

§ 12 Bestätigungsfehler (*confirmation bias*)

- 547 Unter dem Oberbegriff Bestätigungsfehler (*confirmation bias*) wird eine Vielzahl verschiedener Phänomene zusammengefasst. Dementsprechend unterschiedlich sind die Definitionen des Begriffs in der Literatur.<sup>855</sup> Hier soll unter Bestätigungstendenz das Phänomen verstanden werden, dass Informationen, die eine von der Versuchsperson zumindest in Betracht gezogene Hypothese bestätigen, von ihr eher gesucht, wahrgenommen, stärker gewichtet oder besser in Erinnerung behalten werden als Informationen, die gegen die Hypothese sprechen.<sup>856</sup> Noch allgemeiner formuliert liegt eine Bestätigungstendenz vor, wenn die Neigung besteht, eine zumindest in Betracht gezogene Hypothese beizubehalten und eine korrespondierende Abneigung, sie zugunsten einer neuen Hypothese aufzugeben.<sup>857</sup>
- 548 Während man sich weitgehend einig ist, dass Bestätigungstendenzen real sind,<sup>858</sup> besteht über die Ursachen der beobachteten Phänomene bei weitem keine Einigkeit. Können Menschen Hypothesen nicht neutral und unvoreingenommen testen oder wollen sie nicht? Wenn kognitive Faktoren zumindest mitverantwortlich sind für die beobachteten Tendenzen: Sind Bestätigungstendenzen darauf zurückzuführen, dass Menschen nicht in der Lage sind, mehr als eine Alternative gleichzeitig zu beachten?<sup>859</sup> Oder darauf, dass Menschen das Vorhandensein einer Eigenschaft leichter erkennen und erinnern können als ihr Fehlen?<sup>860</sup> Oder kann erst eine Interaktion verschiedener Phänomene, wie einer positiven Teststrategie und der leichteren Verarbeitung positiver Informationen, Bestätigungsfehler erklären?<sup>861</sup>
- 549 Nicht nur keine Einigkeit, sondern Streit herrscht bezüglich der Frage, ob Bestätigungstendenzen als „Fehler“ bezeichnet werden dürfen; d.h. ob sie „irrational“ sind und zu Resultaten führen, die suboptimal sind. Während bis Mitte der 80-er Jahre des letzten Jahrhunderts die Ansicht vorherrschte, dass Bestätigungstendenzen ein schweres Defizit menschlichen Denkens mit negativen Konsequenzen sind, hat sich seither das Bild gewandelt. Zahlreiche Autoren weisen darauf hin, dass das, was vorschnell als Bestätigungsfehler bezeichnet wurde, ausserhalb des Labors oft eine der Umwelt angepasste Strategie ist, die

---

<sup>855</sup> JOSHUA KLAYMAN, Varieties of Confirmation Bias, in: JEROME BUSEMEYER/REID HASTIE/DOUGLAS L. MEDIN (Hrsg.), Decision Making From a Cognitive Perspective, New York 1995, 385-418, 385: “There are almost as many operational definitions of *confirmation bias* as there are studies”; RAYMOND S. NICKERSON, Confirmation Bias: A Ubiquitous Phenomenon in Many Guises, Review of General Psychology 1998, 175-220, 220.

<sup>856</sup> Nach KLAUS FIEDLER, Beruhen Bestätigungsfehler nur auf einem Bestätigungsfehler?, Psychologische Beiträge 1983, 280-286, 282.

<sup>857</sup> KLAYMAN, FN 855, 386.

<sup>858</sup> KLAYMAN, FN 855, 385; NICKERSON, FN 855, 177; FIEDLER, FN 856, 282.

<sup>859</sup> CLIFFORD R. MYNATT/MICHAEL E. DOHERTY/WILLIAM DRAGAN, Information Relevance, Working Memory, and the Consideration of Alternatives, The Quarterly Journal of Experimental Psychology 1993, 759-788.

<sup>860</sup> JONATHAN ST. B. T. EVANS, Bias in Human Reasoning: Causes and Consequences, Hove 1989.

<sup>861</sup> KLAYMAN, FN 855, 399.

mit minimalem Aufwand wenn nicht zu optimalen, so doch zu vernünftigen Resultaten gelangt.<sup>862</sup>

- 550 Eine Bestätigungstendenz kann erstens daraus resultieren, dass man von vorneherein nur solche Informationsquellen auswählt, von denen bestätigende Informationen zu erwarten sind. Zweitens kann eine Bestätigung zu Unrecht erfolgen, wenn ambivalente Informationen, die sowohl für wie gegen die Hypothese sprechen können, konsequent als bestätigend codiert werden. Drittens kann eine unangebrachte Bestätigung erfolgen, wenn man zwar sowohl bestätigende wie nichtbestätigende Informationen registriert, letztere aber nicht (oder nur ungenügend) zur Beurteilung der Hypothese heranzieht.<sup>863</sup> Nach dieser Systematik lassen sich Bestätigungstendenzen bei der Informationssuche und bei der Informationsinterpretation unterscheiden. Sie dient als Gerüst der folgenden Übersicht, die sich massgeblich auf die Übersichtsartikel von NICKERSON und KLAYMAN stützt.<sup>864</sup> In Anbetracht der beinahe unüberschaubaren Literatur zum Hypothesentesten muss diese Übersicht notwendigerweise verkürzt und unvollständig ausfallen. Nach der Darstellung der empirischen Belege für Bestätigungstendenzen werden ihre Folgen beleuchtet, ehe die Resultate der eigenen Studie rapportiert und diskutiert werden; schliesslich wird dargestellt, was gegen den Einfluss des *confirmation bias* unternommen werden kann.

## I. Informationssuche

### A. Pseudodiagnostizität

- 551 Der Begriff „Pseudodiagnostizität“ wurde von DOHERTY et al. geprägt.<sup>865</sup> Wie vorne, S. 133 ff., dargelegt wurde, beurteilt sich die abstrakte Beweiskraft eines Indizes I danach, wie viel häufiger (seltener) es bei Zutreffen der Hypothese H vorliegt als bei Nichtzutreffen von H (abstrakte Beweiskraft =  $P(I | H)/P(I | \neg H)$  = Likelihood-Quotient). Die meisten Menschen scheinen dieses Konzept wie bereits erwähnt nicht intuitiv zu verstehen. In einer typischen Studie zur Pseudodiagnostizität werden die Versuchspersonen gebeten, anzugeben, welcher Kategorie eine Instanz wahrscheinlich angehört; beispielsweise von welchem Fundort ein Gefäss stammt.<sup>866</sup> Ein Gefäss aus glattem Ton mit gebogenen Henkeln kann beispielsweise entweder von der Muschel-Insel oder von der Korallen-Insel kommen. Welche Informationen werden benötigt, um zu entscheiden, ob das Gefäss eher von der der Muschel- oder von der Korallen-Insel stammt? Nachdem die Versuchspersonen sich aufgrund der bekannten Merkmale eine erste Meinung zur Herkunft des Gefässes gebildet hatten, konnten sie weitere Informationen über die Verteilung der Eigenschaften

---

<sup>862</sup> VOLKER GADENNE, Der Bestätigungsfehler und die Rationalität kognitiver Prozesse, Psychologische Beiträge 1982, 11-25; JOSHUA KLAYMAN/YOUNG-WON HA, Confirmation, Disconfirmation, and Information in Hypothesis Testing, Psychological Review 1987, 211-228; JAMES FRIEDRICH, Primary Error Detection and Minimization (PEDMIN) Strategies in Social Cognition: A Reinterpretation of Confirmation Bias Phenomena, Psychological Review 1993, 298-319.

<sup>863</sup> GADENNE, FN 862, 13.

<sup>864</sup> NICKERSON, FN 855; KLAYMAN, FN 855.

<sup>865</sup> MICHAEL E. DOHERTY/CLIFFORD R. MYNATT/RYAN D. TWENEY/MICHAEL D. SCHIAVO, Pseudodiagnosticity, Acta Psychologica 1979, 111-121.

<sup>866</sup> DOHERTY/MYNATT/TWENEY/SCHIAVO, FN 866, 114 f.

