

*Vorlesung (HS-2013):
Urteilsfehler
(Teil 1)*

Autor: Siegfried Macho

Datum der letzten Änderung: Mittwoch, 23. Oktober 2013

Inhaltsverzeichnis

1. Grundlagen	1
1.1 Urteilsfehler und Rationalität	1
1.1.1 Epistemische vs. instrumentelle Rationalität	4
1.1.2 Kritikpunkte an der Forschung über Urteilsfehler und am Rationalitätskonzept	5
1.2 Psychologische Mechanismen: »Hot« und »cold cognition«	7
2. Kontingenz- und Kausalurteile	9
2.1 Die klassische Untersuchung von Chapman und Chapman zur illusorischen Korrelation	9
2.2 Subjektive Theorien und fehlerhafte Kontingenz- und Kausalurteile	11
2.3 Wahrnehmung von Muster in Zufallssequenzen	15
2.3.1 Die »heisse Hand« im Basketball	17
2.3.2 »Heisse Hand« und Gambler's Fallacy	19
2.3.3 Sind die »heisse« Hand und die Gambler's Fallacy rational?	20
2.4 Ignorierung des Phänomens der Regression zum Mittel	21
2.4.1 Das Prinzip der Regression zum Mittel	21
2.4.2 Die durch das Ignorieren der Regression zum Mittel bedingten Urteilsfehler	22
2.4.2.1 Der Fluglehrer	22
2.4.2.2 Erklärungen in Sport und Beruf	22
2.4.2.3 Fairness gegenüber benachteiligten Gruppen	23
2.5 Fehlerhafte Gewichtung von relevanter Information	25
2.5.1 Beurteilung von Information in Kontingenztabelle	25
2.5.2 Aufmerksamkeits- und Salienzeffekte	28
2.5.2.1 Das ANOVA Modell von Kelly	28
2.5.2.2 Nichtbeachtung subtiler Einflussgrößen auf das eigene Verhalten	31
2.5.2.3 Nichtbeachtung von Konsensusinformation	33
2.5.2.4 Überbewertung irrelevanter Einflussgrößen auf das eigene Verhalten	34
2.5.2.5 Der fundamental Attributionsfehler	34
2.5.2.6 Asymmetrie von Attributionen	36
2.5.2.7 Probleme von Erklärung mittels Salienz- und Aufmerksamkeitseffekten	37

2.6 Methodische Aspekte: Klassische Fehler und Paradoxien in der Beurteilung von Kontingenz und Kausalität	38
2.6.1 Die Bedeutung der Kenntnis von (Un-) Abhängigkeiten zwischen Variablen	38
2.6.2 Das Problem der Erschliessung kausaler Beziehungen aufgrund von Abhängigkeiten zwischen Variablen	39
2.6.2.1 Elimination von Konfundierung durch Randomisierung und Parallelisierung	43
2.6.2.2 Die Bedeutung von Assoziationen für die Prüfung von Kausalstrukturen	45
2.6.3 Das Simpson Paradoxon	47
2.6.4 Der ökologische Fehlschluss	52
2.6.5 Regression zum Mittel und Lords Paradoxon	54
2.7 Zusammenfassung	56
2.8 Übungen	57
3. Gedächtnisurteile	62
3.1 Klassische Studien zu Gedächtnisfehlern	62
3.1.1 Fehlinformationen und Gedächtnisfehler	63
3.1.2 Assoziative Gedächtnisillusionen: Das DRM-Paradigma	66
3.2 Beurteilung von Stabilität und Veränderung	69
3.3 Der Rückschaufehler (Hindsight Bias)	72
3.4 Retrospektive Evaluation affektiver Episoden	73
3.5 Zusammenfassung	75
4. Wahrscheinlichkeitsurteile und statistische Schlussfolgerungen	77
4.1 Heuristics and Biases Ansatz von Tversky and Kahnemann	77
4.1.1 Die Wirkungsweise der Verfügbarkeitsheuristik	79
4.1.1.1 Imagination und Verfügbarkeit	81
4.1.1.2 Lebhaftigkeit, persönliches Erleben und Verfügbarkeit	83
4.1.2 Die Wirkungsweise der Repräsentativitätsheuristik	85
4.2 Fehlerhafter Umgang mit Wahrscheinlichkeiten	88
4.2.1 Subjektiver Gewichtung von Wahrscheinlichkeiten	88
4.2.2 Probability Matching	90
4.2.2.1 Erklärung von Probability Matching	93
4.2.2.2 Individuelle Unterschiede beim Probability Matching	93
4.2.3 Umgang mit bedingten Wahrscheinlichkeiten	94

4.2.3.1	Asymmetrie bedingter Wahrscheinlichkeiten	95
4.2.3.2	Der Einfluss von Kausalen und diagnostischen Schlüssen auf die Beurteilung bedingter Wahrscheinlichkeiten	97
4.2.3.3	Das Problem nonmonotoner Wahrscheinlichkeiten	100
4.2.4	Unterbewertung und Ignorierung von Basisraten	101
4.3	Probabilistische Schlussfolgerungen	103
4.3.1	Das Bayes-Theorem in Wettquotienten-Format	108
4.3.2	Exkurs: Das Ziegenproblem	112
4.4	Die Kritik am »Heuristics and Biases« Programm	114
4.4.1	Normativität von Axiomen und Regeln der Wahrscheinlichkeitsrechnung	114
4.4.2	Anwendbarkeit der Axiome und Regeln der Wahrscheinlichkeitsrechnung	115
4.4.3	Fehlerhaft angewendete normative Modelle	118
4.4.4	Verbesserung von Wahrscheinlichkeitsurteilen	123
4.4.4.1	Verbesserung des Wahrscheinlichkeitsurteils durch Erhöhung der Salienz des Zufallscharakters	124
4.4.4.2	Verbesserung des Wahrscheinlichkeitsurteils durch Verwendung von Häufigkeiten	124
4.4.4.3	Natürliche Häufigkeiten	126
4.4.4.4	Erhöhung der Verständlichkeit der Problemstruktur	128
4.5	Übungen	130
5.	Literatur	132
6.	Bücher zum Thema	139
7.	Anhang: Paik-Diagramme zu den Beispielen zum Paradoxon von Simpson	142

1. Grundlagen

In diesem einleitenden Kapitel werden einige Begriffe und Prinzipien diskutiert, welche für die nachfolgende Behandlung der Thematik von Bedeutung sind.

1.1 Urteilsfehler und Rationalität

Wir betrachten im Folgenden menschliche Urteilsfehler als Formen von irrationalem Verhalten. Hier stellen sich nun sofort einige Fragen:



Fragen:

1. *Wie kann man beurteilen ob ein Urteilsfehler vorliegt?*
2. *Was bedeutet rationales bzw. irrationales Verhalten?*

Zur Beantwortung der ersten Frage betrachten wir folgenden Grundsatz:



Prinzip 1-1: *Identifikation von Urteilsfehlern:*

Ein Urteil kann als fehlerhaft identifiziert werden, falls die folgenden Bedingungen erfüllt sind:

1. *Es existiert ein normatives Prinzip.*
2. *Dieses normative Prinzip ist auf die aktuelle Situation anwendbar.*
3. *Die Person verstösst gegen das normative Prinzip.*

Bevor wir ein konkretes Beispiel diskutieren, soll noch konkretisiert werden, dass mit *rationalem Verhalten* gemeint ist.



Prinzip 1-2: *Rationales Verhalten:*

Ein Verhalten ist rational, wenn es den erwarteten persönlichen Nutzen (expected subjective utility) maximiert.

Oder anders ausgedrückt:

Ein Verhalten ist rational, wenn es in optimaler Weise zur Erreichung der persönlichen Ziele beiträgt.

Betrachten wir nun ein einfaches Beispiel, welches das Gesagte illustriert.



Bsp. 1-1: *Transitivität von Präferenzen:*

Ein wichtiges Prinzip die Rationalität persönlicher Präferenzen betreffend besagt, dass Präferenzen transitiv sein müssen. Konkret besagt dieses Prinzip:

Falls eine Person Alternative A gegenüber Alternative B und B gegenüber Alternative C bevorzugt, so muss sie auch A gegenüber C bevorzugen.

In Symbolen:

$$A \succ B \wedge B \succ C \rightarrow A \succ C$$

hierbei gilt:

$A \succ B$ bedeutet, dass Alternative A gegenüber B bevorzugt wird ;

Das Symbol \wedge repräsentiert das logische UND (der Satz $X \wedge Y$ ist genau dann wahr, wenn sowohl X als auch Y wahr sind).

Das Symbol \rightarrow repräsentiert die logische Implikation (der Satz $X \rightarrow Y$ ist genau dann wahr, wenn entweder X falsch oder Y wahr ist).

Es ist sehr einfach zu zeigen, dass ein Person, welche dieses Prinzip verletzt zu einer so genannten »Geldpumpe« (»money pump«) gemacht werden kann, d.h. es kann eine Serie von Transaktionen konstruiert werden, sodass die Person ständig an Geld verliert.

Um dies zu demonstrieren betrachten wir folgende konkrete Situation.

Gegeben seien drei Statistiklehrbücher: »Der Bortz« (B), »der Eid« (E) und »der Hayes« (H).

Person P habe die folgende intransitive Präferenzordnung:

1. $B \succ E$
2. $E \succ H$
3. $H \succ B$

Nehmen wir nun an, ich besitze eines der drei Bücher (sagen wir B) und besagte Person besitze die anderen beiden (E und H).

Da P H gegenüber B bevorzugt, wird sie bereit sein, einen kleinen Betrag (sagen wir Sfr. 1.- [die genaue Grösse des Betrages ist unwichtig]) zu geben, um B gegen H einzutauschen. Ebenso wird sie bereit sein, einen kleinen Betrag zu geben (sagen wir wiederum Sfr. 1.-), um H gegen E einzutauschen (da sie E gegenüber H bevorzugt).

Schliesslich wird bereit sein, einen kleinen Betrag zu geben (sagen wir wiederum Sfr. 1.-), um E gegen B einzutauschen (da sie B gegenüber E bevorzugt).

Damit sind wir am Ausgangspunkt angelangt (nur ist P um Sfr. 3.- ärmer) und das Spiel kann von vorne beginnen.

Wir sehen also: Indem Person P das Prinzip der Transitivität von Präferenzen verletzt, schadet sie sich selbst (maximiert daher nicht ihren persönlichen Nutzen) und handelt daher irrational.



Bemerkung zu den grundlegenden Axiomen der Wahrscheinlichkeitstheorie und das »Dutchbook« - Argument

Die elementaren Axiome der Wahrscheinlichkeitstheorie (siehe Anhang) können mit Hilfe analoge Argumente begründet werden, indem gezeigt wird, dass bei Nichtbefolgung eines Axioms ein System von Wetten – ein so genanntes »Dutchbook« – konstruiert werden kann, dass in jedem Fall (d.h. gleichgültig, welches der möglichen Zufallsereignisse eintritt) zu einem Verlust führt. Hierbei ist jede Wette für sich genommen fair und somit akzeptable (eine einfache Darstellung findet man bei Hacking, 2001).

Die Formulierung in Prinzip 1-2 enthält eine »Subtilität«, die es zu beachten gilt. Es wird nämlich nicht einfach davon gesprochen, dass der persönliche Nutzen zu maximieren ist, sondern dass der *erwartete* persönliche Nutzen maximiert werden soll. Um dies zu verstehen, definieren wir zuerst, was unter erwartetem persönlichem Nutzen zu verstehen ist.



Konzept 1-1: Erwarteter persönlicher Nutzen:

Gegeben:

- Konsequenzen K_1, K_2, \dots, K_m , die sich aus der Wahl einer Alternativen ergeben.
- Persönlicher Nutzen u_1, u_2, \dots, u_m der einzelnen Konsequenzen. Der Nutzen einer Konsequenz kann positiv oder negativ sein
- Wahrscheinlichkeiten p_1, p_2, \dots, p_m der einzelnen Konsequenzen.

Der erwartete persönliche Nutzen $E_U(X)$ einer Alternative X entspricht der Summe der einzelnen Nutzenkomponenten multipliziert mit der Wahrscheinlichkeit:

$$E_U(X) = u_1 \cdot p_1 + u_2 \cdot p_2 + \dots + u_m \cdot p_m$$

Der erwartete (oder durchschnittliche) Nutzen wird als zu maximierende Zielgrösse gewählt, da der sich ergebende Nutzen vom Zufall abhängt. Es kann daher vorkommen, dass eine Person eine schlechte Wahl trifft (gemäss dem zu erwartenden Nutzen), jedoch Glück hat, während eine andere Person eine bessere Wahl trifft aber aufgrund von Pech ein schlechteres Ergebnis gewärtigen muss. Um diese Abhängigkeit vom Zufall (der nicht durch die Person zu kontrollieren ist) zu umgehen, wird der erwartete Nutzen betrachtet.

1.1.1 Epistemische vs. instrumentelle Rationalität

Aufgrund des oben Gesagten kann das Vorliegen von Urteilsfehlern mit Hilfe zweier Kriterien bestimmt werden: (i) Das Urteil widerspricht normativen Prinzipien, oder (ii) das Urteil verhindert eine Optimierung des erwarteten Nutzens. In der Literatur wurden für diese beiden Kriterien unterschiedliche Bezeichnungen verwendet (Stanovich, 2010).



Konzept 1-2: Epistemische vs. instrumentelle Rationalität :

Epistemische (theoretische) Rationalität liegt vor, falls ein Urteil oder eine Entscheidung den als normativ richtig anerkannten Prinzipien gehorcht bzw. nicht widerspricht.

Instrumentelle (praktische) Rationalität eines Urteils oder einer Entscheidung ist gegeben, falls das Urteil/die Entscheidung den erwarteten Nutzen optimiert.

Man kann sich nun die folgende Frage vorlegen:



Frage:

Führen epistemische und instrumentelle Rationalität immer zu identischen Urteilen oder Entscheidungen?

Die Antwort auf diese Frage ist nicht ganz einfach und in der so genannten Rationalitätsdebatte innerhalb der Psychologie gibt es dazu unterschiedliche Meinungen (vgl. Stainton, 2006, Stanovich, 2010). Das folgende Beispiel legt jedoch nahe, dass beide Arten von Rationalität durchaus zu unterschiedlichen Ergebnissen führen können.



Bsp. 1-2: Wohlbefinden und Glaube an ein Jenseits

Der Glaube an ein Jenseits widerspricht einem grundlegenden Rationalitätsprinzip, das unter dem Namen *Occam's Razor* bekannt ist [nach dem englischen Mönch und scholastischem Philosophen William von Ockham (vermutlich 1288-1347)].

Dieses Prinzip besagt, dass Entitäten nicht über das Notwendige hinaus vermehrt werden dürfen (»Entia non sunt multiplicanda praeter (sine) necessitatem«).

