine Intuition bei der Würdigung von Beweisen nennenswert verbessern könnte.

[15]

Ausführlich befasst *Schweizer* sich mit der Frage, wie intuitive Kohärenzbildung („konstruktive Intuition“) vonstattengeht

(S. 272 ff.). Verbreitet wird vertreten, dass die Kohärenzbildung aufgrund der Verbindung vorliegender Beweismittel zu einer stimmigen Geschichte erfolge (narrative Kohärenz). *Schweizer* weist darauf hin, dass dieses Vorgehen die Gefahr mit sich bringt, nicht nur die Geschichte den Beweismitteln, sondern auch die Interpretation der Beweismittel der Geschichte anzupassen. Diese Befürchtung deckt sich mit Ergebnissen der psychologischen Forschung (kognitive Dissonanz, Bestätigungsfehler).¹⁷ Überdies könne, wie *Schweizer* unter Verweis auf empirische Untersuchungen von Bennett/Feldman¹⁸ darlegt, eine stimmige, aber schlecht belegte Geschichte einer besser belegten, aber schlechter strukturierten Geschichte vorgezogen werden.

[16]

Demgegenüber gehen Theorien kognitiver Kohärenz davon aus, dass der menschliche Verstand anstelle inkohärenter mentaler Repräsentationen kohärente Repräsentationen konstruiere (S. 285). Der Prozess der Kohärenzmaximierung sei nach Parallel-Constraint-Satisfaction-Modellen (PCS) ein Optimierungsproblem - Kohärenz die möglichst gute gleichzeitige Erfüllung von Restriktionen zwischen verschiedenen Elementen (S. 285). Ein zentraler Punkt, der in PCS-Modellen vielfach empirisch belegt wurde, besteht darin, dass es zu Kohärenzverschiebungen während des Entscheidungsprozesses kommt, dass sich also die Interpretation von für die Entscheidung wesentlichen Aussagen während des Entscheidungsfindungsprozesses (!) ändere und damit die Entscheidung kausal beeinflusst (S. 296). So neigten in einer Untersuchung von *Simon/Snow/Read*¹⁹ Versuchspersonen, die den Verdächtigen freisprechen wollten, tendenziell dazu, schon im Lauf ihres Entscheidungsfindungsprozesses entlastenden Aussagen zunehmend stärker zuzustimmen - allgemeiner ausgedrückt: Die Beurteilung der Indizien bewegt sich schon während des Prozesses der Beweiswürdigung in Richtung der Entscheidung, die gerade getroffen wird. Tatsächlich gibt es empirische Hinweise darauf, dass die Kohärenzverschiebungen schon die Entscheidung beeinflussen, sie also nicht erst der Rechtfertigung einer schon getroffenen Entscheidung dienen (S. 297 ff.). Klassische Vorstellungen von kognitiver Dissonanz, wie sie von Festinger begründet wurden, lassen sich mit diesem Befund nicht vereinbaren, da sie von einer Anpassung erst nach der Entscheidungsfällung ausgehen. Besonders problematisch für die Rationalität richterlicher Entscheidungsfindung (nicht nur bezüglich der Beurteilung von Tatsachen-, sondern auch von Rechtsfragen!) ist der Befund, dass auch die Beurteilung von Aussagen, die rational betrachtet nichts miteinander zu tun haben, in dieselbe, zur Entscheidung passende Richtung verschoben wird (S. 300). Diese Kohärenzverschiebungen erfolgen in der Regel unbewusst, der Entscheider ist also subjektiv der Auffassung, seine Entscheidung rational aufgrund objektiver Aspekte getroffen zu haben.

[17]

Aus diesem empirischen Befund leitet *Schweizer* im Anschluss an Glöckler/Engel²⁰ die Vermutung ab, dass ein höheres Beweismaß unter Umständen nur zu einer stärkeren Kohärenzverschiebung statt zu einer niedrigeren Quote an überzeugten Richtern führen könnte (S. 307). Seine weitere Vermutung, dass die Kohärenzverschiebungen sich bei einer holistischen Würdigung mehrerer Indizien vermeiden ließen, wenn der Sachverhalt anhand eines Bayes-Netztes modelliert wird, überprüft *Schweizer* empirisch, indem er dieselben Probanden einen Sachverhalt zunächst holistisch hat würdigen lassen (also abgefragt hat, welche Wahrscheinlichkeit für eine bestimmte Ursache - im konkreten Fall: die Täterschaft einer bestimmten Person bei einem Diebstahl oder einer Unterschlagung - spricht), danach den