## Besonderer Teil

der Gefässe beider Inseln erwerben. Die meisten Versuchspersonen, die sich aufgrund des ersten Eindrucks für die Korallen-Insel entschieden hatten, fragten nur oder vorwiegend nach Informationen zu den Gefässen von der Korallen-Insel.<sup>867</sup> Entscheidend ist aber immer das Verhältnis der Merkmale; d.h. die Information, dass 80 % der Gefässe von der Korallen-Insel gebogene Henkel haben, ist nur diagnostisch, wenn ich weiss, dass nicht auch 80 % der Gefässe von der Muschel-Insel gebogene Henkel haben. Wenn ich nicht danach frage, werde ich es nie erfahren. Wie vorne erwähnt, scheinen Menschen zu meinen, die Kenntnis der Typizität oder Repräsentativität einer Eigenschaft für die Kategorie genüge, um auf die Zugehörigkeit der Instanz zur Kategorie zu schliessen.<sup>868</sup> DOHERTY et al. haben die Tendenz ihrer Versuchspersonen, sich mehr Informationen zu der Insel zu beschaffen, die sie ursprünglich favorisierten, *confirmation bias* genannt.<sup>869</sup> Diese Bezeichnung ist nicht unumstritten.<sup>870</sup> Richtigerweise spricht man wohl besser einfach von einer ineffizienten Suchstrategie, die weder bestätigende noch widerlegende Informationen liefern kann.

- 552 Der Anteil der nachgefragten pseudodiagnostischen Informationen sinkt, je mehr die Versuchspersonen Zweifel an der ursprünglichen Hypothese haben.<sup>871</sup> MYNATT et al. erklären dies damit, dass die Versuchspersonen eher bereit sind, eine alternative Hypothese in Betracht zu ziehen, wenn die ursprünglich Hypothese von den vorhandenen Daten nur schwach gestützt wird. Pseudodiagnostizität resultiert ihrer Meinung nach daraus, dass Menschen nicht in der Lage sind, mehr als eine Hypothese gleichzeitig zu verfolgen. Verschiedene Hypothesen werden seriell, nacheinander geprüft. Erst der Wechsel von der einen Hypothese zur anderen bewirkt, dass Informationen zu den Eigenschaften der Instanzen der anderen Kategorie abgefragt werden.<sup>872</sup>
- 553 FEENEY und EVANS weisen darauf hin, dass „pseudodiagnostische“ Information wertvoll sein kann, wenn das Indiz, das die ursprüngliche Hypothese stützt, sehr häufig ist.<sup>873</sup> Angenommen, ich weiss nicht mehr, ob meine Schwester ein Auto der Marke X oder Y hat. Ich weiss aber, dass das Auto ein Radio hat und eine Höchstgeschwindigkeit von 280 km/h erreicht. 80 % der Autos der Marke X haben ein Radio. Streng logisch müsste ich mich danach erkundigen, wie viele Autos der Marke Y ebenfalls ein Radio haben. Da aber die meisten Autos heute ein Radio haben, wird die Antwort auf diese Frage nicht viel weiterhelfen. Die Frage danach, wie viele der Autos der Marke X eine Höchstgeschwindigkeit von mehr 280 km/h erreichen, ist zwar streng genommen pseudodiagnostisch; wenn ich aber erfahre, dass sehr viele Autos der Marke X so schnell sind, potentiell wertvoller als

---

<sup>867</sup> DOHERTY/MYNATT/TWENEY/SCHIAVO, FN 866, 117.

<sup>868</sup> KLAYMAN, FN 855, 397.

<sup>869</sup> DOHERTY/MYNATT/TWENEY/SCHIAVO, FN 866, 116.

<sup>870</sup> DAWES, FN 652, 12088.

<sup>871</sup> MYNATT/DOHERTY/DRAGAN, FN 859, 774; JONATHAN ST.B.T. EVANS/SIMON VENN/AIDAN FEENEY, *Implicit and Explicit Processes in a Hypothesis Testing Task*, *British Journal of Psychology* 2002, 31-46.

<sup>872</sup> MYNATT/DOHERTY/DRAGAN, FN 859, 774 f.

<sup>873</sup> AIDAN FEENEY/JONATHAN ST.B.T. EVANS/SIMON VENN, *A Rarity Heuristic for Hypothesis Testing*, in: LILA R. GLEITMAN/ARAVID K. JOSHI (Hrsg.), *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*, Mahwah 2000, 119-124; ob eine Eigenschaft selten ist, können die Versuchspersonen auch aufgrund ihres Hintergrundwissens ahnen, AIDAN FEENEY/JONATHAN ST.B.T. EVANS/JOHN CLIBBENS, *Background Beliefs and Evidence Interpretation*, *Thinking and Reasoning* 2000, 97-124.

die Häufigkeit von Autoradios bei Y.<sup>874</sup> FEENEY et al. können nachweisen, dass Menschen bei der Suche nach Informationen berücksichtigen, wie wertvoll die gesuchte Information potenziell sein kann. Ist das bekannte Indiz selten, werden weniger häufig pseudodiagnostische Informationen gesucht als wenn das bekannte Indiz häufig ist; wie erwähnt ist die Suche nach pseudodiagnostischer Information im zweiten Fall nicht von vorneherein irrational.

## B. Bevorzugte Suche nach bestätigenden Informationen

- 554 Eines der klassischen Experimente, das zur Begründung der These, dass Menschen nicht nach falsifizierenden Informationen suchen, herangezogen wird, stammt von PETER WASON.<sup>875</sup> Den Versuchspersonen wird ein Zahlentripel 2, 4, 6 vorgelegt. Sie erhalten die Aufgabe, die Regel herauszufinden, nach der die Folge gebildet wurde. Hierzu sollen sie weitere Zahlentripel bilden, die ihrer Meinung nach der Regel entsprechen. Nach jedem Versuch erhalten sie vom Versuchsleiter eine Rückmeldung darüber, ob ihre produzierte Folge zu der wirklichen Regel passt oder nicht. Die Versuchspersonen dürfen ihre produzierten Folgen, ihre Hypothesen und die Antworten des Versuchsleiters schriftlich festhalten. Den Versuchspersonen wird weiter erklärt, dass sie die gesuchte Regel erst dann als Ergebnis bekannt geben sollen, wenn sie aufgrund ihrer Versuche ziemlich sicher seien, sie gefunden zu haben. Die zu findende Regel lautet, dass jede folgende Zahl grösser als die vorangehende Zahl ist.
- 555 72 % der Versuchspersonen fanden die Lösung, aber nur 21 %, ohne mindestens eine falsche Hypothese zu nennen. Die meisten Versuchspersonen bilden eine bestimmte Hypothese (z.B. „Folgende Zahl = Vorhergehende Zahl + 2“) und produzieren dann ausschliesslich positive Instanzen für diese Hypothese (d.h. Zahlentripel, die der Hypothese entsprechen, wie 10, 12, 14).<sup>876</sup> Kaum eine Versuchsperson formuliert negative Instanzen für ihre Hypothese. WASON und JOHNSON-LAIRD ziehen daraus den Schluss, dass sich die Versuchspersonen nicht gemäss der Methode des kritischen Rationalismus (POPPER) verhalten, weil sie nur versuchen, ihre Hypothese zu verifizieren.<sup>877</sup>
- 556 Das besprochene Experiment ist aber nicht geeignet, dies zu beweisen. Wie GADENNE zu Recht anführt, ist die Strategie, positive Instanzen einer Theorie zu testen, nicht notwendigerweise eine Bestätigungstendenz. Die Versuchspersonen wissen nicht, ob sie eine positive oder negative Rückmeldung durch den Versuchsleiter erhalten werden. Jede positive Instanz einer Hypothese ist geeignet, die Hypothese zu falsifizieren.<sup>878</sup> Dass die Versuchspersonen vorwiegend bestätigendes Feedback erhalten haben, liegt daran, dass die von den meisten formulierte Theorie „Folgende Zahl = Vorhergehende Zahl + 2“ enger ist als die gesuchte Theorie „ansteigende Zahlen“. Die Vorgabe des Zahlentripels 2, 4, 6 hat

---

<sup>874</sup> FEENEY/EVANS/VENN, FN 873, 121.

<sup>875</sup> PETER C. WASON, On the Failure to Eliminate Hypothesis in a Conceptual Task, Quarterly Journal of Experimental Psychology 1960, 129-140.

<sup>876</sup> WASON, FN 875, 135.

<sup>877</sup> PETER C. WASON/PHILIP N. JOHNSON-LAIRD, Psychology of Reasoning, London 1972, 210.

<sup>878</sup> GADENNE, FN 862, 18 f.

## Besonderer Teil

sicher dazu beigetragen, dass die Versuchspersonen eine enge Hypothese entwickelten. Die Wahrscheinlichkeit, dass ein nach der Regel „Folgende Zahl = Vorhergehende Zahl + 2“ gebildetes Zahlentripel aus der Menge der Zahlen zwischen 0 und 1'000 gezogen wird, wenn die Regel der Ziehung „ansteigende Zahlen“ lautet, ist sehr gering im Vergleich zur Wahrscheinlichkeit, dass ein beliebiges ansteigendes Zahlentripel gezogen wird (z.B. 18, 179, 657). Die Hypothese „ansteigende Zahlen“ wird durch das Zahlentripel 2, 4, 6 natürlich nicht widerlegt, erscheint aber doch unwahrscheinlich.