Gemäss diesem Prinzip ist ein Glaube an ein Jenseits unsinnig, da hierbei etwas angenommen wird, wofür es keine anerkannten Hinweise gibt. Der Glaube an ein Jenseits ist daher vom Standpunkt der epistemischen Rationalität untragbar.

Auf der anderen Seite kann ein Glauben an ein Jenseits durchaus den persönlichen Nutzen maximieren, denn zahlreiche Studien aus dem Bereich der positiven Psychologie zeigen, dass Personen, die an ein Jenseits glauben, sich oft eines besseren psychischen Wohlbefindens erfreuen und meist besser mit Schicksalsschlägen umgehen können.

Daher ist der Glaube an ein Jenseits vom Standpunkt der instrumentellen Rationalität nicht unbedingt ungerechtfertigt (Wenn man annimmt, dass der Gewinn an Wohlbefinden die eventuell mit dem Jenseitsglauben verbundenen Kosten – Sünde, die Ausführung bestimmter Rituale – ausgleicht).

Man beachte, dass hier nur eine punktuelle Abweichung von der Norm vorliegt. Ein Individuum, welches das normative Prinzip in lebensrelevanten Situationen ignoriert, kann sein Leben gefährden.



Abschliessende Bemerkung:

Die gegebene Unterscheidung zwischen epistemischer und instrumenteller Rationalität umfasst im Wesentlichen alle wichtigen Formen von Rationalität.

In der Literatur finden sich noch andere Formen von Rationalität. So sprechen Kahneman und Frederick (2002) von *Kohärenzrationalität*, womit sie meinen, dass die Glaubensinhalte einer Person kohärent sein müssen. Hier handelt es sich eindeutig um einen Spezialfall von epistemischer Rationalität.

1.1.2 Kritikpunkte an der Forschung über Urteilsfehler und am Rationalitätskonzept

Die traditionelle Forschung zu Urteilsfehler ist in den letzten Jahren zunehmend kritisiert worden. Die Kritik geht dabei in drei verschiedene Richtungen:

1. *Kritik der normativen Prinzipien:* Diese Art von Kritik zielt auf die Korrektheit der angenommenen normativ korrekten Prinzipien, welche als Standard zur Evaluation von Urteilen verwendet werden. So argumentiert Cohen (1981), dass es sich bei normativen Prinzipien um Konventionen und damit um subjektive Setzungen handle.
2. *Problem der Anwendbarkeit bestimmter normativer Prinzipien:* Diese Art von Kritik weist darauf hin, dass in vielen der aktuellen Untersuchungen entweder die normativen Prinzipien nicht anwendbar seien oder falsch angewendet würden (siehe z.B. Birnbaum, 1983; Gigerenzer, 1996; Gigerenzer & Murray, 1987).
3. *Problem der ökologischen Repräsentativität:* Diese Kritik zielt darauf ab, dass die in Experimenten gefundenen Urteilsfehler ein Artefakt darstellen, welches dadurch zustande kommt, dass die gestellten Aufgaben die Alltagsrealität der untersuchten Personen nicht adäquat widerspiegeln (Siehe z.B. Gigerenzer & Hoffrage, 1995; Juslin, 2001).

Wir werden auf diese Kritikpunkte zur gegebenen Zeit eingehen. Im Speziellen werden wir die ersten beiden Kritikpunkte im letzten Kapitel **xxxx** genauer behandeln. An dieser Stelle sei jedoch ein

kleines Beispiel präsentiert, welche das Problem der Anwendbarkeit normativer Prinzipien illustriert,



Bsp. 1-3: Erwarteter Nutzen bei Einzelereignissen:

Jemand bietet Ihnen die folgende Wette an:

Es wird ein Würfel geworfen. Falls dieser die Augenzahl 5 oder 6 zeigt, so erhalten Sie Sfr. 6000.-, andernfalls zahlen Sie Sfr. 600.-

Viele Personen würden eine derartige Wette ablehnen, obwohl der erwartete Nutzen (Gewinn) positiv ist:

$$E_U = \frac{1}{3} \cdot 6000 - \frac{2}{3} \cdot 600 = 1600$$

Der Grund für die Ablehnung der Wette ist offensichtlich: Für viele Personen ist der Verlust von Sfr. 600.- relativ schmerzlich und ein Verlust ist wahrscheinlicher als ein Gewinn.

Nun lässt sich aber einwenden, dass Wahrscheinlichkeiten für Einzelereignisse nicht definiert sind (zumindest gemäss der frequentistischen Theorie der Wahrscheinlichkeit). Damit ist auch obige Berechnung des zu erwartenden Nutzens wenig sinnvoll.

Dem korrespondiert die Erfahrung, dass viele Personen (vermutlich die meisten) folgende modifizierte Wette akzeptieren:

Es wird ein Würfel geworfen. Falls dieser die Augenzahl 5 oder 6 zeigt, so erhalten Sie Sfr. 600.-, andernfalls zahlen Sie Sfr. 60.-.

Dieses Spiel wird 10× wiederholt und die Gewinne und Verluste aus den einzelnen Durchgängen werden addiert.

Diese Wette hat exakten den gleichen erwarteten Nutzen wie die erste. Dennoch erscheint sie den meisten Personen als viel attraktiver.

Man beachte, dass bei der modifizierten Wette bereits ein Gewinn von Sfr. 60.- gegeben ist, falls in nur einem der 10 Durchgänge die Augenzahl 5 oder 6 erscheint. Die Wahrscheinlichkeit hierfür beträgt:

$$P(\text{Mindestens einmal eine 5 oder 6 in 10 Spielen}) = 1 - \left(\frac{2}{3}\right)^{10} = 0.983$$

Es gibt noch einen zweiten, verwandten Gesichtspunkt: Die zweite Wette weist eine um den Faktor 10 geringere Varianz auf als die erste Wette: 968000 gegenüber 9680000.

Zusammenfassend könnte man daher wie folgt argumentieren:

Das scheinbar irrationale Verhalten einer Person, welche die obige Wette ablehnt, entpuppt sich bei genauerer Analyse als nicht irrational. Vielmehr handelt es sich um fehlerhafte Anwendung des Konzepts der Wahrscheinlichkeit.

oder:

Personen sehen nicht nur auf den erwarteten Nutzen sondern ihre Wahl wird auch von der Varianz des Nutzens beeinflusst. Es macht wenig Sinn, Personen, welche auch die Varianz einbeziehen, als irrational zu bezeichnen.

Interessanterweise wird dieser letztere Aspekt, wonach auch die Varianz in die Überlegungen einbezogen werden sollte, in der Literatur kaum diskutiert (Die einzige mir bekannte Ausnahme ist Rode, Cosmides, Hell, & Tooby, 1999).

1.2 Psychologische Mechanismen: »Hot« und »cold cognition«

In der Sozialpsychologie gibt es die Unterscheidung zwischen »hot« und »cold cognition«. Der Term »hot cognition« wurde von Abelson (1963) eingeführt.



Konzept 1-3: Hot vs. cold cognition:

Unter *hot cognition* versteht man kognitive Prozesse (z.B. Urteils- und Entscheidungsprozesse), die durch Motivation, Emotion, Stimmungen oder andere Formen von Erregungszuständen (z.B. Stress) beeinflusst sind.

Im Falle von *cold cognition* liegen keine derartigen Einflussfaktoren vor.

Die Unterscheidung zwischen hot und cold cognition muss als eine graduelle betrachtet werden, da Urteile immer eine motivationale Basis haben auch wenn es sich »nur« um die Motivation handelt, ein möglichst gutes Urteil zu fällen.

Wenn man von Beeinflussung des Urteils durch die Motivation einer Person spricht, so meint man meistens, dass eine *gerichtete Motivation* vorliegt, welche dazu führt, dass das Urteil in eine bestimmte Richtung gelenkt wird (Kunda, 1990), um bestimmte Werte und Meinungen aufrecht erhalten zu können, deren Aufgabe als eine Bedrohung des Selbst erlebt wird. Typische Beispiele von derartigen gerichteten Verzerrungen des Urteils sind Fehlattritionen zur Vermeidung eines negativen Selbstbildes (»Nicht ich war unfähig, sondern die Aufgabe war zu schwierig«), die Abwertung, das Ignorieren und die fehlerhafte Interpretation von Tatsachen, welche bestimmten Vorstellungen widersprechen (Leugnen des Holocaust, Leugnen der biologischen Evolution), oder Fehlinterpretationen zur Vermeidung von kognitiver Dissonanz.

Im Folgenden werden wir uns vorwiegend – aber nicht ausschliesslich – mit Urteilsfehlern befassen, welche nicht durch gerichtete Motivationen bedingt sind oder unter anderen verzerrenden Bedingungen (wie Stress) erfolgen. Der Grund hierfür liegt darin, dass wir uns vorwiegend mit den grundlegenden kognitiven Beschränkungen, welche menschlichen Urteilen zugrunde liegen, beschäftigen wollen. Einen Überblick über gerichtete Urteile aufgrund motivationaler Ursachen findet man in Kunda (1990). Janis und Mann (1979) präsentieren Untersuchungen zu Entscheidungen unter Stress und anderen verzerrenden Einflüssen.



Bemerkung

Auch bei Vorliegen motivationaler Ursachen, die zur Urteilsverzerrung beitragen, haben diese oft nur einen geringen Einfluss auf das Urteil. Der Grund liegt darin, dass die meisten Personen versuchen, in ihren Urteilen möglichst rational zu sein und daher offenkundige Urteilsfehler vermeiden (Kunda, 1990).

In den nächsten Kapiteln werden menschliche Urteilsfehler bei verschiedenen Arten von Urteilen untersucht. Hierbei werden zuerst empirische Ergebnisse zu den relevanten Phänomenen präsentiert. Dem folgt eine Besprechung der psychologischen Mechanismen, welche die Urteilsfehler verursachen. Schliesslich werden Methoden und Konzepte bereitgestellt, die es ermöglichen, die Situation zu analysieren und die korrekten Antworten zu finden.

2. Kontingenz- und Kausalurteile

In der Geschichte der Menschheit finden sich zahlreiche Beispiele von fehlerhaften Urteilen über Zusammenhänge (=Kontingenzurteile) und von Urteilen über fehlerhafte Ursache-Wirkungsbeziehungen (=Kausalurteile). Zugleich findet man aber auch Beispiele, wo wichtige Ursache-Wirkungszusammenhänge nicht erkannt wurden.

Prägnante Beispiele fehlerhafte Kontingenzurteile bilden die Astrologie, wonach es einen Zusammenhang zwischen der Planetenkonstellation bei der Geburt eines Menschen und dessen Persönlichkeitsmerkmalen gibt, oder verschiedener (anderer) Arten von diagnostischen Verfahren wie Graphologie oder Rorschachtest (um nur die »seriösesten« zu nennen. Weniger seriös sind z.B. Tests, welche aufgrund der Schlafposition, der Wohnungseinrichtung, den Essgewohnheiten etc. Schlüsse auf die Persönlichkeit oder Psychopathologie von Personen ziehen). Zu nennen sind hier auch verschiedene Arten von Typologien wie jene von Ernst Kretschmer (1888-1964), der einen Zusammenhang zwischen Körperbau und Charakter postulierte.

Was Beispiele von fehlerhaften Kausalurteilen betrifft, so ist die Anzahl Legion. Vor allem im Bereich der Psychologie und der Medizin findet man gehäuft Beispiele von angeblichen Wirkungen auf das Verhalten (z.B. Wirkung des Vollmondes) oder auf die Gesundheit (z.B. Wirkung von Handystrahlen, Homöopathie, Akupunktur und anderer angeblicher Heilmittel). Interessant ist in diesem Zusammenhang auch, dass sich manche angebliche Heilmittel in der Realität als schädlich herausgestellt haben (z.B. der Aderlass [Ernst & Singh, 2009]).

Was das Übersehen von wichtigen Kausalzusammenhängen betrifft, so wäre die Unterschätzung der Bedeutung von Hygiene in Krankenhäusern für die Vermeidung von Todesfällen zu erwähnen. Den Vorreiterinnen wie Florence Nightingale (1820-1910) und Ignaz Semmelweis (1818-1865) wurde zuerst mit grosser Skepsis begegnet.

Die folgende Untersuchung legt einen zentralen Mechanismus, der für viele der oben genannten Urteilsfehler verantwortlich ist, bloss.

2.1 Die klassische Untersuchung von Chapman und Chapman zur illusorischen Korrelation

Im Jahre 1969 führten Chapman und Chapman eine klassische Studie zur so genannten *Illusorischen Korrelation* durch. Die Untersuchung verwendete als Basis den Rorschach-Test.

Beim Rorschachtest handelt es sich um 10 Karten mit Tintenflecken. Der zu untersuchenden Person wird jeweils eine Karte vorgelegt mit der Aufforderung irgendwelche Gestalten und Formen in den Flecken zu »erkennen«.

Es hat sich gezeigt, dass homosexuelle häufiger als heterosexuelle Männer Monster auf Karte Nr.4 »sehen«. Daher kann die Wahrneh-

mung von Monster auf Karte 4 als diagnostisch valider Hinweis für Homosexualität betrachtet werden (es gibt noch weitere valide Hinweise für Homosexualität).

Neben diesen diagnostisch validen Hinweisen gibt es eine Menge möglicher *plausibler* Hinweise, wie z.B. wahrgenommene Personen mit weiblicher Kleidung, Personen mit unklarem Geschlecht, die Wahrnehmung männlicher Genitalien oder anale Formen. Letztere Hinweise können als plausibel bezeichnet werden, weil »der gesunde Hausverstand« nahe legt, dass Homosexuelle vermehrt derartige Formen in die Kleckse hineinprojizieren.

In Wirklichkeit gibt es jedoch keinen statistischen Zusammenhang zwischen plausiblen Hinweisen und Homosexualität (sie sind daher empirisch nicht valide). Die Chapmans führten nun eine Untersuchung, durch, die aus mehreren Stufen bestand:

Phase 1:

32 praktizierenden Kliniker, die angegeben hatten, schon eine Reihe von Protokollen homosexueller Männer analysiert zu haben, erhielten eine Liste von Hinweisen (=Antworten von Personen), die sowohl diagnostisch valide, als auch plausible (invalide) Hinweise bezüglich des Vorliegens von Homosexualität beinhaltete.

Die 5 häufigsten Antworten, die von den Klinikern als relevant betrachtet wurden, waren plausible, aber empirisch nicht valide Hinweise. Nur 2 Kliniker erwähnten überhaupt die empirisch validen Hinweise.