Beweiswert der einzelnen Beweismittel anhand ihres Likelihood-Quotienten hat bewerten lassen. Davon ausgehend hat *Schweizer* dann rechnerisch ermittelt, welche Endwahrscheinlichkeit aufgrund der Likelihood-Quotienten eigentlich rechnerisch angemessen wäre (S. 311 ff.); im Anschluss an die Abfrage der Likelihood-Quotienten wurden die Probanden dann noch einmal nach der holistischen Wahrscheinlichkeit gefragt. Tatsächlich ergaben sich zwischen den holistischen A-posteriori-Wahrscheinlichkeiten sowohl vor als auch nach Abfrage der Likelihood-Quotienten zwischen denjenigen, die den Verdächtigen für schuldig hielten („Verurteiler“) und denjenigen, die ihn für unschuldig hielten („Freisprecher“) signifikante Unterschiede. Der Unterschied zwischen beiden Gruppen fällt aber weitaus geringer aus, wenn man die A-posteriori-Wahrscheinlichkeit mit Hilfe des Bayes-Netzes berechnet. Die intuitive, holistische Bewertung weicht also insoweit erheblich von der eigentlich rechnerisch angemessenen Bewertung ab, was Zweifel an der Rationalität der in der Praxis allgemein üblichen weitgehend intuitiven Beweiswürdigung angezeigt erscheinen lässt. Das Within-subjects-Design der Untersuchung, bei der dieselben Probanden sowohl nach der holistischen Würdigung als auch nach den einzelnen Likelihood-Quotienten befragt wurden, erklärt sich daraus, dass es *Schweizer* gerade darum ging, mögliche Unterschiede zwischen einer holistischen und einer errechneten Beweiswürdigung bei denselben Personen festzustellen. Spannend wäre insoweit aber auch ein Between-Subjects-Design (verbunden mit der von *Schweizer* bei einer anderen Untersuchung [S. 574 ff., insbesondere S. 582 f.] verwendeten Rangordnungsmethode), bei dem die Werte von unterschiedlichen Personengruppen abgefragt werden - denn wenn die Kohärenzverschiebungen schon während der Überzeugungsbildung erfolgen, ist die Annahme naheliegend, dass die Bewertung der Likelihood-Quotienten schon durch die vorherige Abfrage der Täterschaftswahrscheinlichkeit verzerrt ist.

VI. Erfahrungssätze

[18]

Den vierten Teil bildet eine Auseinandersetzung mit der Bedeutung von „Erfahrungssätze[n] in der Beweiswürdigung“. Ausgangspunkt des vierten Teils ist die Feststellung, dass durch Bayes-Netze zwar die innere Kohärenz der verschiedenen richterlichen Teilüberzeugungen gesichert werden kann, damit aber noch nicht sichergestellt ist, dass diese Teilüberzeugungen auch mit der Wirklichkeit übereinstimmen: Auch eine Summe völlig absurder Überzeugungen kann in sich kohärent sein, wie *Schweizer* an einem Beispiel zur frühneuzeitlichen Hexenverfolgung verdeutlicht (S. 349 f.). Zur Binnenrationalität sollte also zusätzlich auch noch eine Verankerung in der Wirklichkeit hinzukommen. Die Annahme einer bestimmten Wahrscheinlichkeit eines Ereignisses sollte daher grundsätzlich der relati

- 104 -

Effer-Uhe, GVRZ 2025, 100-108

- 105 -

ven Häufigkeit entsprechen, mit der Aussagen gleicher Art in der Vergangenheit wahr gewesen sind (Frequency Principle), was wiederum voraussetzt, dass man definiert, wann Aussagen zur „gleichen Art“ gehören (Problem der richtigen Referenzklasse). Dieses Referenzklassenproblem ist für die Rechtsanwendung - schon bei der Festlegung von Anfangswahrscheinlichkeiten - von elementarer Bedeutung: Der Schluss von einer relativen Häufigkeit auf die subjektive Wahrscheinlichkeit ist nur dann sinnvoll, wenn für alle Mitglieder der Referenzklasse die gleiche Wahrscheinlichkeit besteht. Beim o.g. Beispielsfall des Mordes auf dem Kreuzfahrtschiff kann man, wenn wir keine näheren Anhaltspunkte haben, die für einen bestimmten Täter sprechen, von einer grundsätzlich gleichen Anfangswahrscheinlichkeit für alle Menschen auf dem Schiff ausgehen. (Schon das ist allerdings stark vereinfacht: So wird es vielleicht unter den 2.000 Personen 50 Kleinkinder geben, die aus Altersgründen nicht in