557 KLAYMAN und HA kritisieren WASON, weil die von ihm konstruierte Situation, in der die (nahe gelegte) Anfangshypothese enger ist als die gesuchte Hypothese, in Wirklichkeit kaum vorkomme. Meistens ist die getestete Hypothese zu weit, und die gesuchte Hypothese ist enger. In diesem Fall, wenn zudem positive Instanzen der Theorie seltener sind als negative Instanzen, ist die Suche nach positiven Instanzen der Theorie durchaus rational, weil sie häufiger zur Falsifizierung der Hypothese führt.<sup>879</sup> Dass die positiven Instanzen einer Theorie seltener sind als die negativen, ist ebenfalls eine realistische Annahme: man sucht nach den Persönlichkeitsmerkmalen, die zur Straffälligkeit führen, oder nach Gründen, die zu ausserordentlichem geschäftlichem Erfolg führen, nicht nach Merkmalen oder Gründen, die zur Einhaltung der Gesetze oder durchschnittlichem Erfolg führen. Unter den Fällen, die von der Hypothese nicht vorausgesagt werden, nach falsifizierenden Instanzen zu suchen, gleicht hier der sprichwörtlichen Suche nach der Nadel im Heuhaufen.<sup>880</sup> KLAYMAN und HA schlagen daher vor, die Strategie der Versuchspersonen in WASONs Zahlentripel-Aufgabe besser „positive Teststrategie“ statt „Bestätigungsstrategie“ zu nennen.<sup>881</sup>

558 Das Zahlentripel-Experiment ist nicht geeignet, eine Bestätigungstendenz nachzuweisen, weil die Versuchspersonen nicht wissen können, ob die nachgefragte Information (die Antwort des Versuchsleiters) bestätigend oder widerlegend sein wird. Experimente, bei denen die Versuchspersonen von vorneherein wissen, ob die nachgefragte Information ihre Hypothese stützt oder nicht, leiden nicht an diesem methodischen Defizit und sind daher geeignet, eine Bestätigungstendenz bei der Informationssuche nachzuweisen. Zahlreiche Studien weisen eine solche verzerrte Informationssuche sowohl bei Individuen wie bei Gruppen nach.<sup>882</sup> Die bestätigende Tendenz ist umso stärker, je mehr die Versuchsperson an der zu überprüfenden Hypothese „hängt“. Dies wird – z.B. durch JONAS et al. – dissonanztheoretisch erklärt. Vereinfacht gesagt entstehen kognitive Dissonanzen dann, wenn zwei Wahrnehmungen eines Menschen mit einer subjektiven Hypothese über deren Beziehung im Widerspruch stehen. Wer an eine Theorie glaubt, hofft, dass die Fakten die Theorie stützen. Tun sie dies nicht, entsteht eine Dissonanz zwischen Theorie und Fakten, die aufgelöst werden muss. Um diese Dissonanz zu vermeiden, sucht man unbewusst gar nicht erst nach Informationen, die im Widerspruch zur Hypothese stehen könnten.

---

<sup>879</sup> KLAYMAN/HA, FN 862, 217.

<sup>880</sup> Ist aber in einem zweiten Schritt, wenn die Hypothese aufgrund der positiven Teststrategie nicht falsifiziert wurde, dennoch notwendig, da ansonsten der Likelihood-Quotient nicht gebildet werden kann.

<sup>881</sup> KLAYMAN/HA, FN 862, 212

<sup>882</sup> Nachweise bei EVA JONAS/STEFAN SCHULZ-HARDT/DIETER FREY, Konfirmatorische Informationssuche bei simultaner vs. sequentieller Informationsvorgabe, Zeitschrift für Experimentelle Psychologie 2001, 239-247.

- 559 Bei sequentieller Informationsvorgabe – d.h. die Versuchspersonen erhalten eine Information und sagen dann, welche weitere Information sie erhalten möchten – ist der *confirmation bias* stärker als bei gleichzeitiger Vorgabe der Informationen (d.h. die Versuchspersonen wählen aus den zur Verfügung stehenden Informationen diejenigen aus, die sie erfahren möchten).<sup>883</sup> Gruppen, deren Mitglieder nicht alle die gleiche Präferenz teilen, suchen weniger ausgeprägt nach bestätigender Information. Die Heterogenität der Gruppe als solche, z.B. ob die Gruppe aus unterschiedlichen Persönlichkeiten oder Personen beider Geschlechts zusammengesetzt ist, hat hingegen keinen Einfluss auf den *confirmation bias*.<sup>884</sup>
- 560 CLOYD und SPILKER haben die bestätigende Informationssuche im juristischen Kontext untersucht.<sup>885</sup> Ihre Versuchspersonen waren 71 Steuerberater der (damals noch) Big Five Buchprüfungsfirmen, die im Schnitt rund 2,5 Jahre Berufserfahrung hatten. Allen wurde der gleiche Sachverhalt vorgelegt, gemäss dem ein fiktiver Klient ein Grundstück verkauft hatte. Der Hälfte der Versuchspersonen wurde gesagt, dass der Klient beim Verkauf einen Gewinn von einer halben Million Dollar gemacht, der anderen Hälfte, dass er einen Verlust in gleicher Höhe erlitten habe. Der Klient, der einen Gewinn gemacht hat, möchte als „Investor“ im Sinne des anwendbaren Steuerrechts behandelt werden, weil dann der Gewinn als Kapitalgewinn steuerfrei ist. Umgekehrt möchte der Klient, der einen Verlust erlitten hat, als „Händler“ qualifiziert werden, weil er dann den geschäftsmässigen Verlust vom Einkommen absetzen kann. Der Sachverhalt liess beide Qualifikationen zu.<sup>886</sup>
- 561 Die Versuchspersonen hatten eine halbe Stunde Zeit, eine Datenbank nach einschlägigen Urteilen zu durchsuchen und die Urteile abzuspeichern, die sie in einem Memorandum zur Frage zitieren würden. Die speziell erstellte Datenbank enthielt gleich viele Urteile, in denen ein Gericht erkannt hatte, dass ein Steuerpflichtiger als Investor zu behandeln ist, wie Urteile, in denen erkannt wurde, dass er als Händler zu behandeln ist.<sup>887</sup>
- 562 Die Versuchspersonen in der „Investor“ Gruppe studierten während 16,6 Minuten Urteile, in denen ein Steuerpflichtiger als Investor qualifiziert wurde, aber nur während 6,2 Minuten Urteile, in denen entschieden wurde, dass der Steuerpflichtige ein Händler ist. Sie speicherten im Schnitt 4,5 „Investor“ Fälle gegenüber nur 1,6 „Händler“ Fälle ab. Die Gruppe „Händler“ andererseits betrachtete während 12,3 Minuten „Händler“ Fälle und nur während 7,6 Minuten „Investor“ Fälle. Die Versuchspersonen in dieser Gruppe speicherten im Schnitt 3,9 „Händler“ und 1,6 „Investor“ Fälle ab.<sup>888</sup> Die Versuchspersonen wurden anschliessend gefragt, wie gross die Chance sei, dass ein Gericht die vom Klienten bevorzugte Sicht einnehmen würde. Die Gruppe „Händler“ sah eine Chance von 49 %, dass der

---

<sup>883</sup> JONAS/SCHULZ-HARDT/FREY, FN 882, 244.

<sup>884</sup> STEFAN SCHULZ-HARDT/DIETER FREY/KRISTIANE FAGO/GÜLER KICI, Selektive Informationssuche und Gruppenheterogenität: Der Einfluss verschiedener Formen der Gruppenheterogenität auf Selbstbestätigungsprozesse bei Entscheidungen, Gruppendynamik 1999, 161-174.

<sup>885</sup> C. BRYAN CLOYD/BRIAN C. SPILKER, The Influence of Client Preferences on Tax Professionals' Search for Judicial Precedents, Subsequent Judgments and Recommendations, The Accounting Review 1999, 299-322.

<sup>886</sup> CLOYD/SPILKER, FN 885, 307.

<sup>887</sup> CLOYD/SPILKER, FN 885, 307.

<sup>888</sup> CLOYD/SPILKER, FN 885, 309.