Phase 2:

Naive Vpn wurden gebeten, den Zusammenhang zwischen Homosexualität und den verschiedenen Hinweisen abzuschätzen. Wenig überraschend zeigte sich, dass der Zusammenhang zwischen plausiblen Hinweisen und Homosexualität als »relativ stark«, der zwischen den diagnostisch validen Hinweisen und Homosexualität hingegen als »sehr schwach« eingeschätzt wurde.

Phase 3:

Naiven Vpn wurden 30 Rorschach-Karten zusammen mit homosexuellen »Symptomen« (z.B. »fühlt sich sexuell durch andere Männer angezogen«) oder neutralen Symptomen (z.B. »fühlt sich oft einsam und traurig«) präsentiert. Die Rorschach-Kleckse befanden sich auf der einen Seite der Karten, die Symptome und die Antworten der (angeblichen) Testpersonen auf der anderen Seite.

Die Antworten beinhalteten valide oder invalide aber plausible, sowie neutrale Hinweise. Die Vpn sollten *lernen*, welche Antworten diagnostisch für homosexuelle Symptome sind. Die Serie war derart konstruiert, dass es keinen objektiven Zusammenhang zwischen den Antworten und den homosexuellen Symptome gab.

Den Vpn wurde immer eine Karte gezeigt, die sie 60 Sekunden lang studieren konnten. Nach Präsentation der 30 Karten mussten sie angeben, welche Antwort mit Homosexualität gepaart war. Die Vpn

sahen – ähnlich wie die »erfahrenen« Kliniker – einen Zusammenhang zwischen den plausiblen Antworten und den homosexuellen Symptomen (obwohl es objektiv keinen derartigen Zusammenhang gab).

Phase 4:

Der Zusammenhang zwischen den validen Antworten und den homosexuellen Symptomen wurde variiert: Der Zusammenhang war entweder 50%, 67%, 83%, oder 100%.

Die Änderung des Zusammenhanges bewirkte keine Veränderung im Urteil der naiven Vpn: Sie erkannten nicht den Zusammenhang zwischen den validen Antworten und den homosexuellen Symptomen.

Nachdem die plausiblen Antworten von den Karten entfernt worden waren, erlernten die Vpn jedoch die wahren Zusammenhänge.

Fasst man die zentralen Ergebnisse der Untersuchung zusammen, so ergibt sich folgendes Bild:

1. Sowohl die »erfahrenen« Kliniker, wie die naiven Vpn »erkannten« aufgrund von vor gefassten Meinungen nicht existierende Zusammenhänge (*Illusorische Korrelation*).
2. Dieselbe vor gefasste Einstellung verhinderte, dass existierende, aber unplausible Zusammenhänge erkannt wurden.

2.2 Subjektive Theorien und fehlerhafte Kontingenz- und Kausalurteile

Die Untersuchung von Chapman und Chapman (1969) weist auf einen zentralen Mechanismus für fehlerhafte Kontingenz- und Kausalurteile hin:



Kognitiver Mechanismus 2-1: Die Bedeutung subjektive Theorien für die Wahrnehmung von Zusammenhängen

Fehlerhafte subjektive (plausible) Theorien über (kausale) Zusammenhänge können folgende negative Wirkung auf das Urteil haben:

- (i) »Wahrnehmung« nicht existierender Zusammenhänge
- (ii) Übersehen existierender Zusammenhänge.

Falsche subjektive Theorien sind – wie sich in den folgenden Abschnitten herausstellen wird – nicht nur für fehlerhafte Kontingenz- und Kausalurteile verantwortlich, sondern auch für andere Verzerrungen, wie z.B. die einseitige Verarbeitung von Evidenz (vgl. Kapitel **xxxx**).

Hier stellt sich nun eine zentrale Frage:



Frage:

Wie kommen Personen zu falschen subjektiven Theorien?

Die meisten Theorien über Zusammenhänge werden kulturell vermittelt. Daher handelt es sich hier um eine der Hauptursachen für die

Verbreitung falscher Theorien. Im Speziellen fungieren oftmals so genannte »Experten« als Quellen für falsche Theorien. Ein gutes Beispiel hierfür bietet der zweifache Nobelpreisträger (für Chemie und Frieden) Linus Pauling (1901-1994). Basierend auf fehlerhaften Studien behauptete er, dass vermehrter Vitamin C Konsum vor Krebs schützt. Eine Behauptung, die sich durch eingehende Studien der Mayo Klinik nicht belegen liess. Viele Beispiele für fehlerhafte Theorien im Bereich der Psychologie (*Wir nutzen nur 10% unseres Gehirns, Traumatische Ereignisse werden verdrängt, Lernen im Schlaf ist möglich, Ärger soll man herauslassen, etc.*) werden in Lilienfeld, Lynn, Rusco und Beyerstein (2010) präsentiert und entlarvt.

Eine zweite wichtige Quelle von fehlerhaften subjektiven Theorien ist der so genannte »gesunde Menschenverstand«. Dieser basiert grösstenteils auf augenscheinlichen Fakten und Zusammenhängen (»Gemäss dem gesunden Menschenverstand ist die Erde eine Scheibe« soll Einstein gesagt haben). Für den gesunden Menschenverstand scheint die Ähnlichkeit zwischen Ursache und Wirkung ein wichtiges Kriterium zu sein. Gilovich und Savitsky (2002) präsentieren eine Reihe von Beispielen aus verschiedensten Bereichen, welche die Bedeutung der Ähnlichkeit für Kausal- und Kontingenzurteile demonstrieren.



Bsp.2-1: Bedeutung der Ähnlichkeit für Kausal- und Kontingenzurteile

Medizin:

- Homöopathie: Gleiches muss durch Gleiches geheilt werden.
- Bei männlichen Potenzproblemen verwende man zerriebenes Nashorn (das bekanntlich fest und stark ist).
- Lunge des Fuchses (bekannt für seine Ausdauer) gegen Asthma.
- Mikroorganismen sind zu klein um grossen Schaden anzurichten.
- Bei Schäden bestimmter Organe (z.B. der Leber), sollen die entsprechenden Organe in roher Form von gesunden Tieren verspeist werden: z.B. Gehirnteile bei mentalen Erkrankungen, Herzkonzentrate bei Herzerkrankungen, oder rohe Mägen bei Magengeschwüren.

Negative Wirkung von Speisen:

- Fettige Kartoffelchips führen zu fettiger Haut.
- Stark gewürzte Speisen führen zu Herzbrennen.

Astrologie:

- Zwillinge haben 2 Gesichter (ambige Persönlichkeit).
- Steinböcke sind hartnäckig und dickköpfig.

- Löwen sind stolze, energische Führer.
- Krebse haben eine harte Schale und einen weichen Kern.
- Waage Menschen sind ausgeglichen und harmonisch.

Graphologie:

- Zonale Theorie: Oberer Teil der Handschrift enthüllt intellektuelle, mittlerer Teil praktische und unterer Teil instinktive Eigenschaften des Schreibers.
- Distanz der Unterschrift zum Text gibt Auskunft darüber, wie sehr sich die Person vom Geschriebenen distanziert.
- Bleibt Handschrift am linken Rand, so haftet die Person stark an der Vergangenheit. Wird der rechte Rand beschrieben so weist dies auf eine zukunftsorientierte Person.

Ein wichtiger Aspekt von Alltagstheorien, die zum Grossteil auf dem gesunden Menschenverstand basieren, besteht in ihrer Vagheit. Dies zeigt sich in exemplarischer Weise im so genannten Barnum Effekt.



Konzept 2-1: Barnum Effekt (Forer Effekt):

Der *Barnum Effekt* – nach dem Amerikanischen Entertainer Phineas Taylor Barnum (1810-1891) – beschreibt die Tendenz von Individuen oberflächliche und vage Beschreibungen ihrer eigenen Person als höchst zutreffend zu bewerten.

Im Jahre 1948 präsentierte der Psychologe Bertram R. Forer einen Persönlichkeitstest und anschliessend gab er den Teilnehmern eine Persönlichkeitsbeschreibung und liess sie auf einer Skala von 0 bis 5 (sehr zutreffend) einschätzen, wie gut die Beurteilung auf ihre eigene Person zutrifft.

Die Beschreibung lautete wie folgt:

You have a great need for other people to like and admire you. You have a tendency to be critical of yourself. You have a great deal of unused capacity which you have not turned to your advantage. While you have some personality weaknesses, you are generally able to compensate for them. Disciplined and self-controlled outside, you tend to be worrisome and insecure inside. At times you have serious doubts as to whether you have made the right decision or done the right thing. You prefer a certain amount of change and variety and become dissatisfied when hemmed in by restrictions and limitations. You pride yourself as an independent thinker and do not accept others' statements without satisfactory proof. You have found it unwise to be too frank in revealing yourself to others. At times you are extroverted, affable, sociable, while at other times you are introverted, wary, reserved. Some of your aspirations tend to be pretty unrealistic. Security is one of your major goals in life.

Alle Personen erhielten die gleiche Beschreibung. Die durchschnittliche Einschätzung betrug: 4.26.

Verwendete Strategien:

Die verschiedenen Aussagen in der vorangegangenen Passage lassen sich in unterschiedliche Kategorien einordnen, wobei eine Aussage auch gleichzeitig mehreren Kategorien zugeordnet werden kann:

1. Gemeinplätze, die mehr oder weniger auf alle Menschen zutreffen und daher geringen diagnostischen Wert haben:
You have a great need for other people to like and admire you.
Da jeder Mensch in einer gewissen Weise bewundert und geliebt werden will, sagt dies wenig aus.
2. Schmeicheleien:
You have a great deal of unused capacity which you have not turned to your advantage.
Sätze wie diese schmeicheln den Lesern und werden daher gern akzeptiert.
3. Teils-Teils-Behauptungen, die mehrere Interpretationen zulassen:
At times you are extroverted, affable, sociable, while at other times you are introverted, wary, reserved.
4. Behauptungen mit Einschränkungen:
While you have some personality weaknesses, you are generally able to compensate for them.
Die Allgemeinheit einer Behauptung wird durch einen Zusatz eingeschränkt.
5. Gefühl von einem »Experten« durchschaut zu werden: Viele der Aussagen vermitteln dem Leser oder der Leserin das Gefühl durchschaut zu werden.

Aufgrund ihrer Vagheit sind Theorien des gesunden Menschenverstandes daher oft schwer zu widerlegen. Die Vagheit von subjektiven Theorien dürfte auch mit ein Grund sein für eine gewisse »Beliebigkeit« von Alltagserklärungen des menschlichen Verhaltens.



Bsp.2-2: Beliebigkeit von Erklärungen:
(Ross, Lepper, Strack, & Steinmetz 1977):

Die Vpn wurden gebeten, sich in die Lage eines klinischen Psychologen zu versetzen und zu versuchen, ein bestimmtes Verhalten von Patienten mittels Hintergrundinformation zu verstehen und vorherzusagen.

Die Hintergrundinformation bestand aus der detaillierten Lebensgeschichte der Patienten. Die Vpn wurden gebeten, einige kritische Ereignisse im Leben der Person aufgrund seiner Lebensgeschichte zu erklären. (Selbstmord, Beitritt zu einer Sekte, Kandidatur für ein politisches Mandat, etc.).

Ergebnis:

Die Vpn hatten wenig bzw. keinerlei Probleme die Ereignisse zu erklären. Die Erklärungen hatten jedoch – wie Nisbett und Ross meinen – eine erschreckende Beliebigkeit:

*»Eine Analyse der durch die Vpn abgegebenen schriftlichen Erklärungen war instruktiv. Die Leichtigkeit, mit der Vpn von nahezu jedem realen Ereignis aus der vergangenen Geschichte des Patienten auf beinahe jedes hypothetische spätere Ereignis schlossen, war einfach **alarmierend**. So wurde zum Beispiel, die Entscheidung, in die Navy zu gehen, von Vpn als Grund herangezogen, um sowohl die Kandidatur für ein politisches Amt als auch den nachfolgenden Selbstmord zu erklären! Im ersten Fall wurde der Dienst in der Navy als Symptom für 'Geselligkeit' und den 'Wunsch zu dienen', das einen Inhaber eines politischen Amtes angeblich charakterisiert, gesehen. Im zweiten Fall wurde es als Hinweis auf die Disposition des Patienten betrachtet, 'andere durch Weglaufen zu bestrafen', was seinen Schatten auf den nachfolgenden Selbstmord voraus wirft (**Jede Ähnlichkeit zwischen dem Verhalten der Vpn in diesem Experiment und lebenden oder toten klinischen Psychologen ist rein zufällig**).«
(Nisbett & Ross, 1980, Seite. 185)*

*Bemerkung*

Das Problem der Vagheit und Mehrdeutigkeit betrifft nicht nur Alltagstheorien. Auch manche »wissenschaftlichen« Theorien im Bereich der Psychologie und anderer Geisteswissenschaften sind davon betroffen.

Falsche subjektive Theorien sind also eine der zentralen Ursachen für fehlerhafte Kausal- und Kontingenzurteile. Es gibt jedoch noch weitere kognitive Mechanismen, die zu einer verzerrten Wahrnehmung von Zusammenhängen führen. Der folgende Mechanismus steht in engem Zusammenhang mit fehlerhaften subjektiven Theorien.

2.3 Wahrnehmung von Muster in Zufallssequenzen

Eine Stärke von Menschen (und von höheren Tieren) besteht in ihrer Fähigkeit zur Verarbeitung komplexer Muster. So existiert bis heute keine Maschine, deren Fähigkeit, Gesichter oder komplexe sprachliche Gebilde zu verarbeiten, an jene des Menschen heranreicht. Diese ausgeprägte Fertigkeit zur Erkennung von Mustern hat jedoch auch eine negative Seite: Es werden Muster gesehen, wo es keine Regelmäßigkeiten gibt.

Wagenaar (1970) präsentierte seinen Vpn Serien mit weissen und schwarzen Punkten. Die Wahrscheinlichkeit der Abfolge von weissen und schwarzen Punkten wurde variiert: Die Wahrscheinlichkeit des Wechsels von einem Punkt einer Farbe zu einem Punkt der anderen Farbe variierte in .10 Schritten von .20 bis .80. D.h. in einer Serie war die Wahrscheinlichkeit des Wechsels der Farbe .20, in einer anderen

.30 usw. Bei einer Serie mit zwei gleichen Alternativen beträgt die Wahrscheinlichkeit des Wechsels .50.

Die Vpn sollten angeben, welche Serie sie als zufällig mit gleichen Alternativen betrachteten. Im Durchschnitt wurde eine Serie mit der Wahrscheinlichkeit des Wechsels von .60 als zufällig angesehen. Die Vpn glaubten, dass der Wechsel zwischen den Farben öfter erfolgen sollte als dies tatsächlich der Fall ist.



Kognitiver Mechanismus 2-2: Repräsentativität und Zufall

Für den Laien bedeutet Zufall Unvorhersagbarkeit und möglichst unsystematisches ungeordnetes Auftreten von Ereignissen.