Betracht kommen; bei den etwas älteren Kindern ist die Täterschaft zwar nicht ganz ausgeschlossen, aber wohl doch ex ante sehr viel unwahrscheinlicher als bei den Erwachsenen auf dem Schiff.) Verlegt man den Mordfall auf das Festland, wird die Gruppe potentieller Täter sehr inhomogen: Wir können bei einer Tat in Köln ex ante nicht wissen, ob der Täter Kölner, Nordrhein-Westfale, Deutscher oder auch nur Europäer ist. Trotzdem wird man ohne weitere Hinweise die Wahrscheinlichkeit eines Täters aus Köln für sehr viel größer halten als die Wahrscheinlichkeit eines Täters aus Palermo, obwohl Köln und Palermo eine vergleichbare Einwohnerzahl haben. (Wüssten wir über das Opfer, dass es selbst aus Palermo stammt, würde sich diese Bewertung möglicherweise umkehren.) Generell kann man zwar mit dem Referenzklassenproblem so umgehen, dass man verschiedene Unterklassen bildet, die in sich möglichst homogen sind, und dann für jede der einzelnen Unterklassen eine Wahrscheinlichkeit annimmt, die dem Frequency Principle für diese Unterklasse entspricht. Diese Bildung von Unterklassen hilft aber dann nicht weiter, wenn sich die einzelnen Mitglieder der Gesamtpopulation nicht prinzipiell, sondern nur graduell unterscheiden (wie es z.B. bei der Täterschaftswahrscheinlichkeit zwischen einem Kölner, einem Leverkusener, einem Düsseldorfer etc. der Fall ist). Aber auch bei prinzipiellen Unterschieden zwischen verschiedenen Untergruppen ist das Problem der richtigen Referenzklasse nicht trivial zu lösen: Oft kommen nämlich mehrere Kriterien zur Abgrenzung der Referenzklassen in Betracht, und dann können sich die verschiedenen Referenzklassen, die man bilden könnte, überschneiden. Man könnte z.B. für den Mord auf dem Kreuzfahrtschiff die Anfangswahrscheinlichkeit gesondert für männliche und weibliche Personen auf dem Schiff bestimmen, aber auch für Vorbestrafte und Nicht-Vorbestrafte, für die Reisegruppe des Mordopfers und diejenigen, die nicht zur Reisegruppe gehören usw. Jede dieser Eigenschaften ist für die Anfangswahrscheinlichkeit relevant, aber je nach Zuschnitt der Referenzklasse gelangt man dann zu unterschiedlichen Anfangswahrscheinlichkeiten (S. 354). Die Vorschläge in der induktiven Logik zur Lösung des Referenzklassenproblems reichen von der Wahl der engsten Klasse, für die verlässliche Statistiken vorliegen,²¹ bis zur Wahl der weitesten homogenen Klasse.²² Als Faustregel für die Praxis empfiehlt *Schweizer*, von der Referenzklasse auszugehen, die in den meisten Merkmalen mit dem relevanten Einzelfall übereinstimmt, bei gleich vielen übereinstimmenden Merkmalen diejenige, von der eine größere Homogenität zu erwarten ist (S. 366). Im Ergebnis bleibt es aber dabei, dass es keine logisch zwingende Wahl einer „richtigen“ Referenzklasse gibt. Allerdings ist die Frage für die praktische richterliche Tätigkeit in vielen Fällen weit weniger virulent, als man annehmen könnte: Denn wenn mehrere Referenzklassen in Betracht kommen, kann man eine wählen und dann das zusätzliche Merkmal bzw. die zusätzlichen Merkmale als belastende oder entlastende Indizien würdigen - die entsprechenden Wahrscheinlichkeiten müssen nötigenfalls plausibel geschätzt werden, aber das ist keine spezifische Problematik nur der Anfangswahrscheinlichkeit.²³

[19]

In der Praxis ist eine richterliche Überzeugungsbildung in den meisten Fällen nur mit Hilfe von Erfahrungssätzen möglich, die häufig keine zwingenden, sondern nur statistische Erfahrungssätze sind und oftmals auch nicht auf sauberen statistischen Daten beruhen (S. 370), sondern z.B. auf eigenen Erfahrungen des Gerichts, allgemeiner Lebenserfahrung oder mehr oder weniger belastbaren Alltagstheorien. Eine Entscheidungsfindung ohne Rückgriff auf Alltagstheorien sei, so *Schweizer*, nicht möglich, obwohl es keine sichere Abgrenzung zwischen brauchbaren und unbrauchbaren Alltagstheorien gebe: Die Umwelt sei zu komplex, um komplett ohne Vorurteile auszukommen (S. 383). Umso wichtiger sei es, auf Alltagswissen beruhendes Erfahrungswissen explizit zu machen, um eine kritische Untersuchung zu ermöglichen (S. 384). Gerade dann, wenn ein Erfahrungssatz ohne weitere Indizien den Schluss auf eine subjektive Wahrscheinlichkeit oberhalb der Beweismaßgrenze zulasse, sei eine Revisibilität aufgrund der damit verbundenen Gefahr für die Wahrheitsfindung sachgerecht (S. 406 f.)

- und überprüft wird (entgegen anderslautenden Äußerungen der Rechtsprechung²⁴) gerade der Beweiswert des Erfahrungssatzes (S. 409).

[20]

Im Hinblick auf den konkreten Inhalt von vielen Erfahrungssätzen weist *Schweizer* darauf hin, dass oftmals nicht hinreichend deutlich wird, ob ein Erfahrungssatz sich auf die A-posteriori-Wahrscheinlichkeit oder auf die Likelihood bezieht (S. 388): So könne man den Erfahrungssatz, dass bei einer Kollision mit einem überholenden Fahrzeug beim Rechtsabbiegen (gemeint ist wohl: beim Linksabbiegen) höchstwahrscheinlich der Schulterblick unterlassen wurde, einerseits dahingehend verstehen, dass es um die Wahrscheinlichkeit geht, dass der Schulterblick bei gegebener Kollision unterlassen wurde (das ist, was das Gericht im Regelfall interessieren wird), oder dahingehend, dass er sich auf die Wahrscheinlichkeit bezieht, mit der bei gegebener Unterlassung des Schulterblicks eine Kollision erfolgt - entweder also die Wahrscheinlichkeit des unterlassenen Schulterblicks im Fall einer Kollision, andererseits die Wahrscheinlichkeit der Kollision im Fall des unterlassenen Schulterblicks. Beide Wahrscheinlichkeiten darf man nicht verwechseln, das droht aber bei ungenauer Formulierung des Erfahrungssatzes.