## Besonderer Teil

Klient als Händler qualifiziert würde, während die Gruppe „Investor“ eine Chance von 66 % sah, dass der Klient als Investor eingestuft würde. Die tatsächliche Wahrscheinlichkeit sollte selbstverständlich nicht von der Präferenz des Klienten abhängen. Weniger als 11 % der Versuchspersonen gaben eine Wahrscheinlichkeit zwischen 33 % und 66 % an, die bei einem ausgewogenen Sachverhalt und neutraler Beurteilung zu erwarten wäre.<sup>889</sup> Eine Pfadanalyse zeigt, dass die unterschiedliche Beurteilung in erster Linie auf die verzerrte Informationssuche zurückzuführen ist.<sup>890</sup>

- 563 In einer weiteren Studie verwendeten CLOYD und SPILKER einen Sachverhalt, der eigentlich nur den Schluss zulies, dass der Klient als Händler einzustufen ist (vier neutrale Experten beurteilten unabhängig die Wahrscheinlichkeit, dass der Klient in diesem Fall von einem Gericht als „Investor“ behandelt würde, als zwischen 10 und 20 %). Die Experten waren der Meinung, dass eine Steuerabrechnung als „Investor“ in diesem Fall sogar zu Sanktionen führen würde. Allen Versuchspersonen der zweiten Studie wurde gesagt, dass der Klient eine Behandlung als Investor vorziehen würde, weil er einen Gewinn gemacht hatte. 46 % der Versuchspersonen empfahlen, dass der Klient als Investor abrechnen sollte. Diese Versuchspersonen verbrachten signifikant mehr Zeit als die Versuchspersonen, die die Abrechnung als Händler empfahlen, mit der Suche nach Urteilen, in denen Gerichte erkannt hatten, dass der Steuerpflichtige als Investor zu betrachten ist und speicherten signifikant mehr „Investor“-Urteile ab als die 54 % der Versuchspersonen, die zur richtigen Entscheidung gelangten.<sup>891</sup> Aufgrund der Korrelation zwischen Suchverhalten und Empfehlung darf geschlossen werden, dass das verzerrte Suchverhalten die (allzu aggressive) Empfehlung verursachte.<sup>892</sup>
- 564 Für Juristen möglicherweise tröstlich ist, dass eine weitere Untersuchung von CLOYD/SPILKER zeigt, dass Rechtsstudierende bei der Beurteilung eines ambivalenten Sachverhalts weniger als Studierende eines Buchhaltungslehrganges (*Master of Accounting*) zur confirmatorischen Informationssuche neigen. Offenbar scheint die Ausbildung – zumindest an amerikanischen *Law Schools*, die nicht in jeder Hinsicht vergleichbar ist mit der Ausbildung an einer schweizerischen Rechtsfakultät – geeignet zu sein, gewissen „Wahrnehmungsstörungen“ entgegenzuwirken.<sup>893</sup>
- 565 KASSIN und Kollegen können zeigen, dass Verhörende, die von der Schuld des Verhörten überzeugt sind, aggressivere Einvernahmetechniken verwenden und mehr Druck auf den Verhörten ausüben; speziell wenn dieser (entgegen ihrer Vorstellung) unschuldig ist.<sup>894</sup>

---

<sup>889</sup> CLOYD/SPILKER, FN 885, 310.

<sup>890</sup> CLOYD/SPILKER, FN 885, 313.

<sup>891</sup> CLOYD/SPILKER, FN 885, 316.

<sup>892</sup> Dieser Schluss von Korrelation auf Kausation ist hier zulässig, weil in dem Laborexperiment die übrigen Variablen kontrolliert werden konnten.

<sup>893</sup> C. BRYAN CLOYD/BRIAN C. SPILKER, Confirmation Bias in Tax Information Search: A Comparison of Law Students and Accounting Students, *Journal of the American Taxation Association* 2000, 60-71.

<sup>894</sup> SAUL M. KASSIN/CHRISTINE C. GOLDSTEIN/KENNETH SAVITSKY, Behavioral Confirmation in the Interrogation Room: On the Dangers of Presuming Guilt, *Law and Human Behavior* 2003, 187-203.

Dies führt dazu, dass neutrale Beobachter das Verhalten der Verhörten als defensiver einstufen und die Verhörten als geringfügig schuldiger betrachten.<sup>895</sup>

## II. Informationsbewertung

### A. Ambivalente Informationen werden als Bestätigung interpretiert

566 Nicht jedes Beweismittel spricht eindeutig für oder gegen eine Hypothese. Gerade im Recht sind Informationen häufig, die so oder anders interpretiert werden können. Siehe nur das folgende Gedankenexperiment von CURT GOETZ:<sup>896</sup>

Sie gehen nachts spazieren und finden einen Mann in einer Blutlache liegen. Sie knien nieder ... ziehen ein Messer aus der Brust ... In diesem Augenblick hören sie Schritte, und plötzlich kommt ihnen ihre Situation zu Bewusstsein: ... der Tote kann nicht mehr für sie zeugen – und sie rennen davon, wie von Furien gejagt. Aber man fängt sie.

567 CURT GOETZ wünscht dem Betroffenen, dass er auf psychologisch geschulte Richter trifft. Man könnte auch sagen, auf Richter, die sich bewusst sind, dass Menschen die Neigung haben, ambivalente Informationen im Sinne der von ihnen als wahrscheinlich erachteten Hypothese zu interpretieren.<sup>897</sup> Es ist offensichtlich, dass die Informationen in GOETZ' Gedankenexperiment ganz unterschiedlich bewertet werden können, je nachdem, ob man an die Schuld oder Unschuld des Gefassten glaubt.

568 Eines der bekanntesten Experimente zur Interpretation von Fakten im Lichte der eigenen Theorie stammt von LORD, LEPPER und ROSS.<sup>898</sup> LORD und Kollegen wählten Versuchspersonen aus, die entweder Befürworter oder Gegner der Todesstrafe waren. Die Versuchspersonen erhielten zwei Studien zur Lektüre, die scheinbar empirische Belege für oder gegen die Todesstrafe lieferten. Sowohl Befürworter wie Gegner der Todesstrafe bewerteten die Studie, die ihre Position scheinbar stützte, als überzeugender und besser durchgeführt als die Studie, die gegen ihre Position sprach. Zweitens führte die Lektüre beider Studien dazu, dass die Überzeugungen der beiden Gruppen noch mehr polarisiert wurden; d.h. sowohl Befürworter als auch Gegner waren nach Kenntnisnahme aller Beweismittel – auch derjenigen, die ihrer Position widersprachen – von ihrer Position überzeugter als zuvor. Diese Polarisation entstand dadurch, dass die Überzeugung der Versuchspersonen jedes Mal erheblich zunahm, wenn sie ein Argument *für* ihre Position lasen, aber nur unwesentlich abnahm, wenn sie ein Argument gegen die Position lasen. KOEHLER hat in einer ähnlichen Studie Anhänger und Gegner parapsychologischer Phänomene gebeten, die Qualität von wissenschaftlichen Untersuchungen solcher Phänomene zu be-

---

<sup>895</sup> KASSIN/GOLDSTEIN/SAVITSKY, FN 894, 198.

<sup>896</sup> Zitiert nach BENDER/NACK, FN 605, X.

<sup>897</sup> Zu dieser Tendenz LEE ROSS/CRAIG A. ANDERSON, Shortcomings in the Attribution Process, in: TVERSKY/KAHNEMAN/SLOVIC (Hrsg.), 129-152, 144 ff.; KLAYMAN, FN 855, 394.

<sup>898</sup> CHARLES G. LORD/LEE ROSS/MARK R. LEPPER, Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence, Journal of Personality and Social Psychology 1979, 2098-2109.



## Besonderer Teil

werten. Beide Gruppen bewerteten Studien, die ihre Position stützten, als qualitativ besser.<sup>899</sup>

- 569 Zwei Drittel der Versuchspersonen von KOEHLER gaben an, dass ihre Bewertung von ihrer Einstellung zu den untersuchten Phänomenen nicht beeinflusst werde; 85 % meinten, dass ihre Bewertung nicht von der Einstellung beeinflusst werden sollte.<sup>900</sup> KOEHLER weist (wie bereits LORD et al.) darauf hin, dass beide Aussagen falsch sind.<sup>901</sup> Es ist normativ richtig, die Resultate einer Studie mit einer gesunden Portion Skepsis zu betrachten, wenn sie einem theoretisch gut fundierten, durch zahlreiche unabhängige Experimente bestätigten (richtiger: nicht falsifizierten) Modell widersprechen. Wenn ein unbekannter Wissenschaftler berichtet, dass Schweine fliegen können, ist die Wahrscheinlichkeit, dass seine Beobachtung falsch ist, grösser, als dass alle Zoologen der Menschheitsgeschichte falsch lagen. Problematisch wird die Skepsis dann, wenn der Urteilende zwischen zwei Theorien entscheiden muss, die beide durch die vorliegenden (ambivalenten) Daten gestützt werden können, wenn man die Daten nur „richtig“ liest und keine der beiden Theorien *a priori* eine grössere Überzeugungskraft besitzt. Dies ist aber genau die Situation, in der sich der Richter oft befindet, wenn er über die Stichhaltigkeit der Theorien von Anklage und Verteidigung, von Kläger und Beklagtem, entscheiden muss. In diesem Fall fällt der Entscheid – in Grenzfällen – regelmässig auf die Seite, die der Richter, und sei es auch nur als „Arbeitshypothese“, ursprünglich als zutreffend erachtet.<sup>902</sup> Dies kann, siehe hinten, S. 188 ff., beispielsweise die These sein, die durch die zuerst vorgebrachten Beweismittel gestützt wird.