Eine durch Zufall generierte Sequenz sollte daher keine längeren Abfolgen von gleichartigen Teilsequenzen enthalten, oder anders ausgedrückt: jede Teilsequenz einer Zufallssequenz muss *repräsentativ* sein für das was als prototypisch für ein Zufallergebnis angesehen wird: unsystematisch, unvorhersehbar.

Personen haben also einen Prototyp (eine Art Minitheorie) davon was Zufall ist. Wenn eine Sequenz diesem nicht entspricht, d.h. diesem nicht ähnlich ist, so kann wird sie nicht als zufällig betrachtet.

Aufgrund des beschriebenen Mechanismus werden Sequenzen mit einer Änderungswahrscheinlichkeit von .50 als zu geordnet angesehen. Die Annahme, Ungeordnetheit und eine wirre Abfolge müsse sich in jeder Teilsequenz einer Zufallssequenz widerspiegeln, ist *falsch*. Vielmehr ist das Auftreten einer längeren Sequenz gleichartiger Ereignisse beim Werfen einer Münze oder von anderen zufallsbedingten Prozessen durchaus wahrscheinlich.

Diese irrige Annahme ist eine Facette jener Phänomene, die Tversky und Kahneman (1971) als den *Glauben an das Gesetz der kleinen Zahl* bezeichnen.



Kognitiver Mechanismus 2-3: Glaube an das Gesetz der kleinen Zahl

Der Glaube an das Gesetz der kleinen Zahl besteht in der Überschätzung der Repräsentativität kleiner Stichproben für die Population. Im Speziellen gilt:

- Leute schätzen die Ähnlichkeit kleiner Stichproben untereinander, bzw. die Ähnlichkeit zwischen Sample und Population als zu hoch ein. D.h. auch kleine Stichproben repräsentieren adäquat die Population und eine weitere gezogene kleine Stichprobe ist ähnlich zusammengesetzt wie eine zuvor gezogene kleine Stichprobe.

- Überschätzung der Repräsentativität kleiner Stichproben in Abhängigkeit von der Populationsgrösse:
Ein Sample wird als umso repräsentativer angesehen, je kleiner das Grössenverhältnis zwischen Sample und Population ist. Von zwei Samples wird jenes, welches aus einer kleineren Population stammt als repräsentativer betrachtet.
- Überschätzung der Repräsentativität von Teilen der Stichprobe: Personen nehmen an, dass alle Teile eines Samples die Populationsverhältnisse widerspiegeln.
- Personen haben die Tendenz, reine Zufallsprozesse als selbst korrigierend zu betrachten. Diese Tendenz nennt sich *Gambler's Fallacy*. Zufallsprozesse mit unabhängigen Ereignissen sind nicht selbst korrigierend, sondern *ausdünnend*. Letzteres bedeutet, dass innerhalb langer Sequenzen eingebettet Folgen von gleichartigen Ereignissen an Bedeutung für die Ermittlung der relativen Häufigkeiten verlieren.



Bemerkung

Der Name *Gesetz der kleinen Zahl* ist eine Anspielung auf das *Gesetz der grossen Zahl*, wonach sich – global gesprochen – mit zunehmender Grösse eine Stichprobe immer mehr der Population annähert.

Der Glaube in das Gesetz der kleinen Zahl zeigt, dass Laien (und auch – wie Tversky und Kahneman feststellten – einige Experten) eine Reihe von falsche Vorstellungen über die Natur von Zufallsprozessen hegen, speziell wenn es sich um kleine Stichproben handelt. Die Wirkung von derart fehlerhaften Vorstellungen findet man häufig bei Interpretationen von Sportereignissen, wo der Zufall oft eine nicht zu vernachlässigende Rolle spielt. Das folgende Beispiel demonstriert dies in eindringlicher Weise.

2.3.1 Die »heisse Hand« im Basketball

Die meisten Basketballexperten (Spieler und Betreuer) und auch viele Laien hegen die Vorstellung von einer so genannten »heissen Hand« im Basketball: Wenn ein Spieler einen Korb (oder mehrere Körbe) erzielt hat, so wird er »heiss«, d.h. seine Trefferwahrscheinlichkeit steigt. Gilovich, Vallone, & Tversky (1985) führten eine Untersuchung durch, um zu prüfen, was es mit dieser Annahme auf sich hat.



*Bsp.2-3: Die »heisse Hand im Basketball«
(Gilovich, Vallone, & Tversky, 1985):*

1. Die Autoren demonstrierten zuerst, dass diese Annahme tatsächlich verbreitet ist: Von 100 Stanford und Cornell Studenten, die sich für Basketball interessierten, gaben 91% an, dass ein Spieler eine bessere Chance hat den nächsten Wurf zu verwandeln, wenn er die letzten 2 oder 3 Schüsse verwandelt hat.
2. Sie werteten die Korbstatistiken für die 9 wichtigsten Spieler der Philadelphia 76er aus 48 Heimspielen der Saison 1980-81 aus:
 - (i) Die Wahrscheinlichkeit einen Treffer zu erzielen war nach einem, zwei, oder drei unmittelbar zuvor erzielten Treffern im Durchschnitt etwas geringer als nach einem, zwei oder drei Fehlwürfen:

$$P(\text{Treffer}|\text{1 Fehlwurf}) = .54 \quad P(\text{Treffer}|\text{1 Treffer}) = .51$$

$$P(\text{Treffer}|\text{2 Fehlwürfe}) = .53 \quad P(\text{Treffer}|\text{2 Treffer}) = .50$$

$$P(\text{Treffer}|\text{3 Fehlwürfe}) = .56 \quad P(\text{Treffer}|\text{3 Treffer}) = .46$$



Bemerkung zur Notation:

Das Symbol $P(X|Y)$ repräsentiert die Wahrscheinlichkeit, dass das Ereignis X eintritt, wenn das Ereignis Y eingetreten ist.

- (ii) Anzahl und Länge der beobachteten »Runs«, d.h. der aufeinander folgenden gleichartigen Ereignisse (seien es nun Fehlwürfe oder Treffer) unterschied sich nicht signifikant von solchen, die durch einen Zufallsprozess mit unabhängigen Durchgängen erzeugt wurden.
3. Es gab auch keine »heissen« und »kalten Nächte«, d.h. Nächte, in denen Spieler eine »heisse« Hand hatten gegenüber Nächten, in denen dies nicht der Fall war.
4. Auch bei Freiwürfen fand sich kein Hinweis, dass nach einem erzielten Korb beim ersten Freiwurf die Wahrscheinlichkeit eines Treffers für den zweiten Wurf steigt.
5. Ein Experiment mit 14 Studenten und 12 Studentinnen der Cornell Basketballmannschaft, welche aus bestimmten Positionen Korbwürfe durchführten, erbrachte keinerlei Hinweise für eine Abhängigkeit des Ergebnisses des letzten Wurfes.
6. Die Studenten und Studentinnen waren auch nicht in der Lage, das Ergebnis ihres nächsten Wurfes oder das der Kollegen korrekt vorher zu sagen.

7. Eine Beurteilung von Sequenzen, wie z. B. (X= Treffer, O = Fehlwurf)

XOXOXOOOXXOXOXOOXXXOX,

durch die 100 Stanford und Cornell Studenten (siehe 1) ergab, dass erst ab einer Wahrscheinlichkeit des Wechsels von $\geq .70$ die Sequenz als zufallserzeugt eingeschätzt wurde und nicht als »Streak« Sequenz, d.h. eine mit mehr oder weniger längeren »Runs«, die durch »heisse« oder »kalte Hände« erzeugt wurden.

Die Untersuchung von Gilovich, Vallone & Tversky (1985) zeigt eindeutig, dass das Phänomen der »heissen Hand« nicht existiert, obwohl die meisten Experten daran glauben. Die Tatsache, dass Sequenzen mit einer Wahrscheinlichkeit des Wechsels von .50 mehrheitlich als »Streak« Sequenzen eingeschätzt werden, weist auf die Wirkung der Repräsentativität bzw. des Glaubens an das Gesetz der kleinen Zahl hin, wonach sich der Zufall in allen Teilsequenzen einer Zufallssequenz widerspiegeln sollte.



Bemerkungen:

1. Die Nichtexistenz der »heissen Hand« beim Basketball wurde in bestimmten Kreisen mit Unglauben aufgenommen (Larkey, Smith, & Kadane, 1989). An den statistischen Fakten, welche die Unabhängigkeit der Treffer Sicherheit vom letzten Erfolg zeigen, kann jedoch nicht gerüttelt werden (Tversky & Gilovich, 1989a,b).
2. In anderen Sportarten (Billiard, Golf, Hufeisenwerfen, Pfeilwerfen, Kegeln) scheint jedoch so etwas wie eine »heisse Hand« zu existieren.

2.3.2 »Heisse Hand« und Gambler's Fallacy

Gilovich, Vallone & Tversky (1985) erklären den Glauben an die »heisse Hand« im Basketball mit Hilfe der Repräsentativität, indem die Personen kurze Serien von erfolgreichen Würfeln als Hinweis auf die »heisse« Hand deuten, obwohl keine statistischen Abhängigkeiten vorliegen und die Serien durchaus zufallsbedingt sein können.

Andererseits wird die Gambler's Fallacy mit Hilfe des Glaubens der Personen erklärt, wonach der Zufall ausgleichend wirkt. Es stellt sich daher die Frage:



Frage:

Warum glauben die Personen im Falle des Basketballs an eine Erhöhung der Wahrscheinlichkeit des Auftretens eines gleichartigen Ereignisses und im Falle der Gambler's Fallacy an eine Verringerung?

Die Antwort auf diese Frage liegt wiederum im Vorliegen einer subjektiven Theorie: Im Falle des Basketballs gibt es eine subjektive Theorie, welche die »heisse Hand« als plausibel erscheinen lässt, da das Ergebnis nicht rein zufallsbedingt ist, sondern auch noch von Fähigkeiten des Sportlers abhängt. Im Falle eines reinen Zufallsprozesses wie beim Münzwurf oder beim Roulette spielen menschliche Fähigkeiten keine Rolle. Hier dominiert daher die subjektive Theorie, welche die Personen über Zufallsprozesse haben (nämlich, dass sich diese ausgleichen).

2.3.3 Sind die »heisse« Hand und die Gambler's Fallacy rational?

Ayton und Fischer (2004) behaupten, dass eine Strategie, welche von der Anzahl »Streaks« (=Sequenzen gleicher Ergebnisse, seien es Treffer oder Fehlwürfe) abhängt, *instrumentell rational* sein kann. So könnte, die Entscheidung, welchem Spieler der Ball zugespielt wird, von der Anzahl der »Streaks« des Spielers abhängen, indem z.B. einer Person mit vielen Sequenzen von Treffern der Ball öfter zugespielt wird. Diese Argumentation ist nicht nachvollziehbar, denn wie bereits Gilovich, Vallone und Tversky (1985) feststellten, macht es wohl wenig Sinn den Ball im Spiel vermehrt jener Person zuzuspielen, die angeblich die »heisse Hand« hat. Ebenso wenig scheinen mir andere Strategien sinnvoll, welche auf dem Glauben an die »heisse« Hand beruhen, obwohl diese nicht existiert. Da Ayton und Fischer (2004) keine genaueren Angaben machen, gibt es keinerlei Evidenz, dass der Glaube an die »heisse Hand« einen Vorteil bringt.

Auch bezüglich der Gambler's Fallacy gibt es Meinungen, welche diese als gerechtfertigt ansehen. So meint Pinker (1997), dass wir im Alltag kaum stochastische unabhängige Ereignisse (wie beim Roulette oder Münzwurf) antreffen. Vielmehr sind wir im Alltag mit begrenzten Populationen und mit Stichproben ohne Zurücklegen konfrontiert. In diesem Fall macht es Sinn anzunehmen, dass mit jedem gesehenen Ereignis eines bestimmten Typs die Wahrscheinlichkeit dieses Ereignisses sinkt: Wenn ich schon 100 Waggons eines Zuges beobachtet habe, so ist es sinnvoll anzunehmen, dass bald das Ende des Zuges erreicht sein wird. Gegen dieses Argument ist wenig einzuwenden. Nur lässt sich damit nicht die Behauptung rechtfertigen, dass es sich bei der Gambler's Fallacy um rationales Verhalten handelt, denn die einzelnen Ausfälle beim Roulette und Münzwurf sind (mit sehr grosser

Annäherung) voneinander unabhängig und daher ist das Modell nicht anwendbar.

Wir wenden uns als nächstes einem weiteren Mechanismus zu, welcher zur fehlerhaften Beurteilung von Zusammenhängen führt.

2.4 Ignorierung des Phänomens der Regression zum Mittel

Das Phänomen der Regression zum Mittel wurde von Sir Francis Galton (1822-1911) entdeckt. Er fand in seinen eugenischen Studien dass extreme Charakteristika der Eltern nicht auf die Kinder übertragen werden. Zum Beispiel haben sehr grosse oder kleine Väter Söhne, die eher von mittlerer Grösse sind. Galton interpretierte diese Ergebnisse zu erst völlig falsch, indem er annahm, dass es genetische Ursachen für dieses Phänomen gäbe. Konkret nahm er an, dass die Merkmale des Kindes teilweise von Eltern und teilweise von früheren Vorfahren bestimmt seien. Er konstruierte also eine kausale Theorie, um das Phänomen zu erklären. Dann fand er jedoch, dass das Phänomen auch in die umgekehrte Richtung funktioniert: Väter von sehr grossen / kleinen Kindern sind eher von mittlerer Grösse.

Im Folgenden wird zuerst das Prinzip der Regression zum Mittel diskutiert und anschliessend werden Urteilsfehler, welche auf der Ignorierung des Prinzips beruhen, geschildert.

2.4.1 Das Prinzip der Regression zum Mittel

Bevor eine formale Beschreibung des Prinzips gegeben wird, präsentiere ich ein anschauliches Beispiel.



Bsp.2-4: Regression zum Mittel

Angenommen Sie nehmen im Rahmen Ihres Studiums an einer Prüfung mit multiple-choice Fragen Teil. Der Einfachheit halber gebe es pro Frage zwei Antworten, eine falsche und eine korrekte.

Nach absolviertem Test werden die besten und die schlechtesten 10% der Teilnehmer einem zweiten Test mit neuen Fragen aus dem gleichen Fragenpool unterzogen.

Ergebnis:

- Die 10% besten Studentinnen werden im zweiten Test im Durchschnitt ein schlechteres Ergebnis erzielen als im ersten.
- Die 10% schlechtesten Studentinnen werden im zweiten Test im Durchschnitt ein besseres Ergebnis erzielen als im ersten.