- 105 -

Effer-Uhe, GVRZ 2025, 100-108

- 106 -

[21]

Die Lehre von den Erfahrungssätzen leidet zusätzlich darunter, dass oftmals nicht sauber zwischen der Wahrscheinlichkeit, dass ein Erfahrungssatz zutrifft (Unsicherheit über den Erfahrungssatz), und der Beweiskraft des Erfahrungssatzes, wenn er zutrifft (Unsicherheit des Erfahrungssatzes), unterschieden wird (S. 392). Beide Wahrscheinlichkeiten sind durch die Produktregel zu kombinieren (S. 393) - ein Erfahrungssatz, der mit einer Wahrscheinlichkeit von 90 % zutrifft und eine Wahrscheinlichkeit von 70 % für eine bestimmte Ursache-Wirkung-Kombination mit sich bringt, kann also nur noch einen 63 %igen Überzeugungsgrad von diesem Kausalverhältnis vermitteln ($0,9 \times 0,7 = 0,63$).

VII. Beweismaß und Fehlerkosten

[22]

Im fünften Teil schließlich geht es *Schweizer* um die Frage, wie das angemessene Beweismaß gestaltet sein sollte und ob dieses Beweismaß als Regelbeweismaß schon de lege lata angebracht ist. *Schweizer* vertritt mit einer zumindest im deutschen Recht extremen Mindermeinung das zivilprozessuale Regelbeweismaß der überwiegenden Wahrscheinlichkeit bzw. Überzeugtheit (also eine Überzeugtheit größer 50 %). Im Strafprozess geht er dagegen davon aus, dass ein erheblich höheres Beweismaß der vollen Überzeugung angemessen ist. Die Begründung für das Beweismaß der überwiegenden Wahrscheinlichkeit liegt für ihn in rechtsökonomischen bzw. entscheidungstheoretischen Überlegungen zu den Fehlerkosten: Je höher das Beweismaß ist, desto häufiger wird es zu falsch-negativen Entscheidungen kommen (also zur „falsch-negativen“ Verneinung der beweisbedürftigen Tatsachenbehauptung, obwohl die Behauptung tatsächlich wahr ist: Fehler erster Art oder Alpha-Fehler), je niedriger das Beweismaß ist, desto häufiger werden falsch-positive Entscheidungen sein („falsch-positive“ Bejahung der beweisbedürftigen Tatsachenbehauptung, obwohl die Behauptung unwahr ist:

Fehler zweiter Art oder Beta-Fehler). Die Reduzierung der Quote des einen Fehlers bringt also zwingend eine Erhöhung der Quote des anderen Fehlers mit sich - je höher die Sensitivität der Tatsachefeststellung, desto niedriger die Spezifität. Entscheidungstheoretisch sei es sinnvoll, das Beweismaß so anzusetzen, dass die erwarteten Fehlerkosten minimiert werden, und das sei bei einem Beweismaß der Fall, bei dem ab einer subjektiven Wahrscheinlichkeit der Tatbestandsverwirklichung größer 50 % der Beweis als geführt anzusehen ist. Das sei im Strafprozess anders, da die Verurteilung eines Unschuldigen hier schwerer wiege als der Freispruch eines Schuldigen (S. 443).

VIII. Fazit

[23]

Dass die Überzeugung für die Tatbestandsverwirklichung für *Schweizer* über 50 % liegen muss, führt allerdings nicht dazu, dass bei jedem einzelnen von mehreren Tatbestandsmerkmalen, deren Vorliegen bestritten wird, eine Wahrscheinlichkeit von 51 % ausreicht - es geht ja um die Wahrscheinlichkeit der Verwirklichung des ganzen Tatbestands, und bei mehreren Tatbestandsmerkmalen, deren Vorliegen unsicher ist, verringert sich die Gesamtwahrscheinlichkeit nach der Produktregel (S. 489 f.): Wenn zwei voneinander unabhängige Tatbestandsmerkmale A und B jeweils mit einer Wahrscheinlichkeit von 70 % vorliegen, beträgt die Wahrscheinlichkeit der Verwirklichung des gesamten Tatbestands schon nur noch 49 % ($0,7 \times 0,7 = 0,49$). *Schweizers* Überlegungen vermögen zwar rechnerisch zu überzeugen, sind aber gleichwohl kaum mit der auf einzelne Behauptungen und nicht auf gesamte Tatbestände bezogenen Formulierung von § 286 Abs. 1 ZPO („ob eine tatsächliche Behauptung für wahr oder für nicht wahr zu erachten sei“) zu vereinbaren,²⁵ wie auch *Schweizer* (S. 501) konzediert.