### **B. Bestätigende Informationen werden stärker gewichtet als widersprechende Informationen**

- 570 Beweismittel, die für eine von der Versuchsperson gehaltene These sprechen, werden besser wahrgenommen, erinnert und stärker gewichtet als Beweismittel, die gegen die These sprechen.<sup>903</sup> Nach GADENNE und OSWALD werden hypothesenbestätigende Informationen zu stark gewichtet, während hypothesekonträre Informationen erst als wichtig wahrgenommen werden, wenn sie der Hypothese deutlich widersprechen. Auch bei stärkster konträrer Information bleibt die anfängliche Hypothese immer noch die stärkste Hypothese; erst der deutliche Hinweis auf eine alternative Erklärung vermag den Effekt (weit-

---

<sup>899</sup> JONATHAN J. KOEHLER, The Influence of Prior Beliefs on Scientific Judgements of Evidence Quality, *Organizational Behavior and Human Decision Making Processes* 1993, 28-55; siehe auch TOM PYSZCZYNSKI/JEFF GREENBERG, Toward an Integration of Cognitive and Motivational Perspectives on Social Inference: A Biased Hypothesis Testing Model, in: LEONARD BERKOWITZ (Hrsg.), *Advances in Experimental Social Psychology*, Vol. 30, San Diego etc. 1987, 297-340, 329.

<sup>900</sup> KOEHLER, FN 899, 45 f.

<sup>901</sup> KOEHLER, FN 899, 30 f.; LORD/ROSS/LEPPER, FN 898, 2106.

<sup>902</sup> DEREK J. KOEHLER, Explanation, Imagination, and Confidence in Judgment, *Psychological Bulletin* 1991, 499-519.

<sup>903</sup> NICKERSON, FN 855, 178; KLAYMAN, FN 855, 395; RICHARD NISBETT/LEE ROSS, *Human Inference: Strategies and Shortcomings of Social Judgment*, Englewood Cliffs 1980, 181 f.; MARGRIT E. OSWALD, Hypothesentesten: Suche und Verarbeitung hypothesenkonformer und hypothesekonträrer Informationen, in: HELL/FIEDLER/GIGERENZER (Hrsg), FN 11, 189-212, 197; jeweils mit zahlreichen Hinweisen.

gehend) zum Verschwinden zu bringen.<sup>904</sup> Gemäss der bereits vorne, S. 155, erwähnten Studie von SCHUM und MARTIN neigen Menschen dazu, bei der globalen Bewertung der Beweismittel konträre Beweismittel schlicht zu ignorieren oder als hypothesenbestätigend zu interpretieren.<sup>905</sup> Man hat die Tendenz, sich darauf zu konzentrieren, wie gut die Beweismittel eine Hypothese stützen, und vergisst dabei zu überlegen, wie gut die gleichen Beweismittel auch mit einer alternativen Hypothese in Einklang stehen.<sup>906</sup>

571 LINDA JOHNSON legte 109 Steuerberatern vier Gerichtsurteile zu einer umstrittenen steuerlichen Abgrenzungsfrage vor. Der Sachverhalt blieb bei jedem Urteil immer derselbe, aber der Ausgang des Verfahrens wurde so manipuliert, dass in jeder Versuchsgruppe zwei andere Fälle zu einem für den Klienten günstigen Schluss kamen. Die Versuchspersonen wurden gebeten, die Relevanz der Urteile – die in erster Linie davon abhängt, ob die Fälle wegen des übereinstimmenden Sachverhalts gute Präjudizien sind – für den von ihnen zu beurteilenden Fall zu bewerten. Urteile, die zu einem für den Klienten positiven Schluss kamen, werden als relevanter beurteilt; diese Tendenz beeinflusste wiederum die Beurteilung der Prozesschancen durch die Steuerberater.<sup>907</sup>

572 Eine Studie von BABCOCK et al. belegt ebenfalls den Einfluss, den die Perspektive auf die Beurteilung der Prozesschancen haben kann. Wie vorne, S. 91, dargelegt, hängt es nach dem ökonomischen Modell von Vergleichsverhandlungen in erster Linie von der Einschätzung des Prozessausgangs durch die Parteien ab, ob die Parteien sich aussergerichtlich einigen. Die korrekte Voraussage des richterlichen Urteils hängt wiederum von den Fähigkeiten der Parteien ab, die Stärken und Schwächen ihres Falles unvoreingenommen zu beurteilen. Wenn bereits die Wahrnehmung des Falles verzerrt ist, werden auch die Urteilerwartungen systematisch verzerrt sein, und es kommt seltener zum Vergleich. BABCOCK et al. haben genau diese verzerrte Einschätzung der eigenen Prozesschancen nachgewiesen.<sup>908</sup> Wenn alle ihre Versuchspersonen *zuerst* die Unterlagen zum Streitfall studierten und *dann* zufällig in die Gruppen „Beklagte“ oder „Kläger“ eingeteilt wurden, einigten sich 94 % der Paare auf einen Vergleich. Wurden die Versuchspersonen *vor* der Lektüre der Unterlagen in die Gruppen „Beklagte“ und Kläger eingeteilt, so einigten sich nur 72 % der Verhandelnden; ein signifikanter Unterschied.<sup>909</sup> Die Differenz zwischen dem vom Kläger und vom Beklagten erwartetem Urteil war im ersten Fall im Schnitt nur \$ 6'936, im zweiten Fall aber \$ 18'555. Die Lektüre des Sachverhalts aus der Sicht einer

---

<sup>904</sup> VOLKER GADENNE/MARGRIT OSWALD, Entstehung und Veränderung von Bestätigungstendenzen beim Testen von Hypothesen, Zeitschrift für experimentelle und angewandte Psychologie 1986, 360-374.

<sup>905</sup> SCHUM/MARTIN, FN 770, 164.

<sup>906</sup> DALE GRIFFIN/AMOS TVERSKY, The Weighing of Evidence and the Determination of Confidence, Cognitive Psychology 1992, 411-435; Nachdruck in: GILOVICH/GRIFFIN/KAHNEMAN (Hrsg.), 230-249, 238.

<sup>907</sup> LINDA M. JOHNSON, An Empirical Investigation of the Effects of Advocacy on Preparers' Evaluations of Judicial Evidence, Journal of the American Taxation Association 1993, 1-22.

<sup>908</sup> LINDA BABCOCK/GEORGE LOEWENSTEIN/SAMUEL ISSACHAROFF/COLIN CAMERER, Biased Judgments of Fairness in Bargaining, American Economic Review 1995, 1337-1342.

<sup>909</sup> BABCOCK/LOEWENSTEIN/ISSACHAROFF/CAMERER, FN 908, 1340.

Partei hatte die Wahrnehmung der Versuchspersonen so verändert, dass ihre Einschätzung des Prozessausgangs systematisch zu ihren Gunsten verzerrt war.<sup>910</sup>

### III. Folgen der Bestätigungstendenzen

#### A. Festhalten selbst an einer diskreditierten Hypothese (*belief perseverance*)

573 Als *belief perseverance* wird das Phänomen bezeichnet, dass Leute selbst an einer völlig diskreditierten Theorie festhalten.<sup>911</sup> In der ursprünglichen Studie von ROSS et al. wurden Versuchspersonen gebeten, echte von unechten Abschiedsbriefen von Selbstmördern zu unterscheiden. Der „Erfolg“ der Versuchspersonen war dabei rein zufällig und hing von der Gruppe ab, der sie zugeteilt worden waren. Auch nachdem sie darüber aufgeklärt worden waren, beurteilten die Versuchspersonen in der „Erfolgs“ Gruppe ihre Fähigkeit, ähnliche Aufgaben zu lösen, höher als die Versuchspersonen in der „Misserfolgs“ Gruppe. Zahlreiche weitere Studien bestätigten den ursprünglichen Befund von ROSS et al.<sup>912</sup>

#### B. Einfluss der Reihenfolge der Beweismittel (*primacy effect*)

574 Die ursprüngliche Hypothese, die die Wahrnehmung und/oder Gewichtung von Beweismitteln beeinflusst, muss erst einmal generiert werden. Nicht in allen Fällen hat der Urteilende von vorneherein eine Meinung zum Thema; gerade im juristischen Bereich dürfte er häufig vorerst unvoreingenommen sein. Die Forschung zum *primacy effect* legt nahe, dass die Hypothese, die von dem oder den ersten Beweismitteln favorisiert wird, zur Arbeitshypothese wird und die Bewertung der weiteren Beweismittel beeinflusst.