Begründung:

Das Ergebnis im ersten Test ist durch zwei Faktoren beeinflusst: (1) Das Wissen der Personen und (2) Zufall (Glück). Das Wissen der Personen wird diesen auch im zweiten Test helfen, aber man kann nicht erwarten, dass ihnen das Glück auch im zweiten Test in diesem Ausmass »hold ist«.

Bemerkung:

Das Phänomen ist am stärksten, falls das Wissen für das Ergebnis überhaupt keine Rolle spielt und das Ergebnis im ersten Test rein durch Zufall entstanden ist.

2.4.2 Die durch das Ignorieren der Regression zum Mittel bedingten Urteilsfehler

Das oben dargestellte historische Beispiel von Francis Galton demonstriert bereits die fehlerhaften Schlüsse, die sich bei Ignorierung des Phänomens der Regression zum Mittel typischerweise ergeben: Es werden irgendwelche kausale Theorien fingiert um das Phänomen zu erklären. Im Folgenden betrachten wir einige Alltagsbeispiele, die demonstrieren sollen, wie sich das Ignorieren der Regression zum Mittel im Alltag auswirkt.

2.4.2.1 DER FLUGLEHRER

Kahneman und Tversky (1973) präsentieren ein Beispiel, welches das Phänomen schön illustriert: Ein Flugtrainer der israelitischen Luftwaffe erklärte Daniel Kahneman, dass er nichts von positiver Verstärkung seine Schützlinge halte, denn immer wenn er sie nach einem besonders gelungenem Flugmanöver lobte, so war ihre Leistung beim nächsten Mal schlechter. Tadelte er sie hingegen nach einer schlechten Leistung, so zeigten sie beim nächsten Mal eine bessere Leistung.

Der Trainer erkennt hier offenbar nicht das Phänomen der Regression zum Mittel, demgemäss nach einer sehr guten Leistung wahrscheinlich eine schlechtere folgt und umgekehrt.

2.4.2.2 ERKLÄRUNGEN IN SPORT UND BERUF

Die meisten sportlichen Leistungen sind sowohl durch Geschick als auch durch Zufall beeinflusst. Es scheint jedoch so zu sein, dass die meisten Personen (darunter viele Sportexperten) die Rolle des Zufalls unterschätzen. Im Speziellen wird die Bedeutung der Regression zum Mittel nicht erkannt.

Angenommen ein Neuling (sei es ein Sportler oder eine Mannschaft) zeigt extrem gute Leistungen, sodass alle Sport-Gazetten aufmerksam werden. Plötzlich jedoch nimmt die Leistung ab. In diesem Fall gibt es dann immer zahlreiche Erklärungen, wie z.B. »der Erfolg ist ihm (ihnen) in den Kopf gestiegen« (Nisbett & Ross, 1980). Hierbei wird die Regression zum Mittel übersehen.

Das Phänomen der Regression zum Mittel erklärt z.B. sehr gut, warum in der Deutschen Fussballbundesliga kaum jemals eine Mannschaft mit Ausnahme von Bayern München zweimal hintereinander die Meisterschaft gewinnt (Konkret: Zwischen 2000 bis 2010 wurde Bayern München 6× Meister während die anderen 4 Meisterschaften von verschiedenen Mannschaften geholt wurde: Bremen, Dortmund, Stuttgart und Wolfsburg). Scheinbar hat nur Bayern München als einzige Mannschaft wirklich die Klasse zum Meister.

Ignorieren der Regression zum Mittel findet man auch bei Erklärungen zur Leistung im Beruf (Nisbett & Ross, 1980):

Leute sehen sich oft dazu veranlasst, das Scheitern hervorragender Studenten im Beruf durch Argumente zu erklären, wie: »Sie hat glänzende Fähigkeiten, aber ihr fehlt leider der nötige *Drive*«, »Er hat sich zu sehr in administrative Details verstricken lassen«, oder »Sie erhielt einfach nicht die notwendige Unterstützung und Hilfe ihrer Kollegen«. Eine viel einfachere Erklärung besteht einfach darin, dass zwischen Erfolg im Studium und Erfolg im Beruf nur ein geringer Zusammenhang besteht.

Die ad-hoc Erklärungen verhüllen nur die Tatsache, dass das Scheitern erfolgreicher Studenten im Beruf eher die Regel ist als die Ausnahme.

2.4.2.3 FAIRNESS GEGENÜBER BENACHTEILIGTEN GRUPPEN

In den USA gibt es seit längerem Bemühungen, die benachteiligte schwarze Bevölkerung durch spezielle Massnahmen zu fördern. Eine Massnahme besteht darin, schwarze Studentinnen mit schlechteren Scores als die weissen Kolleginnen zu Studien zuzulassen.

Die *Strebertheorie* besagt, dass schwarze Studentinnen, die im Vergleich zur eigenen Gruppe eine sehr gute Testleistung bringen (die »Streber«) sich nach ihrer Zulassung zum Studium besonders gut entwickeln. Daher sollten diese Personen, die jedoch im Vergleich zur Weissen Gruppe schlechter sind, gegenüber ihren weissen Kolleginnen mit den gleichen Leistungen bevorzugt werden.

Die Theorie der Regression zum Mittel besagt hingegen, dass sich diese Personen aufgrund der Regression zum Mittel eher schlechter entwickeln werden. Konkret sollte die zukünftige Testleistung τ_s einer Schwarzen, die sich im oberen Bereich der relevanten Bezugsgruppe befindet, eher schlechter ausfallen als die aktuell beobachtete Leistung x_s . Für eine weisse Person mit der identischen aktuellen Leistung x_w ($x_s = x_w$) sollte hingegen die zukünftige Leistung τ_w besser ausfallen, da sich diese Person eher im unteren Bereich hinsichtlich der relevanten Bezugsgruppe befindet (siehe Abb. 2-1).

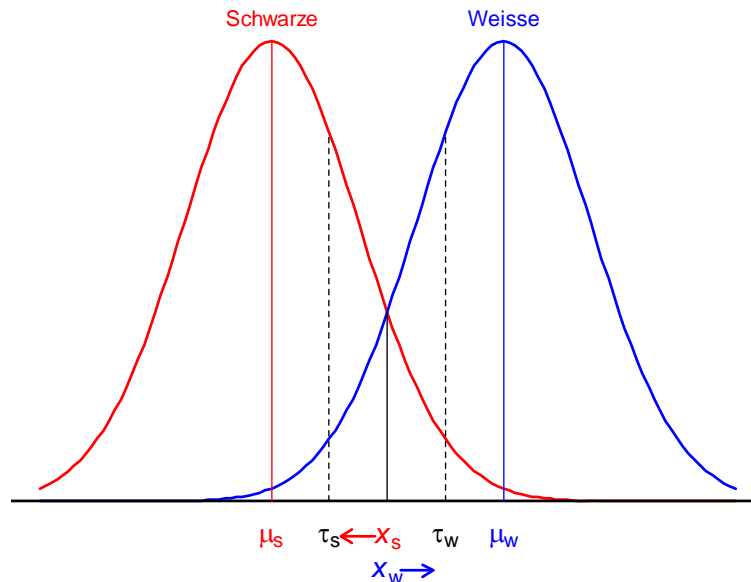


Abb. 2-1: Demonstration der Regression zum Mittel.

Der Harvard Statistik Professor Alan Zaslavsky argumentiert, dass die Personen der benachteiligten Gruppe, sobald sie einmal ausgewählt wurden, die Bürden dieser Gruppe abwerfen. Daher gäbe es auch keinen Effekt der Regression zum Mittel, da die Gruppe nicht mehr die relevante Bezugsgruppe sei. Diese Theorie wurde nun in mehreren Studien geprüft (vgl. Wainer & Brown, 2007).

An einer Studie nahmen mehr als 46'000 Studenten aus 38 Colleges teil. Der Wert des SAT (Scolastic Aptitude Test) wurde zur Vorhersage des HS-GPA (High School Grade Point Average) verwendet. Es zeigte sich ein klarer Regressionseffekt: Für Asiaten und Weiße ergab sich eine Zunahme des HS-GPA gegenüber dem SAT, während sich für Schwarze und Hispanics eine Abnahme ergab.

In einer anderen Studie wurde die Leistung bei den Prüfungen in Medizin aufgrund der Einstiegsprüfung für Weiße und Schwarze vorhergesagt. Wiederum wurde das Modell der Regression zum Mittel bestätigt: Schwarze mit gleichem Wert in der Einstiegsprüfung wie Weiße sind in den nachfolgenden Prüfungen konsistent schlechter als ihre weissen Kolleginnen.

Demnach ist die Theorie von Zaslavsky, wonach die Personen benachteiligter Gruppen nach Auswahl für ein Studium die »Fesseln ihrer Herkunft« abzulegen vermögen, wenig stichhaltig.

Eine detaillierte und formellere Analyse der Regression zum Mittel mit Anwendungen auf die Statistik wird in Abschnitt 2.6.5 präsentiert. Wir wenden uns nun einer anderen Ursache zu, welche für fehlerhafte Kontingenz- und Kausalurteile verantwortlich ist.

2.5 Fehlerhafte Gewichtung von relevanter Information

Zahlreiche Studien belegen, dass Personen Information, welche für die Beurteilung eines (kausalen) Zusammenhanges von Bedeutung sind, entweder ignorieren oder zu wenig gewichten.

2.5.1 Beurteilung von Information in Kontingenztabelle

Eine bequeme Art zur Darstellung von Wahrscheinlichkeitsinformation bildet eine so genannte *Kontingenztabelle* (*Kreuztabelle*). Obwohl diese die gesamte relevante Wahrscheinlichkeitsinformation über den Zusammenhang zweier Attribute enthält, ist es für Laien meist nicht möglich, aus der Tabelle die relevante Information über den stochastischen Zusammenhang der beiden Merkmale zu entnehmen. Das folgende Beispiel soll dies illustrieren.



Bsp.2-5: Information in Kontingenztabelle

Gegeben sei eine neue Methode zur Raucherentwöhnung. Um deren Wirksamkeit zu prüfen wurden $N = 340$ Personen untersucht.

Als Kriterium für den Erfolg galt, dass die untersuchten Personen mindestens 30 Tage lang nicht rauchten.

Eine Gruppe, welche der neuen Methode unterzogen wurden, wurde mit einer Kontrollgruppe ohne Behandlung verglichen.

Tab. 2-1 zeigt die Anzahl Personen, welche bei Anwendung der Methode bzw. ohne deren Anwendung gemäss dem oben gegebenen Kriterium erfolgreich waren oder nicht.

	Erfolgreich	Nicht erfolgreich
Neue Methode angewendet	200	75
Keine Behandlung	50	15

Tab. 2-1: Anzahl erfolgreicher/ erfolgloser Fälle bei Anwendung einer Behandlung bzw. bei Nichtanwendung.

Viele Versuchspersonen, denen diese Tabelle präsentiert wurde, sehen einen positiven Zusammenhang zwischen der neuen Behandlungsmethode und einer erfolgreichen Raucherentwöhnung. In Wirklichkeit ist der Zusammenhang jedoch leicht negativ.



Kognitiver Mechanismus 2-4: Fehlerhafte Gewichtung von Kontingenzinformation

Gegeben:

Eine 2×2 - Kontingenztabelle (4-Felder-Tafel, Kreuztabelle) zur Repräsentation des gemeinsamen Auftretens zwei Ereignisse (Tab. 2-2). Die vier Zellen der Tabelle sind mit den Buchstaben (a)–(d) bezeichnet.

Attribut 1	Attribut 2	
	1	0
1	(a)	(b)
0	(c)	(d)

Tab. 2-2: 2×2 - Kontingenztabelle zur Repräsentation zweier Attribute mit je zwei Ausprägungen, die jeweils mit 1 und 0 bezeichnet werden. Die Buchstaben (a)–(d) bezeichnen die 4 Zellen der Tabelle.

Die Information in den verschiedenen Zellen wird unterschiedlich gewichtet, und zwar wie folgt (Schustack & Sternberg, 1981):

$$(a) > (b) > (c) > (d).$$

D.h. die Information in Zelle (a) wird stärker gewichtet als jene in Zelle (b), diese stärker als die in Zelle (c) und diese wiederum stärker als jene in Zelle (d).

Die Kontingenztabelle von Tab. 2-1 enthält besonders viele Fälle in Zelle (a). Dies verleitet viele Personen dazu, einen positiven Zusammenhang zu vermuten.



Method 2-1: Beurteilung der Information innerhalb von 2×2 - Kontingenztabellen:

Gegeben:

Die 2×2 - Kreuztabelle von Tab. 2-2.

Es gilt:

1. Die Zellen (a) und (d) enthalten Evidenz für einen positiven Zusammenhang zwischen den Attributen, die Zellen (b) und (c) Evidenz für eine negative Assoziation.

Bemerkung:

Hierbei wird angenommen, dass die Ausprägungen 1 für beide Attribute das Vorhandensein und die Werte 0 die Abwesenheit repräsentieren.

2. Für die Beurteilung eines Zusammenhanges ist die Information in allen vier Zellen *in gleicher Weise* relevant.

3. Um zu prüfen, ob ein positiver oder negativer bzw. kein Zusammenhang besteht, berechne man:

$$a \cdot d - b \cdot c$$

4. Ein einfaches normiertes Maß, welches einer Rangkorrelation entspricht, ergibt sich durch *Yules Q*:

$$Q = \frac{a \cdot d - b \cdot c}{a \cdot d + b \cdot c}.$$

Im Zähler befindet sich die Differenz und im Nenner die Summe aus den beiden Produkten.

Bemerkung:

Bei einem perfekten positiven Zusammenhang sind nur die Zellen (*a*) und (*d*) besetzt. Folglich gilt: $Q = 1$.

Im Falle einer perfekten negativen Assoziation sind nur die Zellen (*b*) und (*c*) ungleich 0. Daher gilt: $Q = -1$.

Für die Daten in Tab. 2-1 ergibt sich:

$$Q = \frac{a \cdot d - b \cdot c}{a \cdot d + b \cdot c} = \frac{-750}{6750} = -0.111.$$

Es gibt also eine schwache negative Assoziation zwischen der neuen Methode und dem Behandlungserfolg.

Die klassische Untersuchung von Hamilton und Gifford (1976) demonstriert, wie die unterschiedliche Gewichtung von Informationen die Wahrnehmung von sozialer Gruppen verschiedener Grösse beeinflussen kann.



Bsp.2-6: Illusorische Korrelation und die Wahrnehmung von Minoritäten (Hamilton & Gifford, 1976)

Den Vpn wurden 39 Verhaltensbeschreibungen präsentiert. Diese Beschreibungen wurden auf 2 Gruppen A und B aufgeteilt (Um jegliches Vorwissen auszuschalten, wurden für die Gruppen nur die abstrakten Bezeichnungen A und B verwendet).