[24]

In der Literatur zum deutschen Recht wird eine Festlegung auf einen bestimmten Überzeugungsgrad als Regelbeweismaß meistens vermieden.²⁶ Zwischen den - seltenen - Bezifferungen gibt es zudem erhebliche Unterschiede (z.B. 99,8 %²⁷ gegenüber 75 %²⁸, letzteres allerdings in der Besprechung eines skandinavischen Werks ohne explizite Klarstellung, ob diese Werte auch auf Deutschland übertragen werden können). Für die Schweiz, in der häufiger Prozentwerte angegeben werden, geht *Schweizer* von einem verbreiteten Überzeugungsgrad von 90 % als Regelbeweismaß aus (S. 472). Tatsächlich werden in Deutschland weit auseinanderliegende Überzeugungsgrade unter den Unschärfen der Anastasia-Formel verborgen, was die Diskussion über das Regelbeweismaß erschwert. Ein Überzeugungsgrad von „nur“ 90 % wie in der Schweiz wäre allerdings wohl mit der Anastasia-Formel des BGH schwerlich vereinbar - wer nur mit einer Wahrscheinlichkeit von 90 % annimmt, dass der Vortrag des Beweislastträgers zutrifft, hat sehr wohl noch konkrete, nicht nur theoretische Zweifel.²⁹ Dieses in der praktischen Autorität des BGH wurzelnde Argument spricht natürlich wissenschaftlich nicht zwingend gegen die Annahme eines niedrigeren Regelbeweismaßes - man muss sich aber darüber im Klaren sein, dass dieses Beweismaß dann mit der höchstrichterlichen Rechtsprechung nicht in Einklang zu bringen ist.

[25]

Die rechtsökonomische Argumentation von *Schweizer* ist im Ausgangspunkt bestechend. Er selbst weist allerdings darauf hin, dass es sich nicht um eine rechtsökonomische Argumentation im engeren Sinne handelt, weil es nicht um die Erreichung von Effizienz- oder Wohlfahrtszielen gehe, sondern um die Umsetzung des materiellen Rechts, also um gesetzgeberische Wertentscheidungen, die auch im Prozess zu respektieren seien (S. 488). Hier muss er sich fragen lassen, ob er nicht gleichzeitig selbst Wertentscheidungen des ZPO-Gesetzgebers nicht hinreichend beachtet. Damit setzt sich auch *Schweizer* auseinander. Sein Ausgangspunkt ist die zutreffende Beobachtung, dass der Geset-

zeswortlaut selbst über den Grad der notwendigen Überzeugung nichts aussagt (S. 541). Die Annahme *Schweizers*, dass die Gesetzgebungsmaterialien zur CPO zum notwendigen Beweismaß schweigen,³⁰ ist allerdings bei genauerer Betrachtung unzutreffend. Tatsächlich gibt es sehr deutliche Hinweise, dass der Gesetzgeber der CPO 1877 das Beweismaß der vollen Überzeugung in § 286 hat verankern wollen, wenn auch nicht unmittelbar in den Materialien zur freien Beweiswürdigung.³¹

- 106 -

Effer-Uhe, GVRZ 2025, 100-108

- 107 -

Das ergibt sich daraus, dass er das Institut des Parteieids, der erst 1933 durch die Parteivernehmung ersetzt wurde, in die CPO aufgenommen hat: Derjenige, zu dessen Gunsten schon mehr sprach, konnte danach zum richterlichen Eid zugelassen werden, um die volle Überzeugung zu erreichen, der Eid begründete gerade zwingend (wenn auch nur durch Fiktion) die volle Überzeugung und war nicht mehr der freien Würdigung zugänglich.³² Das wäre aber von vornherein unnötig gewesen, wenn die bloße überwiegende Überzeugung ausreicht. Der Gesetzgeber der CPO 1877 ist also sehr wohl von einem Beweismaß der vollen Überzeugung ausgegangen.

[26]

Die spannende Frage ist, ob diese gesetzgeberische Entscheidung durch die Abschaffung des Parteieids und seine Ersetzung durch die frei zu würdigende Parteivernehmung 1933 ihre Grundlage verloren hat. Die Reform hat die Grundlagen der Überzeugungsbildung massiv verändert: Während bis 1933 die - wenn auch vom Gesetz fingierte - volle Überzeugung durch die Eidesleistung erreicht werden konnte, wenn zuvor eine tatsächliche überwiegende Überzeugung des Gerichts die Eidesleistung ermöglicht hatte, so musste nach 1933 der Beweislastträger das Gericht positiv im Rahmen der stattdessen (wenn auch nach h.M. nur unter einschränkenden Voraussetzungen) möglichen Parteivernehmung überzeugen.³³ Eine Äußerung des Gesetzgebers dazu, ob es gleichwohl beim Beweismaß der vollen Überzeugung bleiben soll, sucht man allerdings in den Gesetzgebungsmaterialien³⁴ vergebens.³⁵ Das ursprüngliche Ziel der Normierung des Beweismaßes der vollen Überzeugung in Verbindung mit der Normierung des Parteieids, dem Richter im Ergebnis die Entscheidung zugunsten der Partei, für die mehr spricht, unabhängig von der Beweislastverteilung zumindest nach Eidesleistung zu ermöglichen, war mit der Reform von 1933 nicht mehr erreichbar.³⁶ Anstelle der nach Eidesleistung fingierten vollen Überzeugung nunmehr eine tatsächliche volle Überzeugung zu fordern, kann ebenfalls nicht überzeugen, weil das praktisch die Anforderungen an den Beweis deutlich verschärft hätte, und für eine derartige Verschärfung wären dann doch Anhaltspunkte im Gesetzeswortlaut oder der Gesetzesbegründung zu erwarten.³⁷

[27]