575 Der *primacy effect* wurde erstmals von SOLOMON ASCH beschrieben. ASCH beschrieb der einen Hälfte seiner Versuchspersonen eine Person als intelligent-fleissig-impulsiv-kritisch-eigensinnig-neidisch, der anderen Hälfte als neidisch-eigensinnig-kritisch-impulsiv-fleissig-intelligent. Die Versuchspersonen in der ersten Gruppe beurteilten die beschriebene Person durchwegs positiver als diejenigen in der zweiten Gruppe.<sup>913</sup> ASCH selber interpretierte die Ergebnisse dahingehend, dass die Bedeutung der späteren Adjektive durch die zuerst wahrgenommenen Adjektive geformt wird; d.h. „eigensinnig“ wird positiv als „eigenständig“ kodiert, weil der betreffende Mensch intelligent ist und daher seinen Mitmenschen zu Recht häufig widerspricht. Diese Interpretation der Ergebnisse hat ihre Anhänger und Gegner; eine Gegenmeinung ist, dass die späteren Adjektive nicht anders codiert werden, sondern schlicht weniger Gewicht erhalten.<sup>914</sup> Die Annahme, dass der *primacy effect* durch die frühe Bildung einer Hypothese entsteht, ist mit beiden Auffassungen kom-

---

<sup>910</sup> Siehe auch GEORGE LOEWENSTEIN/SAMUEL ISSACHAROFF/COLIN CAMERER/LINDA BABCOCK, Self-Serving Assessments of Fairness and Pretrial Bargaining, *Journal of Legal Studies* 1993, 135-159.

<sup>911</sup> LEE ROSS/MARK R. LEPPER/MICHAEL HUBBARD, Perseverance in Self Perception and Social Perception: Biased Attributional Processes in the Debriefing Paradigm, *Journal of Personality and Social Psychology* 1975, 880-892.

<sup>912</sup> Nachweise bei NICKERSON, FN 855, 187; ROSS/NISBETT, FN 903, 175 ff.

<sup>913</sup> SOLOMON ASCH, Forming Impressions of Personality, *Journal of Abnormal and Social Psychology* 1946, 258-290, 270 f.

<sup>914</sup> Nachweise bei NISBETT/ROSS, FN 903, 173 f.

patibel.<sup>915</sup> Die ebenfalls nahe liegende Auffassung, dass die stärkere Beachtung zuerst dargebotener Beweismittel einfach darauf zurückzuführen ist, dass die Aufmerksamkeit der Versuchspersonen nachlässt, gilt als durch Experimente von JONES und GOETHALS widerlegt.<sup>916</sup>

576 Studien, die einen *primacy effect* nachweisen, sind ausserordentlich zahlreich.<sup>917</sup> Manche Studien berichten aber auch von einem *recency effect*, d.h., die zuletzt dargebotenen Informationen beeinflussen das Urteil stärker. Dies erscheint paradox. Der *recency effect* scheint aber in erster Linie ein Gedächtnisphänomen zu sein; er tritt dann auf, wenn den Versuchspersonen relativ zahlreiche Informationen in schneller Folge dargeboten werden. In diesen Fällen erinnern sie sich nicht mehr an die ersten Informationen; diese können folglich auch das Urteil nicht mehr beeinflussen.<sup>918</sup> *Recency effects* treten zudem häufiger auf, wenn die Versuchspersonen nach jedem einzelnen Beweismittel ein Zwischenurteil abgeben, während *primacy effects* vorwiegend dann beobachtet werden, wenn die Beurteilung erst am Schluss, alle Informationen gesamthaft vorliegen, abgegeben wird.<sup>919</sup>

577 Studien mit spezifisch juristischen Fragestellungen haben, wie andere Studien, sowohl *recency* wie *primacy effects* gefunden.<sup>920</sup> PENNINGTON stellt im Gegensatz zu THIBAUT und WALKER einen klaren *primacy effect* fest und schreibt dies der grösseren Realitätsnähe seiner Studie zu.<sup>921</sup> Die neuste verfügbare Studie von COSTABILE/KLEIN findet wiederum *recency effects*, schreibt diese aber selber in erster Linie dem Gedächtnis der Geschworenen zu, die sich an Beweismittel, die erst spät eingeführt wird, im Zeitpunkt des Urteils besser erinnern.<sup>922</sup> Bei der Interpretation amerikanischer Studien zum Einfluss der Reihenfolge der Beweismittel auf Urteile von Geschworenen ist Vorsicht geboten, weil das amerikanische Prozessrecht sich grundlegend vom schweizerischen unterscheidet. In amerikanischen Geschworenenprozessen gilt das Unmittelbarkeitsprinzip und den Geschworenen ist es verboten, Notizen zu machen. Diese beiden Faktoren fördern natürlich *recency effects* stark; man kann sich vorstellen, wie schwierig es ist, sich nach einer Reihe von Zeugeneinvernahmen noch an die Aussagen des ersten Zeugen zu erinnern, wenn man keine Notizen machen darf.

---

<sup>915</sup> NISBETT/ROSS, FN 903, 174.

<sup>916</sup> NISBETT/ROSS, FN 903, 174 f.

<sup>917</sup> Nachweise bei NISBETT/ROSS, FN 903, 172 ff.; NICKERSON, FN 855, 187.

<sup>918</sup> NISBETT/ROSS, FN 903, 172.

<sup>919</sup> ROBIN M. HOGARTH/HILLEL J. EINHORN, Order Effects in Belief Updating: The Belief-Adjustment Model, *Cognitive Psychology* 1992, 1-55, 5.

<sup>920</sup> DONALD C. PENNINGTON, Witnesses and Their Testimony: Effects of Ordering on Juror Verdicts, *Journal of Applied Social Psychology* 1982, 318-333, mit Nachweisen.

<sup>921</sup> PENNINGTON, FN 920, 330; JOHN THIBAUT/LAURENS WALKER, *Procedural Justice: A Psychological Analysis*, Hillsdale 1975, 61.

<sup>922</sup> KRISTI A. COSTABILE/STANLEY B. KLEIN, Finishing Strong: Recency Effects in Juror Judgments, *Basic and Applied Social Psychology* 2005, 47-58.

**C. Zu grosse Sicherheit im Urteil (*overconfidence*)**

- 578 Die ungenügende Berücksichtigung und Gewichtung widersprüchlicher Informationen kann dazu führen, dass sich der Entscheidende seiner Sache zu sicher ist (mehr zur Selbstüberschätzung generell hinten, S. 261 ff.).<sup>923</sup> Frauen, die weniger zur Selbstüberschätzung neigen als Männer, bewerten hypothesenkonträre Informationen als wichtiger und empfinden (daher) die zu beurteilende Frage als schwieriger, was zu entsprechend geringerer empfundener Sicherheit bezüglich der Richtigkeit ihrer Antwort führt.<sup>924</sup> Die Annahme, dass die einseitige Gewichtung der Argumente zu allzu grosser Gewissheit führt, dass das eigene Urteil richtig ist, wird durch Experimente gestützt, in denen die Versuchspersonen gezwungen werden, Gründe aufzulisten, die gegen die von ihnen getroffene Entscheidung sprechen.<sup>925</sup> Wenn sie dies tun, stimmt ihre Einschätzung der Wahrscheinlichkeit, dass ihr Urteil richtig ist, eher mit der tatsächlichen Häufigkeit richtiger Urteile überein.<sup>926</sup>

**IV. Eigene Studie zum Bestätigungsfehler**

**A. Frage zur Pseudodiagnostizität**

- 579 Den Richterinnen und Richtern der Zivil- und Strafgerichte der Kanton beider Basel, Bern und Graubünden wurde folgender Sachverhalt vorgelegt:

Eine Fussgängerin wird von einem Auto angefahren und leicht verletzt. Der Fahrer des Autos begeht Fahrerflucht.

Die Fussgängerin sagt aus, sie sei entweder von einem Ford oder einem Opel angefahren worden. Sie kann nicht sagen, welche der beiden Marken es war, aber sie erinnert sich, dass es sich um einen schwarzen Kombi mit Schweizer Nummernschildern handelte.