26 Beschreibungen, 18 positive und 8 negative, wurden der Gruppe A zugeordnet und 13 Beschreibungen, 9 positive und 4 negative der Gruppe B.

Gruppe A mit der grösseren Anzahl zugeordneter Beschreibungen ist daher die Majoritäts- und Gruppe B die Minoritätsgruppe.

Tab. 2-3 zeigt die Anzahl positiver und negativer Verhaltensbeschreibungen für die beiden Gruppen.

Gruppe	Verhaltens		Σ
	Positiv	Negativ	
A (Majorität)	18	8	26
B (Minorität)	9	4	13

Tab. 2-3: Verteilung der positiven und negativen Verhaltensbeschreibungen über die beiden Gruppen im Experiment von Hamilton & Gifford (1976).

Ergebnis:

Die Majoritätsgruppe A wird in einer nachfolgenden Beurteilung durch die Vpn als positiver eingeschätzt als die Minoritätsgruppe B.

Interpretation:

Die Überbewertung der Information in Zelle (a) führt offenbar dazu, dass die Personen einen positiven Zusammenhang zwischen der Majoritätsgruppe und positivem Verhalten sehen.

Da das Verhältnis von positiven zu negativen Verhaltensweisen (18/8 bzw. 9/4) in beiden Gruppen gleich ist besteht offensichtlich kein Zusammenhang zwischen den Gruppen und den Verhaltensweisen.

Die differentielle Gewichtung von Information in Kontingenztabelle kann als ein Ergebnis von Aufmerksamkeits- oder Salienzeffekten beurteilt werden: Personen schenken der Information in Zelle (a) mehr Aufmerksamkeit als den Informationen in den anderen Zellen, bzw. die Information in Zelle (a) ist salienter. Aufmerksamkeits- und Salienzeffekte spielen im Allgemeinen eine wichtige Rolle für die Wahrnehmung (kausaler) Zusammenhänge.

2.5.2 Aufmerksamkeits- und Salienzeffekte

Salienzeffekte findet man sehr häufig im Bereich der Erklärung von eigenem und fremdem Verhalten. Den Zweig, welcher sich mit diesen Arten von Erklärungen beschäftigt, bezeichnet man als *Attributions-theorie*. Bevor wir uns mit Attributionsfehlern beschäftigen sei kurz eine Darstellung der sehr einflussreichen Theorie von Harold Kelley präsentiert.

2.5.2.1 DAS ANOVA MODELL VON KELLY

Das ANOVA (*Analysis of Variance*) Modell von Harold Kelley (1967) nimmt an, dass wir uns im Alltag wie *Laien-Wissenschaftler* verhalten, indem wir die Daten aus verschiedenen Informationsquellen in optimaler Weise miteinander verrechnen.

Grundsätzlich können die möglichen Ursachen eines bestimmten Verhaltens in drei Kategorien eingeteilt werden:

- 1. Person:** Hierbei handelt es sich um Ursachen, die als innerhalb der Person befindlich betrachtet werden: Wünsche, Absichten, Fertigkeiten etc. sind für das beobachtete Verhalten verantwortlich.
- 2. Objekt:** Der Stimulus bzw. das Objekt ist relevant für das Verhalten.
- 3. Situation:** Die spezielle Situation ist für das Zustandekommen verantwortlich.



Bsp.2-7: Tanzunterhaltung:

Zu erklärendes Phänomen:

Hans trat Maria beim Tanzen ständig auf die Füße.

Personenattribution:

Die Ursache ist in Hans zu suchen (z.B. Hans kommt beim Tanzen ständig aus dem Rhythmus).

Objektattribution:

Die Ursache ist in Hansens Partnerin Maria zu suchen (z.B. sie möchte beim Tanzen immer die Führung übernehmen, was meist zu Unstimmigkeiten mit den Tanzpartnern führt).

Situationsattribution:

Die Ursache ist in der Situation zu suchen (z.B. die Tanzfläche war völlig überfüllt, so dass Hans und Maria ständig von anderen Paaren angerempelt wurden).

Die drei möglichen Erklärungen in Bsp.2-7 lokalisieren die Ursache für das zu erklärende Phänomen an jeweils unterschiedlicher Stelle. Man beachte auch, dass die möglichen Erklärungen, welche sich in den Klammern befinden, Hintergrundwissen über die Person, das Objekt und die Situation voraussetzen.

Wie kann man nun aufgrund von reiner Beobachtung (ohne Hintergrundwissen) die Ursache lokalisieren? Kelley nimmt an, dass die Person drei Informationsquellen verwendet, um zu einem Kausalurteil zu kommen:

- 1. Konsensus:** Verhalten sich andere Personen ähnlich in der gleichen Situation mit gleichem Objekt? [Variation des Effekts mit der Person]
- 2. Distinktheit:** Führen andere Stimuli / Objekte zum gleichen Ergebnis? [Variation des Effekts mit dem Objekt]
- 3. Konsistenz:** Zeigte sich das Verhalten der Person gegenüber dem Objekt auch in anderen Situationen? [Variation des Effekts mit der Situation bzw. über die Zeit hinweg]



Bsp.2-8: Tanzunterhaltung (Fortsetzung von Bsp.2-7):

Konsensusinformation:

Diese Information gibt Auskunft darüber, ob auch andere Personen an jenem Abend Maria auf die Füße traten.

Distinktheit:

Betrifft die Information, ob Hans an jenem Abend nur Maria auf die Füße trat oder auch anderen Tanzpartnerinnen.

Konsistenz:

Betrifft die Frage, ob Hans Maria auch früher schon beim Tanzen auf die Füße trat.

Aufgrund von Information über diese drei Arten von Variationen kann eine Person im Idealfall entscheiden, wo die Ursache für das Ereignis zu suchen ist.



Bsp.2-9: Tanzunterhaltung (Fortsetzung von Bsp.2-7):

Konsensus = gering:

Andere Personen traten Maria nicht auf die Füße.

Distinktheit = gering:

Hans trat ausnahmslos allen seinen Partnerinnen auf die Füße.

Konsistenz = hoch:

Hans trat auch früher schon Maria beim Tanzen auf die Füße.

Schlussfolgerung: Die Ursache ist bei Hans zu suchen.

Gemäss dem Modell von Kelley prüfen also die urteilenden Personen, wie der Effekt mit der Person, dem Objekt oder der Situation variiert. Aufgrund des Musters der Variation lokalisieren sie die Ursache: Man kann dies so betrachten, als ob die Urteilenden eine Art qualitative Varianzanalyse (ANOVA) durchführen, indem sie die unterschiedlichen Varianzquellen für das Verhalten analysieren.

Das Modell von Kelley beschreibt die Dynamik menschlicher Attributionen nur in sehr eingeschränktem Maße korrekt. Im Wesentlichen gibt es folgende Gründe für das Scheitern des Modells:

1. Das Modell ist ein so genanntes *datengetriebenes* Modell, d.h. es beschreibt den Prozess der Kausalzuschreibung als rein durch die vorliegenden Fakten bedingt. Es verkennt daher die Bedeutung der subjektiven Alltagstheorien, die – wie bereits oben gezeigt – einen mächtigen Einfluss auf kausale Erklärungen ausüben.
2. Eine zweite Schwäche des Modells besteht im Ignorieren von Aufmerksamkeits- und Salienzeffekten. Diese führen dazu, dass Personen den Einfluss verschiedener Ursachen falsch beurteilen. Im Speziellen wird wenig salienten – aber relevanten – Ursachen zu

wenig und hervorstechenden – aber irrelevanten – Ereignissen – zu viel Beachtung beigemessen.



Bemerkung:

Eine weitere Schwäche des Modells besteht darin, dass motivationsgeleitete Erklärungen des Verhaltens – wie self-serving biases – ignoriert werden (vgl. Abschnitt 1.2).

2.5.2.2 NICHTBEACHTUNG SUBTILER EINFLUSSGRÖSSEN AUF DAS EIGENE VERHALTEN

Nisbett und Wilson (1977) beschreiben eine Reihe von Untersuchungen, in denen das Verhalten von Personen auf subtile Weise beeinflusst wurde, ohne dass die Personen dies bemerkten. Zur Illustration hier zwei Beispiele.



Bsp.2-10: Nichterkennen eines Halo-Effekts
(Nisbett & Wilson, 1977):

Die Vpn des Experiments sollten folgende drei persönliche Merkmale eines Belgischen Psychologielehrers beurteilen:

- die äussere Erscheinung
- die Manieriertheit
- sowie die sprachliche Eigenheit (europäischer Akzent)

Es gab zwei experimentelle Bedingungen:

In der *Warm-Bedingung* beantwortet der Vortragende Fragen in freundlicher und enthusiastischer Weise.

In der *Kalt-Bedingung* beantwortete der Vortragende die Fragen auf autoritäre und intolerante Weise.

Ergebnisse:

1. Wenig überraschend, erhielt die Person in der *Warm-Bedingung* höhere Sympathie-Beurteilungen als in der *Kalt-Bedingung*.
2. Es zeigte sich aber auch ein so genannter *Halo-Effekt*: Die drei einzuschätzenden persönlichen Eigenheiten wurden in der *Warm-Bedingung* als attraktiv eingeschätzt, während die objektiv gleichen Eigenheiten in der *Kalt-Bedingung* als irritierend betrachtet wurden.
3. Die Vpn glaubten jedoch (irrtümlicherweise), dass die 3 Eigenheiten das Sympathieurteil negativ beeinflusst hätte und nicht umgekehrt.

Interpretation:

Die Personen erkennen offensichtlich nicht die Relevanz des eher subtil fungierenden *Halo-Effekts*.



Konzept 2-2: Halo Effekt (Cooper, 1981; Rosenzweig, 2007):

Der *Halo-Effekt* besteht in einer überhöhten Korrelation von Beurteilungen verschiedener Merkmale eines Objekts.

Der Halo-Effekt setzt sich aus unterschiedlichen Teileffekten zusammen:

- (i) Der Beeinflussung der Wahrnehmung eines Merkmals durch ein anderes Merkmal, oder
- (ii) Der Beeinflussung eines Merkmals durch den Gesamteindruck, oder
- (iii) Der Beeinflussung des Gesamteindrucks durch ein Merkmal.

Alle drei Arten von Effekten werden als Halo-Effekt bezeichnet.

Der Halo-Effekt wurde von Thorndike (1920) so benannt. Er fand, dass Soldaten im 1. Weltkrieg von ihren Vorgesetzten entweder als auf allen Merkmalen überragend oder als generell inferior beschrieben wurden.

Eine wichtige Determinante des Halo-Effekts scheint das Streben nach Konsistenz zu sein.

Halo-Effekte sind immer dann zu erwarten, wenn ein Objekt gleichzeitig auf verschiedenen Merkmalen zu beurteilen ist.

Typische Beispiele:

- Nach dem 11. September 2001 wurde George Bush im Allgemeinen hoch bewertet. Interessanterweise stieg auch die Einschätzung seiner ökonomischen Kompetenz von 47% auf 60% (obwohl Letzteres kaum etwas mit den Vorfällen um 9/11 zu tun hatte).

Im Oktober 2005, als die Unterstützung für den Irakkrieg schwand und nach dem Debakel um den Hurrikan Kathrina sank auch die Zustimmung zu Bush und zwar *gleichermaßen auf allen Skalen*.

- Attraktive Personen werden oft als intelligenter, wärmer, sympathischer etc. beurteilt.
- Wissenschaftliche Manuskripte von Personen aus berühmten Universitäten werden eher akzeptiert.



Bsp.2-11: Nichterkennung eines Ankereffekts (Nisbett & Wilson, 1977):

Die Vpn des Experiments sollten das Verhalten eines typischen Studenten der Universität von Michigan in verschiedenen experimentellen Situationen vorhersagen.

Ein Teil der Vpn erhielt einen *Anker* in Form von Wissen über einen »zufällig ausgewählten Studenten«.

Ergebnisse:

1. Der Anker hatte unterschiedliche Wirkung auf die Ratings der Vpn: Von sehr stark in Richtung Anker bis eher in die entgegen gesetzte Richtung.
2. Die Vpn, bei welchen der Anker eine grosse Wirkung zeigte, beurteilten dessen Einfluss als gleich wie jene, bei denen keinerlei Wirkung des Ankers vorlag.

Interpretation:

Die Personen erkennen offensichtlich nicht wirklich den Einfluss des Ankers auf ihre Einschätzung.

2.5.2.3 NICHTBEACHTUNG VON KONSENSUSINFORMATION

Gemäss dem Modell von Kelley wird Konsensusinformation, d.h. Information darüber, wie sich andere Personen in der gleichen Situation verhalten, für Ursachenzuschreibungen herangezogen. Viele Studien zeigen jedoch, dass Konsensusinformation oft wenig bis überhaupt nicht beachtet wird.

*Bsp.2-12: Nichtbeachtung von Konsensusinformation*

Miller, Gillen, Schenker, & Radlove (1973):

Die Vpn wurden in 2 Gruppen A & B eingeteilt:

Gruppe A:

Die Vpn erhielten die Daten aus Milgrams (1963) Experiment, die zeigen, dass praktische alle Versuchspersonen substantielle E-Schocks anwendeten und dass die Mehrheit (65% der Vpn) das Maximum applizierte.

Gruppe B:

Die Vpn wurden in ihrem naiven Glauben darüber belassen, wie viele Vpn die maximale Schockmenge anwendete.

Bemerkung:

Milgram hatte vor dem Experiment eine Reihe von Psychiatern schätzen lassen, wie viel Prozent der Vpn das Maximum an Schock verabreichen würden. Die Schätzung lag bei 1%.

Anschliessend mussten die Vpn 11 Traits – wie Aggressivität, menschliche Wärme, Liebreiz, etc – von 2 Personen aus Milgrams Experiment, die bis zum Maximum des Schocks »gegangen« waren, einschätzen.

Ergebnis:

Nur auf einer der 11 Traits zeigte sich ein Unterschied zwischen den Einschätzungen der beiden Gruppen.

Die Information, dass die Mehrheit der Teilnehmer an Milgrams Experiment die maximale Stärke des E-Schocks verabreichte, hatte demnach fast keinen Einfluss auf die Beurteilung der Personen (die extrem negativ ausfiel).

Interpretation:

Konsensusinformation ist *Basisrateninformation*, d.h. Information über das Auftreten eines Merkmals, Verhaltens etc. innerhalb der betrachteten Population. Viele Studien zeigen, dass Basisrateninformation nicht ausreichend beachtet wird, wenn nicht spezielle Massnahmen ergriffen werden, um sie salienter zu machen (vgl. Kapitel **xxxx**).