Das gilt aber ebenso für einen Übergang zu einem Beweismaß der bloß überwiegenden Überzeugung, wie *Schweizer* es propagiert, denn bis 1933 war die Entscheidung zugunsten desjenigen, für dessen Behauptung mehr sprach, nur unter der Voraussetzung der Eidesleistung - und damit der Übernahme eines Strafbarkeitsrisikos, das gleichzeitig über das Restitutionsverfahren Raum für eine Verfahrenswiederaufnahme im Fall des Falscheids ließ - möglich.³⁸ Ein solches Strafbarkeitsrisiko mit Wiederaufnahmemöglichkeit besteht dagegen nicht automatisch, wenn man die bloße überwiegende Wahrscheinlichkeit als Regelbeweismaß genügen lässt.³⁹ So überzeugend die entscheidungstheoretische

Begründung *Schweizers* für das Beweismaß der überwiegenden Überzeugung auch zumindest im Ausgangspunkt ist, ist ihr de lege lata nicht zu folgen, schon weil der ZPO-Gesetzgeber mit der Reform 1933 jedenfalls nicht davon abgerückt ist, dass die freie Beweiswürdigung auf die einzelne Behauptung und nicht auf die Verwirklichung des ganzen Tatbestands gerichtet ist. Damit ist eine Gesamtwahrscheinlichkeit für die Verwirklichung des gesamten Tatbestands von 50 % plus X nicht zu vereinbaren. Eine bloß überwiegende Wahrscheinlichkeit bezüglich jedes einzelnen bestrittenen Tatbestandsmerkmals (die auch *Schweizer* bei mehreren streitigen Tatbestandsvoraussetzungen nicht fordert) könnte aber zu einem Überzeugungsgrad bezüglich der Verwirklichung des gesamten Tatbestands führen, der weit, unter Umständen auch sehr weit, unter 50 % liegt, was nicht zu überzeugen vermag.

[28]

Wenn man mit *Schweizer* von den Fehlerkosten ausgeht, kann man auch im Zivilprozess mit Fug und Recht darüber streiten, ob die Kosten eines Fehlers erster Art und diejenigen eines Fehlers zweiter Art tatsächlich identisch sind. Aufgrund eines Status-quo-Bias und der aus ihm folgenden Verlustaversion (der Beklagte nimmt den Verlust bei einer Klagestattgabe als relevanter wahr als der Kläger den ausgebliebenen Gewinn bei einer Klageabweisung) könnte man zu dem Ergebnis kommen, dass die unberechtigte Klagestattgabe subjektiv für die unterlegene Partei schwerer wiegt als die unberechtigte Klageabweisung. Diese Wahrnehmung ist zwar irrational, aber dadurch als Wahrnehmung nicht weniger real. *Schweizer* setzt sich mit der Verlustaversion zwar auseinander, ist aber der Auffassung, dass sie nur die Intuition erklärt, eine Entscheidungsgrenze deutlich über 50 % Überzeugtheit anzusetzen, aber keine normative Grundlage dafür bilden kann (S. 513, 538). Dasselbe Argument lässt sich jedoch auch gegen *Schweizers* Lösung wenden - seine normative Grundlage ist zwar der vom Gesetzgeber im materiellen Recht verankerte Gerechtigkeitsgehalt, den seine Lösung auf entscheidungstheoretischer Basis umsetzen soll. Warum dieser Gerechtigkeitsgehalt aber zwingend an die materiellen Kosten und nicht an die Kostenwahrnehmung durch die Parteien anknüpfen muss, bleibt offen, vor allem angesichts dessen, dass der Gesetzgeber mit § 286 Abs. 1 ZPO gerade keine vollständige Umsetzung der besten materiellen Lösung im Prozess vorgesehen hat. Zutreffend ist allerdings der Hinweis *Schweizers*, dass die Frage, welcher Status als Status quo wahrgenommen wird, gewissen Manipulationen zugänglich ist (S. 513 f.). Geht man von dem empirisch festgestellten Verhältnis zwischen Gewinnchance und Verlustrisiko aus, ab dem ein Spiel mit gleich hohen Chancen für die meisten Menschen akzeptabel wird,⁴⁰ und versucht, das in eine entsprechenden Beweismaßhürde für den Kläger umzusetzen, entspricht das einem Überzeugtheitsgrad von etwa 70 % (S. 513).

[29]

Hinterfragen kann man aber auch den entscheidungstheoretischen Ausgangspunkt *Schweizers*. Er geht davon aus, dass die Fehlerkosten am geringsten sind, wenn das Gericht ab einem