- 580 Die Hälfte der Fragebogen („plausible Arbeitshypothese“) setzen sich wie folgt fort:

Sie wissen folgendes:

- A. Es gibt gleich viele Personenwagen der Marken Opel und Ford in der Schweiz.
- B. 60 % aller in der Schweiz zugelassener Personenwagen der Marke Opel sind Kombis.

Folgende Informationen sind ebenfalls erhältlich:

- C. Der prozentuale Anteil Kombis an der Gesamtheit in der Schweiz zugelassener Personenwagen der Marke Ford.
- D. Der prozentuale Anteil schwarzer Fahrzeuge an der Gesamtheit in der Schweiz zugelassener Personenwagen der Marke Opel.
- E. Der prozentuale Anteil schwarzer Fahrzeuge an der Gesamtheit in der Schweiz zugelassener Personenwagen der Marke Ford.

---

<sup>923</sup> NICKERSON, FN 855, 189.

<sup>924</sup> JANNE CHUNG/GARY S. MONROE, Gender Differences in Information Processing: An Empirical Test of the Hypothesis-Confirming Strategy in an Audit Context, *Accounting and Finance* 1998, 256-279.

<sup>925</sup> ASHER KORIAT/SARAH LICHTENSTEIN/BARUCH FISCHHOFF, Reasons for Confidence, *Journal of Experimental Psychology: Human Learning and Memory* 1980, 107-118.

<sup>926</sup> KORIAT/LICHTENSTEIN/FISCHHOFF, FN 925, 110.

581 Die andere Hälfte („unplausible Arbeitshypothese“) lautete folgendermassen:

Sie wissen folgendes:

- A. Es gibt gleich viele Personenwagen der Marken Opel und Ford in der Schweiz.
- B. 10 % aller in der Schweiz zugelassener Personenwagen der Marke Ford sind schwarz.

Folgende Informationen sind ebenfalls erhältlich:

- C. Der prozentuale Anteil schwarzer Fahrzeuge an der Gesamtheit in der Schweiz zugelassener Personenwagen der Marke Opel.
- D. Der prozentuale Anteil Kombis an der Gesamtheit in der Schweiz zugelassener Personenwagen der Marke Ford.
- E. Der prozentuale Anteil Kombis an der Gesamtheit in der Schweiz zugelassener Personenwagen der Marke Opel.

582 In beiden Fällen wurden die Richterinnen und Richter gefragt, welche der zusätzlichen Informationen C, D oder E ihnen hilft entscheiden, ob die Fussgängerin einen Ford oder einen Opel gesehen hat, wenn keine anderen als die zur Auswahl stehenden Informationen erhältlich sind. In beiden Gruppen ist die Antwort C diagnostisch, die Antwort D pseudodiagnostisch für die Arbeitshypothese und die Antwort E pseudodiagnostisch für die alternative Hypothese. Die Gruppen werden als „plausibel“ und „unplausibel“ bezeichnet, weil das bekannte Indiz „Art des Fahrzeuges“ in der ersten Gruppe bei mehr als 50 % der Instanzen der Kategorie „Opel“ vorliegt, während das bekannte Indiz „Farbe“ in der zweiten Gruppe nur bei 10 % der Instanzen der Kategorie „Ford“ vorliegt. Folglich erhält die *focal hypothesis* (hier als „Arbeitshypothese“ bezeichnet) durch das bekannte Indiz der ersten Gruppe mehr Unterstützung als durch das bekannte Indiz der zweiten Gruppe (in der ersten Gruppe  $P(\text{Kombi} | \text{Opel}) = 0,6$ ; in der zweiten Gruppe  $P(\text{Schwarz} | \text{Ford}) = 0,1$ ). Nach MYNATT, DOHERTY und DRAGAN müsste dies dazu führen, dass die alternative Hypothese eher in Betracht gezogen wird und daher der Anteil pseudodiagnostischer Antworten sinkt (womit auch gleich die Hypothese formuliert wäre).<sup>927</sup>

583 Systematisch lassen sich die bekannten und erhältlichen Informationen in einer Matrix gemäss Tabelle 14 darstellen. Die Gruppe „plausible Arbeitshypothese“ kennt die Information in der Zelle A ( $P(\text{Kombi Opel}) = 0,6$ ). Diagnostisch ist die Information in Zelle B, während die Information in Zelle C weitere Informationen über die Arbeitshypothese („angenommen, es ist ein Opel“) liefert und gemäss üblichem Sprachgebrauch als pseudodiagnostisch bezeichnet wird. Die Zelle D wird in traditionellen Pseudodiagnose-Studien selten gewählt; sie ist pseudodiagnostisch für die Alternativhypothese.

	Opel ( $H_1$ )	Ford ( $H_2$ )
% Kombis ( $I_1$ )	Zelle A: $P(I_1   H_1)$ „Anteil Kombis an Opel“	Zelle B: $P(I_1   H_2)$ „Anteil Kombis an Ford“
% schwarz ( $I_2$ )	Zelle C: $P(I_2   H_1)$ „Anteil schwarz an Opel“	Zelle D: $P(I_2   H_2)$ „Anteil schwarz an Ford“

**Tabelle 14: Systematische Darstellung des Diagnose-Problems**

<sup>927</sup> MYNATT/DOHERTY/DRAGAN, FN 859, 766.

**B. Resultate**

584 190 der 234 Richterinnen und Richter, die den Fragebogen zumindest teilweise ausgefüllt zurückgeschickt hatten, beantworteten die Frage nach der Diagnostizität der Beweismittel. Die vergleichsweise hohe Ausfallquote von 18,8 % ist darauf zurückzuführen, dass relativ viele Richter meinten, keine der angebotenen Informationen würde es ihnen erlauben, die Frage, ob es sich um einen Opel oder einen Ford gehandelt habe, mit hinreichender Sicherheit zu entscheiden. Das ist zweifelsfrei richtig. Allerdings war die Frage, welche der Informationen *helfen* würde, zu entscheiden; und diese Frage kann meines Erachtens sinnvoll beantwortet werden. Sie mag zugegebenermassen wenig realistisch sein, weil man in einem tatsächlichen Gerichts- oder Untersuchungsverfahren den Aufwand, diese – auf jeden Fall nicht sehr diagnostischen – Informationen zu beschaffen, kaum treiben würde.

	Antworten			Total
	Diagnostisch (C)	PD für Arbeitsh. (D)	PD für Alternativh. (E)	
plausible Arbeitshypothese	34 (35 %)	44 (46 %)	18 (19 %)	96
unplausible Arbeitshypothese	38 (40 %)	26 (27 %)	30 (32 %)	94
	72	70	48	190

**Tabelle 15: Resultate zur Frage nach dem diagnostischen Beweismittel (PD = pseudodiagnostisch)**

585 Die Resultate der Frage zur Diagnostizität der Beweismittel sind in Tabelle 15 dargestellt. Der Einfluss der Plausibilität der Arbeitshypothese auf die Antworten ist statistisch signifikant.<sup>928</sup> Der Anteil der diagnostischen Antworten ist in beiden Gruppen ungefähr gleich gross (35 resp. 40 %). Der Trend – mehr diagnostische Antworten bei unplausibler Arbeitshypothese – zeigt in die erwartete Richtung. Ein erheblicher Unterschied zeigt sich bei den pseudodiagnostischen Antworten für die Arbeitshypothese; sie nehmen von 46 % bei plausibler Arbeitshypothese auf 27 % bei unplausibler Arbeitshypothese ab. Das Verhältnis von diagnostischen zu pseudodiagnostischen Antworten kehrt sich beim Wechsel von der plausiblen zur unplausiblen Arbeitshypothese um. Die Resultate von MYNATT, DOHERTY und DRAGAN wurden somit repliziert.<sup>929</sup>

586 Der Abnahme der Anzahl pseudodiagnostischer Antworten bei unplausibler Arbeitshypothese entspricht aber nicht, wie bei MYNATT et al., eine in etwa gleich starke Zunahme bei den diagnostischen Antworten. Stattdessen haben mehr Richterinnen und Richter in der Gruppe „unplausible Arbeitshypothese“ die pseudodiagnostische Information für die Alternativhypothese gewählt; nämlich 32 % gegenüber 19 %. Ein ähnliches Phänomen wurde von EVANS, VENN und FEENEY beobachtet, die ebenfalls einen stabilen Anteil an diagnostischen Antworten, aber eine Verschiebung zu Gunsten der sonst eher selten gewählten

<sup>928</sup>  $\chi^2 = 7,831, p = 0,02.$

<sup>929</sup> MYNATT/DOHERTY/DRAGAN, FN 859, 774.