2.5.2.4 ÜBERBEWERTUNG IRRELEVANTER EINFLUSSGRÖSSEN AUF DAS EIGENE VERHALTEN

Neben der Nichtbeachtung von relevanter aber wenig hervorstechender Information gibt es auch das umgekehrte Phänomen, dass nämlich Information, die sehr auffällig ist, aber das Verhalten nicht beeinflusst, als kausal relevant betrachtet wird. Auch hierzu ein Beispiel:



Bsp.2-13: Fehlerhafte Beurteilung des Einflusses einer Rückversicherung auf die Bereitschaft E-Schocks zu ertragen Nisbett & Wilson (1977):

Die Vpn sollten vorhersagen, welche Schockstärke sie in einer Untersuchung zur Wirkung von E-Schocks ertragen würden.

In einer Version wurde versichert, dass die E-Schocks keine dauerhaften negativen Auswirkungen auf die Gesundheit hätten, in der zweiten Version fehlte diese Passage.

Ergebnisse:

1. Die Passage hatte keinerlei Einfluss auf die vorhergesagte zu ertragende Schockstärke.
2. Personen mit der Passage geben mehrheitlich an, dass diese ihre Vorhersage erhöht hätte.
3. Personen, welche die Passage nicht erhalten hatten, gaben mehrheitlich an, dass die Einbeziehung der Passage einen Effekt auf die Höhe der Vorhersage gehabt hätte.

2.5.2.5 DER FUNDAMENTAL ATTRIBUTIONSFEHLER

Ein weiteres Phänomen, welches sich zumindest teilweise mit Hilfe von Salienzeffekten erklären lässt, bildet der fundamentale Attributionsfehler.

*Konzept 2-3: Fundamentaler Attributionsfehler:*

Der *fundamentale Attributionsfehler* besteht in der Tendenz, die Ursachen für ein Verhalten der Person zuzuschreiben, obwohl gute Gründe vorliegen, welche zeigen, dass die Situation wesentlichen Anteil am Verhalten hat.

Zur Illustration des Phänomens, hier zwei Beispiele:



Bsp.2-14: Fundamentalener Attributionsfehler (Jones & Harris (1967):

Die Vpn erhielten Essays, welche Personen zu einem bestimmten Thema (z.B. Bewertung des Castro-Regimes) verfasst hatten.

Den Vpn wurde entweder mitgeteilt, dass die Autoren der Artikel entweder vom VI dazu angehalten wurde, die Position einzunehmen, oder die freie Wahl hatte.

Obwohl die Vpn wussten, dass ein Autor vom VI dazu eingeteilt worden waren, diese Position einzunehmen, schrieben sie die im Essay vertretene Einstellung dem Autor zu.



Bsp.2-15: Fundamentalener Attributionsfehler (Pietromonaco & Nisbett (1982):

Die Autoren teilten ihre Vpn in 2 Gruppen:

Gruppe A:

Es wurde eine abgekürzte Version der Studie von Darley und Batson (1973) präsentiert.

Gruppe B:

Es wurde eine vollständige Version mit Ergebnissen der Studie von Darley und Batson (1973) präsentiert.

Exkurs: Die Studie von Darley und Batson (1973) zur Hilfsbereitschaft:

Teilnehmer des Experiments waren Seminaristen, die entweder eine Rede über Jobs oder eine Rede über den *Barmherzigen Samariter* halten sollten.

Den Personen wurde durch einen Assistenten mitgeteilt, dass sie in einem anderen Gebäude eine Rede von 3-5 Minuten zum Thema halten sollte, die von einem anderen Assistenten aufgezeichnet würde.

Es gab folgende 3 Bedingungen:

(a) Grosse Eile:

Der Assistent teilte der Vp mit, dass sie sehr spät dran sei, der Assistent im anderen Gebäude sicher schon warte und die Vp sich beeilen möge.

(b) Mittlere Eile:

Der Assistent teilte der Vp mit, dass es Zeit sei, zu gehen und der andere Assistent wahrscheinlich schon bereit sei.

(c) Geringe Eile:

Der Assistent teilte der Vp mit, dass noch einige Minuten Zeit vorhanden seien, die Vp sich aber bereits in das andere Gebäude begeben möge und dort sicher nicht lange zu warten habe.

Um zum Seminarraum im anderen Gebäude zu gelangen, mussten die Personen eine Hof überqueren, in dem Darley und Batson einen schäbig aussehenden, zerlumpten Mann platziert hatten, mit gesenktem Kopf, die Augen geschlossen und ohne Bewegung.

Sobald die Vp vorbeikam, hustet der Mann zweimal und stöhnte, ohne den Kopf zu heben. Falls die Vp stehen blieb und fragte, ob alles in Ordnung sei, bzw. Hilfe anbot, schreckte der Mann auf und sagte:

Oh, danke (Husten)... Nein, es ist alles in Ordnung (Pause). Ich habe nur gerade Atmungsbeschwerden (Husten)... Der Doktor hat mir diese Pillen verschrieben. Ich habe gerade eine genommen... Wenn ich nur hier sitze und mich ein paar Minuten ausruhe, bin ich wieder o.k. ... Trotzdem, Danke für ihre Hilfe.

Ergebnis der Untersuchung von Darley und Batson (1973):

Je grösser die Eile der Vpn desto geringer die Bereitschaft zur Hilfe. Das Thema des Vortrages hatte keinen Einfluss auf die Hilfsbereitschaft.

Ergebnis der Untersuchung von Pietromonaco und Nisbett (1982):

Die Vpn beider Gruppen verwendeten zur Erklärung des Verhaltens die religiöse Einstellung der Person (Dispositionserklärung).

Der fundamentale Attributionsfehler kann zumindest teilweise mit Hilfe eines Salieneffekts erklärt werden, indem man annimmt, dass die urteilenden Personen ihre Aufmerksamkeit vor allem auf die handelnde Person richten und weniger auf die Situation.

Diese unterschiedliche Gewichtung von Person und Situation scheint kulturelle Wurzeln zu haben. So fanden Morris & Peng (1994) Unterschiede im Attributionsstil von Amerikanern und Chinesen dahingehend, dass Chinesen weniger dispositionelle und mehr situationale Erklärungen von Verhalten geben als Amerikaner (Bei Erklärungen im physikalischen Bereich wurden keine Unterschiede zwischen beiden Gruppen gefunden).

2.5.2.6 ASYMMETRIE VON ATTRIBUTIONEN

Ein Phänomen, welches sie gut mit Hilfe von Salienz erklären lässt betrifft die Asymmetrie der Erklärungen des Verhaltens einer Person durch externen Beobachter und durch die Person selbst: Externe Beobachter tendieren zu dispositionellen, personenbezogenen Erklärungen, während die Personen selbst eher zu situationsbezogenen Erklärungen tendieren.

So erklären Personen ihr eigenes aggressives Verhalten meist damit, dass sie von der anderen Person provoziert wurden, während Beobachter das Verhalten eher einer verstärkten Aggressivität der Person

zuschreiben. Die Erklärung mittels Salienz basiert auf der Idee, dass für die handelnde Person das Gegenüber bzw. die Situation hervorsticht, während für den Beobachter die handelnde Person herausragt. Die in diesem Beispiel demonstrierte Asymmetrie könnte jedoch auch mit Hilfe einer *Self-serving Bias* erklärt werden, indem die Person ihr eigenes Verhalten durch Hinweis auf die Situation zu entschuldigen sucht. Die beschriebene Asymmetrie der Attributionen kann jedoch auch in Situationen gefunden werden, für welche eine *Self-serving Biases* ausgeschlossen werden können.



Bsp.2-16: Asymmetrie von Attributionen (Ross, Amabile & Steinmetz, 1977):

Die Vpn wurden zufällig in 2 Gruppen eingeteilt.

Die Vpn der Gruppe A hatten die Aufgabe, sich 10 allgemeine Wissensfragen auszudenken, die so schwierig bzw. speziell waren, dass die Vpn aus Gruppe B diese vermutlich nicht beantworten konnten (z.B. Wie heisst der grösste Gletscher der Welt? [Malaspinagletscher]). Sodann wurden die Fragen gestellt.

Nach jeder Antwort durch Gruppe B wurde von Gruppe A angegeben, ob die Antwort korrekt oder falsch war. Im letzteren Fall wurde die korrekte Antwort gegeben.

Eine Einschätzung des Allgemeinwissens der Vpn der beiden Gruppe durch die Teilnehmer und durch unabhängige Beobachter ergab:

- Die Vpn der Gruppe A (welche die Fragen stellten) wurden sowohl von den Vpn der Gruppe B wie von den unabhängigen Beobachtern als gebildeter eingeschätzt wurden.
- Die Personen der Gruppe A schätzten sich selbst jedoch nicht als klüger ein.

Hier zeigt sich eine klare Asymmetrie der Attributionen, die sich jedoch nicht mit Hilfe einer *Self-Serving Bias* erklären lässt.

2.5.2.7 PROBLEME VON ERKLÄRUNG MITTELS SALIENZ- UND AUFMERKSAMKEITSEFFEKTEN

Im vorangehenden Abschnitt wurde auf die Bedeutung von Salienzeffekten für die Erklärung des eigenen Verhaltens hingewiesen. Im Speziellen wurde gezeigt, dass Personen oft irrelevante aber hervorstechende Information über- und relevante aber wenig auffällige Determinanten unterbewerten.

Erklärungen mit Hilfe von Salienz – obwohl sinnvoll – müssen als unvollständig betrachtet werden, da *Salienz* selbst als ein zu erklärendes Phänomen zu betrachten ist: Warum ist ein Ereignis für uns auffällig und ein anderes nicht? Wir haben gesehen, dass im Falle des fundamentalen Attributionsfehlers die wahrgenommene Salienz des Aktors

ein kulturspezifisches Phänomen darstellt: In westlichen Kulturen wird grosser Wert auf Individualität und persönlicher Verantwortung gelegt. Dies prägt unsere Tendenz, den Personenfaktor über und andere Einflüsse (wie die Situation) unter zu bewerten. In anderen Kulturen, in welchen der Individualität ein geringerer Stellwert eingeräumt wird, werden auch der Personenfaktor geringer bewertet. Hier scheint also die Salienz der Person, bzw. die Aufmerksamkeit, welche der Person entgegen gebracht wird, kein universelles Phänomen, sondern kulturell bedingt zu sein. In anderen Fällen – zum Beispiel im Falle von Attraktivitäts-Cues – scheint die differentielle Salienz biologisch mit bedingt zu sein: Frauen messen in praktisch allen untersuchten Kulturen körperlichen Attraktivitäts-Cues eine geringere Bedeutung bei als Männer.

Zusammenfassend lässt sich also bemerken, dass Erklärungen von Verhalten mit Hilfe von Konzepten wie (differentielle) Aufmerksamkeit, Auffälligkeit oder Ähnlichkeit nur partielle Erklärungen darstellen, wobei das Vorliegen einer differentiellen Salienz, Aufmerksamkeit oder Ähnlichkeit selbst der Erklärung bedarf.

2.6 Methodische Aspekte: Klassische Fehler und Paradoxien in der Beurteilung von Kontingenz und Kausalität

Im vorliegenden Abschnitt geht es Fehler und Paradoxien welche bei der Beurteilung von (kausalen) Zusammenhängen auftreten. Im Gegensatz zur Diskussion in den vorangegangenen Abschnitten handelt es sich hier methodische Fragen der Interpretation von statistischen Ergebnissen. Die hier behandelten Probleme können zu einer völlig falschen Interpretation von Daten führen. Daher sollte jede Person mit einer sozialwissenschaftlichen Ausbildung damit vertraut sein.

Wir beginnen mit einer Besprechung der Bedeutung der Abhängigkeit bzw. Unabhängigkeit von Variablen. Dem folgt eine Diskussion des Problems der Erschliessung von kausalen Zusammenhängen aufgrund von Abhängigkeiten (Korrelationen) zwischen Variablen. Hierauf diskutieren wir das so genannte *Simpson Paradoxon*. Anschliessend behandeln wir nochmals das Problem der Regression zum Mittel und besprechen das *Paradoxon von Lord*. Am Ende diskutieren wir *den ökologischen Fehlschluss* und zeigen, dass es sich hierbei um das Simpson Paradoxon in anderer Form handelt.

2.6.1 Die Bedeutung der Kenntnis von (Un-) Abhängigkeiten zwischen Variablen

Falls zwei Variablen miteinander assoziiert sind, so bedeutet dies, dass die Kenntnis des Wertes der Variable Information über den Wert auf der anderen Variable liefert. Dies ist in mindestens dreierlei Hinsicht von Bedeutung:

1. *Diagnose:* Die Messung einer Variable ermöglicht Aufschlüsse über eine andere – oft nicht direkt beobachtbare – Variable. Als Beispiel wären hier psychologische Tests zu nennen, welche Aufschluss über zugrunde liegende mentale Konstrukte (wie Intelligenz, Persönlichkeit, soziale Intelligenz) geben können.
2. *Vorhersage:* Zusammenhänge zwischen zwei Grössen ermöglichen Vorhersagen über den wahrscheinlichen Wert einer Variable aufgrund der Kenntnis des Wertes der anderen Grösse. Beispielsweise ergibt die Kenntnis der schulischen Leistung Aufschluss über den Studienerfolg.
3. *Kausalität:* Zusammenhänge zwischen Variablen liefern Hinweise für das mögliche Vorliegen (bei Vorhandensein einer Assoziation) oder über die mögliche Abwesenheit einer kausalen Beziehung. So kann der Zusammenhang zwischen bestimmten Persönlichkeitsmerkmalen und beruflichem Erfolg möglicherweise kausal interpretiert werden, wenn man annimmt, dass bestimmte Ausprägungen auf verschiedenen Persönlichkeitsdimensionen, wie Gewissenhaftigkeit, Extraversion, oder Verträglichkeit förderlich oder hinderlich für ein berufliches Weiterkommen sind.

Im wissenschaftlichen Kontext ist Information über die Abwesenheit von Zusammenhängen von grosser Wichtigkeit. Es eröffnet nämlich die Möglichkeit, Gruppen von Variablen isoliert zu studieren. Nur aufgrund Tatsache, dass nicht Alles mit Allem zusammenhängt ermöglicht eine Vereinfachung komplexer Systeme durch Zerlegung derselben in unabhängige Teilsysteme (die natürlich über Schnittstellen miteinander verbunden sein können). Nur auf diese Weise wird eine systematische Erforschung komplexer Systeme wie unser Gehirn überhaupt erst möglich.

2.6.2 Das Problem der Erschliessung kausaler Beziehungen aufgrund von Abhängigkeiten zwischen Variablen

Jede Studentin der Psychologie oder Pädagogik wird mit einer Variante des folgenden (völlig korrekten) Satzes konfrontiert:

Aus Korrelationen lassen sich keine kausalen Relationen erschliessen.