- 107 -

Effer-Uhe, GVRZ 2025, 100-108

- 108 -

Überzeugtheitsgrad von 50 % plus X für die Verwirklichung des gesamten Tatbestands zugunsten des Klägers entscheidet. Das wäre, auch wenn man - anders als hier vertreten - dem Ausgangspunkt der objektiven Berechnung der Fehlerkosten zustimmt, nur dann zutreffend, wenn man davon ausgeht, dass die Gerichte ebenso häufig die Wahrheit einer Behauptung über- wie unterschätzen. Auch das kann man bezweifeln: So haben zwar tendenziell Gerichte Schwierigkeiten damit, intuitiv die Wahr-

scheinlichkeiten mehrerer Indizien in einem Beweising zu kombinieren und gelangen dann eher zu einem zu niedrigen Überzeugungsgrad.⁴¹ Andererseits besteht eine große Geneigtheit der Zivilgerichte dahin, einem (einzigem) Zeugen als Beweismittel grundsätzlich zu glauben,⁴² was tendenziell zu einer „zu hohen“ Überzeugung beim Personalbeweis führen dürfte. Im Sinne einer Fehlerkostenminimierung müsste man also eigentlich bei der Kombination mehrerer Beweismittel in einem Beweising die Grenze des Regelbeweismaßes tendenziell etwas niedriger ansetzen, bei Zeugen, denen weder Sachbeweise noch andere Zeugen zur selben Behauptung entgegenstehen, etwas höher. Darüber, ob das praktikabel wäre, lässt sich streiten.

[30]

Geht man von einem bestimmten erforderlichen Überzeugungsgrad für die einzelne Tatsachenbehauptung aus, der deutlich niedriger als 90 % liegt, kann das dazu führen, dass bei einer ausreichenden Anzahl streitiger Tatsachenbehauptungen eine rechnerische Überzeugtheit unter 50 % von der kumulativen Verwirklichung aller Tatbestandsvoraussetzungen ausreichen kann, um zu einer Klagestattgabe zu gelangen. Diese Folge lässt sich angesichts des Bezugs der Überzeugtheit auf die einzelne Tatsachenbehauptung in § 286 Abs. 1 ZPO de lege lata nicht vermeiden. Immerhin stellt eine solche Vielzahl von streitigen Voraussetzungen in der (zivilrechtlichen) Praxis eher die Ausnahme dar, und selbst bei einer Vielzahl an streitigen Behauptungen wird normalerweise nicht der Überzeugungsgrad hinsichtlich jeder einzelnen davon nur knapp über dem Regelbeweismaß liegen, so dass ein Absinken des Überzeugungsgrads vom kumulativen Vorliegen aller Tatbestandsmerkmale unter 50 % in der Praxis auch bei einer behutsamen Absenkung des Regelbeweismaßes nur in extremen Ausnahmefällen zu erwarten ist.

[31]

Ein weiterer Nachteil an dem von *Schweizer* geforderten Überzeugungsgrad von „nur“ 50 % plus X für die Verwirklichung des Gesamttatbestandes liegt darin, dass damit aufgrund einer verbreiteten Überschätzung der eigenen Erfolgsaussichten⁴³ Klagen provoziert werden könnten, deren Erfolgswahrscheinlichkeit tatsächlich unter 50 % liegt.

[32]

Mit Hilfe einer überzeugend konzipierten empirischen Untersuchung stellt *Schweizer* zuletzt fest, bei welchem Überzeugungsgrad das Regelbeweismaß denn unter Richtern tatsächlich liegt (S. 576). Hier gelangt er zu dem überraschenden Ergebnis, dass bereits ab einem rechnerischen Überzeugungsgrad vom Vorliegen des Gesamttatbestands von nur 63 % etwa 50 % der Probanden einer Leistungsklage stattgeben würden (S. 594).

[33]

Insgesamt handelt es sich um eine lesenswerte Arbeit, die viele Denkanstöße gibt und in vielen Punkten überzeugt. Auch wenn der Rezensent nicht in jedem Punkt der Lösung der Arbeit zu folgen vermag, stellt die Arbeit einen großen Gewinn für die Rechtswissenschaft dar - und nicht zuletzt auch ein leider seltenes Beispiel für eine gelungene interdisziplinäre Herangehensweise. Dem Appell *Schweizers* für eine „interdisziplinäre Wende“ des Beweisrechts (S. 2) ist beizupflichten - nach Auffassung des Rezensenten (der wohl auch *Schweizer* zustimmen würde) täte vor allem eine probabilistische Wende not.

Fußnoten

*)

Der Autor ist Professor für Bürgerliches Recht, Rechtsgeschichte, Rechtstheorie und Rechtspsychologie an der BSP Business and Law School in Berlin.

- 1) Im Volltext zu finden unter https://www.marschweizer.ch/wp-content/uploads/2017/09/schweizer_kognitive_t%C3%A4uschungen_vor_gericht.pdf (Datum des letzten Abrufs: 13.5.2025).
- 2) So für den Strafprozess zutreffend *Wohlens* GA 2016, 725.
- 3) BGH, NJW 1970, 946, 948.
- 4) Zustimmend *Rüssmann* in Adolphsen/Goebel/Haas/Hess/Kolmann/Würdinger, FS für Peter Gottwald zum 70. Geburtstag, 2014, S. 539 (541).
- 5) A.A. z.B. *Brinkmann*, Das Beweismaß im Zivilprozess aus rechtsvergleichender Sicht, 2005, S. 63.
- 6) Einführend dazu *Effer-Uhe/Mohnert*, Psychologie für Juristen, 2019, Rz. 84 ff.
- 7) *Hansjakob/Gundlach/Straub*, Kriminalistisches Denken, 12. Aufl. 2024, S. 405 ff.
- 8) Dazu in Kürze ausführlich *Effer-Uhe*, Würdigung einer Vielzahl von Indizien (erscheint demnächst in NJW 2025).
- 9) Nachweise bei *Greger*, VersR 1980, 1091, 1099.
- 10) *Foerste* in Musielak/Voit, 22. Aufl. 2025, § 286 ZPO Rz. 23.
- 11) Vgl. BGH v. 4.5.1988 - IVa ZR 278/86, MDR 1988, 846 = NJW 1988, 2040, 2041; BGH v. 4.6.2007 - II ZR 147/05, MDR 2007, 1204 = NJW 2008, 76, 77.
- 12) Lesenswert dazu <https://www.newyorker.com/magazine/2024/05/20/lucy-letby-was-found-guilty-of-killing-seven-babies-did-she-do-it> (Datum des letzten Abrufs: 13.5.2025).
- 13)