Antwort „E“ (bei ihnen „D“) feststellten, als sie die Plausibilität der *focal hypothesis* von stark zu schwach manipulierten.<sup>930</sup>

### C. Diskussion

- 587 Die Resultate replizieren die Studie von MYNATT et al. und sprechen dafür, dass der Hauptgrund für die Wahl pseudodiagnostischer Informationen die Schwierigkeit ist, mehrere Alternativen gleichzeitig zu verfolgen. Stattdessen konzentriert sich der Mensch auf die Arbeitshypothese – wobei es bereits genügt, dass sich die verfügbaren Informationen auf diese Hypothese beziehen, damit eine Hypothese zur Arbeitshypothese wird – und sucht nach weiteren Informationen zu dieser Hypothese. Dabei vergisst man leicht, dass sich der abstrakte Beweiswert eines Indizes nur beurteilen lässt, wenn man weiss, wie oft es vorkommt, wenn die Arbeitshypothese nicht zutrifft.
- 588 Die beobachtete Verschiebung von Antworten „D“ zu Antworten „E“ bei der unplausiblen Hypothese wurde von MYNATT et al. nicht beobachtet, entspricht aber den Resultaten von EVANS et al. Diese erklären sie damit, dass die Versuchspersonen, die mit einer unplausiblen Arbeitshypothese konfrontiert sind, zu den Antworten „E“ wechseln, weil sie die Alternativhypothese zu ihrer neuen Arbeitshypothese machen. Die Antworten „E“ sind die pseudodiagnostischen Antworten für Leute, die die Alternativhypothese zu ihrer Arbeitshypothese gemacht haben.<sup>931</sup> So betrachtet verstärken diese Resultate die Überzeugungskraft der Theorie, dass pseudodiagnostische Informationssuche darauf zurückzuführen ist, dass man nicht mehr als eine Hypothese gleichzeitig im Arbeitsgedächtnis behalten kann.
- 589 Warum aber haben MYNATT et al. keine Verschiebung zugunsten von „E“ beobachtet? Eine spekulative, aber plausible, Erklärung ist, dass ihre Manipulation zur Diskreditierung der Plausibilität der Arbeitshypothese nicht ausreichend stark war, um einen Wechsel des Fokus zu bewirken. Bei MYNATT et al. betrug  $P(I_1 | H_1)$  bei der plausiblen Hypothese 0,65 und bei der unplausiblen Hypothese 0,35 (also rund die Hälfte), während bei der vorliegenden Studie  $P(I_1 | H_1)$  bei der plausiblen Hypothese 0,6 und bei der unplausiblen 0,1 – also sechs Mal weniger – betrug. Die Manipulation der Plausibilität der Arbeitshypothese war bei EVANS et al. m.E. ebenfalls stärker als bei MYNATT et al. Die Versuchspersonen bei EVANS et al. sollten aufgrund ganz ähnlicher Informationen, wie sie in der vorliegenden Studie verwendet wurden, herausfinden, welche Marke das Auto ihrer Schwester hat. Die Manipulation der Plausibilität der Arbeitshypothese geschah nicht über die Veränderung von  $P(I_1 | H_1)$ , sondern dadurch, dass den Versuchspersonen erklärt wurde, ihre Schwester lege grossen Wert auf ein sparsames Auto. Die Information ( $I_1 | H_1$ ), die den Versuchspersonen bekannt war, war entweder, dass das Modell Y 14 Liter, 8 Liter oder 6 Liter Benzin pro 100 Kilometer verbrauche.<sup>932</sup> Da es eher unwahrscheinlich ist, dass jemand, der grossen Wert auf ein sparsames Auto legt, ein Auto besitzt, das 14 Liter Benzin pro 100 Kilometer verbraucht, ist die Arbeitshypothese im ersten Fall am unplausibelsten;

---

<sup>930</sup> EVANS/VENN/FEENEY, FN 871, 36.

<sup>931</sup> EVANS/VENN/FEENEY, FN 871, 42.

<sup>932</sup> Im Original 20/25/50 (UK) mpg; EVANS/VENN/FEENEY, FN 871, 35.



die Versuchspersonen schlossen aus dieser Angabe wahrscheinlich, dass es sich beim Auto ihrer Schwester nicht um ein Modell Y handelt und konzentrierten sich auf das Modell X.

## V. Was kann gegen Bestätigungstendenzen unternommen werden?

### A. Kontrafaktisches Denken (*counterfactual reasoning*)

590 Als Strategie zur Vermeidung von Bestätigungstendenzen – und anderen kognitiven Täuschungen – hat sich das im Englischen als *counterfactual reasoning* bezeichnete Denken bewährt, das sich nur ungenügend mit „kontrafaktisches Denken“ ins Deutsche übersetzen lässt.<sup>933</sup> Wenn man kontrafaktisch denkt, stellt man sich vor, welche Bedingungen hätten vorhanden sein müssen, damit ein Ereignis nicht eingetreten wäre.<sup>934</sup> Kontrafaktische Gedanken rekonstruieren die Vergangenheit und werden meist durch Sätze, die mit „wenn (ich) doch nur (nicht)“ beginnen, ausgedrückt. Das Nachdenken darüber „was hätte sein können“ beeinflusst, wie wir über eine zukünftige Situation denken. Es schärft das Bewusstsein dafür, dass man mehr als eine Möglichkeit in Betracht ziehen sollte, ehe man eine Entscheidung fällt.<sup>935</sup> Kontrafaktisches Denken erleichtert es, sich vorzustellen, welche Gründe *gegen* die derzeit favorisierte Hypothese sprechen.<sup>936</sup> Menschen erzählen gerne und gut Geschichten; es fällt uns leicht, im Nachhinein eine Erklärung für eine Beobachtung zu finden, die plausibel und überzeugend klingt.<sup>937</sup> Von der Plausibilität der Erklärung zur Überzeugung, dass die Erklärung richtig ist, ist es nur ein kleiner Schritt. Aktiv und bewusst Gründe zu suchen, die *gegen* die Erklärung/Hypothese sprechen, schärft den Blick für Informationen, die der ersten Erklärung widersprechen und gerne unter den Tisch gewischt werden. Es trägt dazu bei, die vorhandenen Informationen ausgewogener zu gewichten und wirkt daher einer der Ursachen von Bestätigungstendenzen entgegen.

### B. Alternative Erklärungen generieren und testen

591 Alles deutet darauf hin, dass die Beachtung einer alternativen Hypothese den Bestätigungsfehler stark verringert oder gar verschwinden lässt.<sup>938</sup> Erfolgreiche Versuchspersonen zeichnen sich dadurch aus, dass sie mehrere Hypothesen generieren und testen.<sup>939</sup> Die Bevorzugung hypothesenbestätigender Informationen verschwindet, wenn die Versuchspersonen auf eine alternative Hypothese hingewiesen werden.<sup>940</sup> Man sollte daher immer

---

<sup>933</sup> LAURA J. KRAY/ADAM D. GALINSKY, The Debiasing Effect of Counterfactual Mind-Sets: Increasing the Search for Disconfirmatory Information in Group Decisions, *Organizational Behavior and Human Decision Processes* 2003, 69-81.

<sup>934</sup> DANIEL KAHNEMAN/AMOS TVERSKY, The Simulation Heuristic, in: KAHNEMAN/SLOVIC/TVERSKY (Hrsg.), 201-208, 203.

<sup>935</sup> KRAY/GALINSKY, FN 933, 70.

<sup>936</sup> KORIAT/LICHTENSTEIN/FISCHHOFF, FN 925, 109.

<sup>937</sup> KOEHLER, FN 902, 500 ff.

<sup>938</sup> GADENNE/OSWALD, FN 904, 372; KLAYMAN, FN 855, 405; MYNATT/DOHERTY/DRAGAN, FN 859, 775; EVANS/VENN/FEENEY, FN 871, 42

<sup>939</sup> KLAYMAN, FN 855, 405.

<sup>940</sup> GADENNE/OSWALD, FN 904, 372.

versuchen, sich alternative Erklärungen für die beobachteten Fakten vorzustellen und zu testen. Auf keinen Fall sollte man sich frühzeitig gegenüber Dritten oder gar der Öffentlichkeit auf eine Hypothese festlegen; je grösser das *commitment* gegenüber einer Hypothese, desto unangenehmer ist es, die Hypothese wieder aufgeben zu müssen, und desto grösser die unbewusste Tendenz, alles zu tun, um dies zu vermeiden. Alternative Hypothesen können auch durch einen *advocatus diaboli* generiert werden, da es manchmal schwierig ist, sich alternative Hypothesen vorzustellen, wenn man sich bereits auf eine Erklärung eingeschossen hat.