- Thompson, Law, Probability and Risk* (2009) 8, S. 257 f.
- 14) Vgl. dazu *Schweizer, Beweiswürdigung und Beweismaß*, 2015, S. 192 Fn. 617.
- 15) Vgl. *Effer-Uhe*, Würdigung einer Vielzahl von Indizien (erscheint demnächst in NJW 2025).
- 16) Vgl. näher *Effer-Uhe/Mohnert, Psychologie für Juristen*, 2019, Rz. 101 ff.
- 17) Vgl. *Effer-Uhe/Mohnert, Psychologie für Juristen*, 2019, Rz. 25 ff., 57.
- 18) *Bennett/Feldman, Reconstructing Reality in the Courtroom*, 1981, S. 83 ff.
- 19) *Simon/Snow/Read, Journal of Personality and Social Psychology* 2004, 814 ff.
- 20) *Glöckner/Engel, Preprints of the MPI for Research on Collective Goods* 2008, S. 5 f.
- 21) *Reichenbach, The Theory of Probability*, 1949, S. 374; ähnlich *Koch/Rüßmann, Juristische Begründungslehre*, 1982, S. 305 f.
- 22) *Salmon, The Foundations of Scientific Inference*, 1967, S. 91.
- 23) Dazu *Effer-Uhe*, Würdigung einer Vielzahl von Indizien (erscheint demnächst in NJW 2025).
- 24) BGH, NJW 1973, 1411, 1412; zustimmend *Prütting* in *MünchKomm/ZPO*, 7. Aufl. 2025, § 284 Rz. 47.
- 25) Vgl. zur Vereinbarkeit von *Schweizers* Konzeption mit § 286 ZPO auch *Gottwald* in *Fankhauser/Lüchinger/Klingler/Seiler, Das Zivilrecht und seine Durchsetzung - Festschrift für Professor Thomas Sutter-Somm*, 2016, S. 125 (133 f.).
- 26) Kritisch dazu *Effer-Uhe*, ZZP 133 (2020), 339 (342 ff.).
- 27)

- Bender* in Grunsky u.a., FS für Fritz Baur, 1981, S. 247 (258).
- 28) *Bruns*, ZZP 91 (1978), 64 (69 ff.).
- 29) *Effer-Uhe*, ZZP 133 (2020), 339 (344 f.).
- 30) *Schweizer* in *Effer-Uhe/Hoven/Kempny/Rösinger*, Einheit der Prozessrechtswissenschaft?, 2016, S. 341 (355).
- 31) Ausführlich zum Folgenden *Effer-Uhe*, ZZP 133 (2020), 339 (348 ff.).
- 32) Näher *Effer-Uhe*, Die Parteivernehmung - Überlegungen zu einer verstärkten Nutzbarmachung von § 448 ZPO, 2015, S. 142 ff.
- 33) *Effer-Uhe*, ZZP 133 (2020), 339 (351).
- 34) Reichsjustizministerium, Begründung zum Entwurf einer Zivilprozessordnung, 1932, S. 337.
- 35) *Effer-Uhe*, ZZP 133 (2020), 339 (351 f.).
- 36) *Effer-Uhe*, ZZP 133 (2020), 339 (354).
- 37) *Effer-Uhe*, ZZP 133 (2020), 339 (354).
- 38) *Effer-Uhe*, ZZP 133 (2020), 339 (354).
- 39) Im Bereich des Prozessbetrugs gibt es Strafbarkeitslücken, die daraus resultieren, dass für den Parteivortrag abgesenkte Wahrhaftigkeitsanforderungen bestehen: Behaupten darf die Partei schon das, was sie nur für möglich hält (ausführlich *Effer-Uhe*, RW 2015, S. 380 ff.).
- 40)

Vgl. dazu *Tversky/Kahneman*, Quarterly Journal of Economics 1991, 1039 (1053); *Tversky/Kahneman*, Journal of Risk and Uncertainty 1992, 297 (1053), die ein Verhältnis zwischen 2 und 2,5 mit einem Median bei 2,25 angeben.

- 41) *Effer-Uhe*, Würdigung einer Vielzahl von Indizien (erscheint demnächst in NJW 2025).
- 42) *Bender/Schumacher*, Erfolgsbarrieren vor Gericht, 1980, S. 137 f.
- 43) Vgl. dazu *Goodman-Delahunty/Granhag/Hartwig/Loftus*, Psychology, Public Policy, and Law 16 (2010), S. 133, 135 ff.

© Verlag Dr. Otto Schmidt, Köln