Hier stellt sich nun die Frage, warum eine derartige Schlussfolgerung unzulässig ist. Ein erster und sehr einfacher Grund besteht darin, dass bei einem Zusammenhang zwischen X und Y noch nicht die kausale Richtung geklärt ist: $X \rightarrow Y$ oder $X \leftarrow Y$? Oft ist es jedoch der Fall, dass die kausale Richtung aufgrund inhaltlicher Gründe oder aufgrund der zeitlichen Aufeinanderfolge erschlossen werden kann.

Die interessanteren und in der Praxis relevanteren Fälle für die Unmöglichkeit des Erschliessens kausaler Beziehungen aufgrund von Korrelationen sind in Abb. 2-2 skizziert. Die mit X und Y bezeichneten Kästchen symbolisieren beobachtete Variablen. Die mit Z bezeichneten Kreise repräsentieren nicht beobachtete (latente) Variablen. Die Pfeile symbolisieren kausale Einflüsse, wobei die gestrichelten Pfeile andeuten, dass möglicherweise eine kausale Relation vorliegt. Der Bogen mit den Doppelpfeilen stellt einen so genannten Kovarianzbogen dar, d.h. er symbolisiert das Vorliegen einer Kovarianz (oder Korrelation) zwischen den beiden Variablen.

In Abb. 2-2 (a) hat die nicht gemessene Variable Z einen Einfluss auf beiden Grössen X und Y . Dieser Einfluss auf beide Variablen kann möglicherweise die gesamte Korrelation zwischen X und Y erklären. Ob zusätzlich auch noch eine kausale Beziehung zwischen X und Y vorliegt, muss separat geprüft werden.

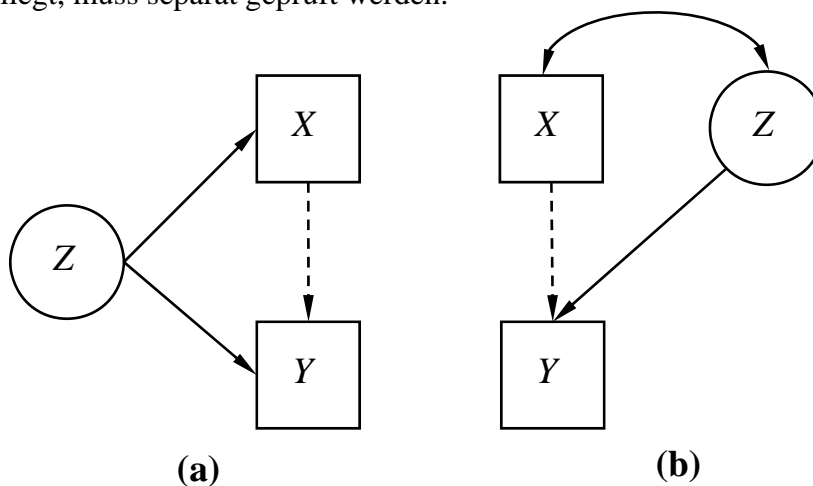


Abb. 2-2: Das Problem der kausalen Interpretation einer Beziehung zwischen zwei Variablen: (a) Beobachtete Abhängigkeit aufgrund einer gemeinsamen versteckten Ursache; (b) Konfundierende Variable

Um eine kausale Beziehung zwischen zwei Variablen X und Y zweifelsfrei zu etablieren muss daher ausgeschlossen werden, dass die beobachtete Assoziation zwischen zwei Variablen nicht durch eine dritte ungemessene Grösse, welche beide Variablen beeinflusst, bedingt ist. Falls sich die beobachtete Assoziation zwischen X und Y nicht aufgrund eines direkten kausalen Einflusses $X \rightarrow Y$ ergibt, sondern durch eine dritte gemeinsame Ursache bedingt ist, so spricht man von einem *Scheineffekt*.



Bsp.2-17: Scheineffekt:

Im Burgenland wurde in den 50er Jahren eine Korrelation zwischen der Anzahl der Störche und der Anzahl der Geburten beobachtet.

Nach derzeitigem Wissensstand gibt es keine kausale Beziehung zwischen Störchen und Geburten.

Die Erklärung ergibt sich aufgrund der Tatsache, dass nach dem Krieg neue Familien gegründet und die zerstörten Häuser wieder aufgebaut bzw. neue gebaut wurden. Die Familiengründungen führten zur Erhöhung der Zahl der Kinder und die neu aufgebauten Häuser boten Nistplätze für die Störche, was deren Anzahl erhöhte (Abb. 2-3).

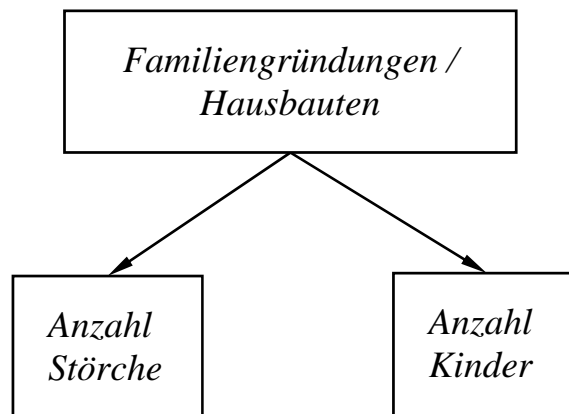


Abb. 2-3: Erklärung des Zusammenhangs zwischen der Zunahme der Störche und der Anzahl Kinder.

Die Idee, wonach eine latente oder mehrere latente Größen die Korrelation zwischen zwei oder mehreren Variablen erklärt, bildet die Basis des faktorenanalytischen Modells der menschlichen Intelligenz oder Persönlichkeit. Dieser Ansatz geht auf Charles Spearman (1863-1945) zurück, der die Korrelation zwischen den Ergebnissen verschiedener Leistungstests durch einen latenten gemeinsamen Faktor g (general intelligence) erklärte (Abb. 2-4).

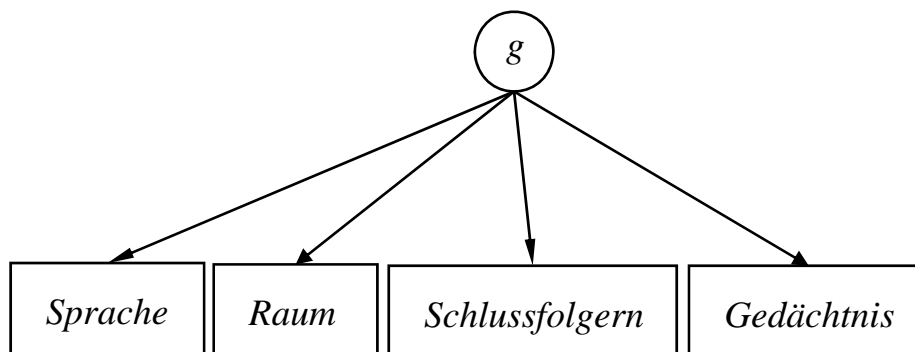


Abb. 2-4: Erklärung der Korrelation zwischen den Ergebnissen verschiedener Tests mit Hilfe eines latenten gemeinsamen Faktors.

Ein zweites Hindernis, welches sich einer Erschließung von kausalen Beziehungen aufgrund von Korrelationen zwischen Variablen entgegenstellt, bildet das Problem der *Konfundierung*.



Konzept 2-4: Konfundierende Variable:

Unter einer *konfundierenden Variable Z* versteht man eine Variable, welche mit der unabhängigen Variablen *X* korreliert ist und einen Einfluss auf die abhängige Variable *Y* ausübt.

In Abb. 2-2 (b) stellt die latente Variable *Z* eine konfundierende Variable dar, welche die Interpretation des Zusammenhanges von *X* und *Y* als kausalen Effekt beeinflusst.



Bsp.2-18: Vollmond und Verkehrsunfälle:

In ihrer Meta-analyse zur Wirkung des Vollmondes stellten Rotton & Kelly (1985) fest, dass es keinerlei relevanter Effekte des Vollmondes auf das menschliche Verhalten gebe. Am Ende des Artikels geben sie – entgegen der allgemeinen Gepflogenheit, wonach weitere Forschung gewünscht wird – ihrem Wunsche Ausdruck, dass ihre Analyse die Forschung zur Wirkung des Vollmondes beenden werde (der natürlich nicht erfüllt wurde, obwohl die neueren Forschungen zum gleichen Ergebnis kommen [siehe Lilienfeld, Lynn, Rusco, & Beyerstein, 2010]).

In einer der untersuchten Studien ergab sich ein Effekt des Vollmondes auf die Anzahl Autounfälle. Dieser Effekt kam dadurch zustande, dass die beobachteten Vollmondphasen auf Wochenende fielen, wo typischerweise das Verkehrsaufkommen erhöht ist. Wird der Einfluss des Wochenzyklus kontrolliert, so ergibt sich keinerlei Effekt des Vollmondes auf die Anzahl der Autounfälle.

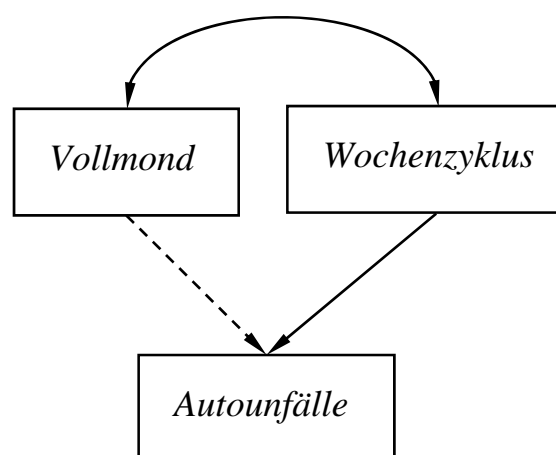


Abb. 2-5: Die Konfundierung der Variablen *Vollmond* und *Wochenzyklus* führt zu einer fehlerhaften Interpretation eines Einflusses des *Vollmonds* auf die Anzahl *Autounfälle*.

Um eine kausale Beziehung zwischen zwei Variablen einwandfrei festzustellen muss daher der Effekt möglicher konfundierender Grös-

sen ausgeschlossen werden. Auf den ersten Blick scheint es daher praktisch unmöglich zu sein, kausale Effekte einwandfrei festzustellen. Hier hilft nun ein Erkenntnis, welches den meisten Studenten der Psychologie ebenfalls geläufig sein sollte:

Kausale Effekte können nur mit Hilfe eines Experiments mit Zufallszuordnung zweifelsfrei festgestellt werden.

Die Bedeutung dieser Aussage soll im Folgenden näher untersucht werden.

2.6.2.1 ELIMINATION VON KONFUNDIERUNG DURCH RANDOMISIERUNG UND PARALLELISIERUNG

Wir gehen von folgender Situation aus: Die unabhängige Variable sei eine neue Behandlungsmethode, deren Wirksamkeit getestet werden soll. Der *Goldstandard* der Prüfung der Methode besteht im Vergleich der Methode mit der bisher besten Methode oder einer anderen Kontrollbedingung (z.B. einer Placebo-Bedingung), wobei idealerweise 2 Bedingungen erfüllt sein sollten:

- Die Zuordnung der Personen zu den Bedingungen ist zufällig.
- Weder die Klientin noch der Therapeut weiß, welche Methode angewendet wird (Doppelblindversuch).

Um zu erkennen, was diese beiden Bedingungen bewirken, betrachte man Abb. 2-6.

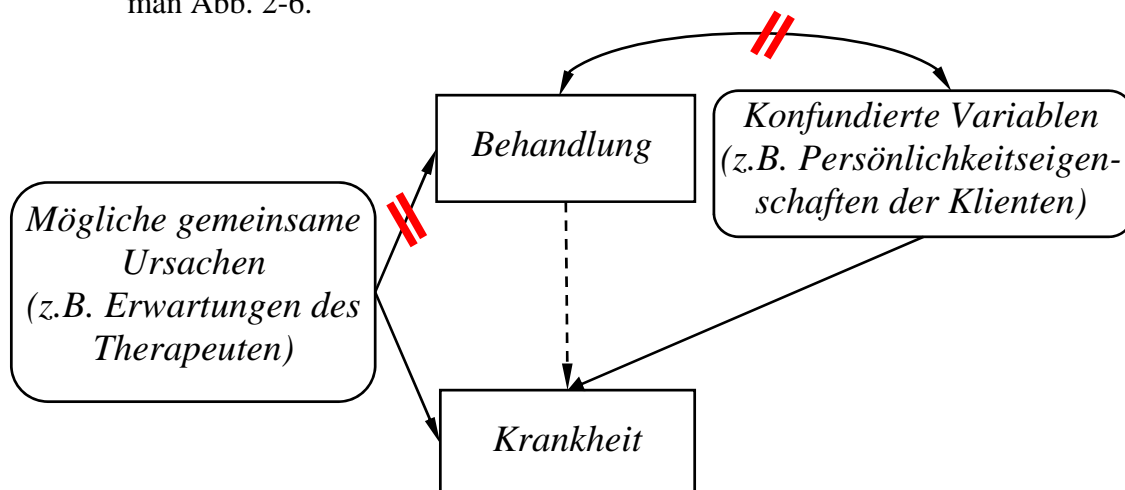


Abb. 2-6: Effekte der Randomisierung und Kontrolle von Versuchsleitereffekten: Die roten Doppelstriche symbolisieren die Unterbrechung der Verbindungen.

Die Kontrolle der Behandlungsbedingung durch den Untersucher bewirkt, dass der kausale Effekt möglicher latenter Variablen auf die Variable *Behandlung* ausgeschaltet wird.

Zentral hierbei ist, dass einzig die *Behandlung* durch den Untersucher variiert wird und nicht auch andere Größen, welche sowohl die *Behandlung* als auch die *Krankheit* beeinflussen können. Aus diesem

Grund müssen Erwartungen des Versuchsleiters ausgeschaltet werden, da diese sowohl die Krankheit als auch die Behandlung beeinflussen können.

Die Zufallszuordnung der Patienten zu den Behandlungsmethoden bewirkt, dass die Korrelation zwischen der Behandlung und möglichen konfundierenden Variablen (wie Persönlichkeitseigenschaften) mit grosser Wahrscheinlichkeit eliminiert wird.

Sobald diese beiden Arten von Verbindungen zwischen der Zielursache und den anderen Variablen unterbrochen wurden, kann der kausale Effekt *Behandlung* → *Krankheit* verzerrungsfrei geschätzt werden, *auch wenn die anderen Variablen noch einen Einfluss auf die Krankheit (=Outcome-Variable) ausüben.*

Kontrastiert man diese Situation mit jener von Beobachtungsstudien, so zeigt sich der grosse Nachteil der letzteren: Aufgrund der fehlenden Randomisierung kann die Korrelation zwischen konfundierenden Variablen und der Zielursache nicht einwandfrei ausgeschlossen werden.

Die Zufallszuordnung von Klienten zu den Behandlungsbedingungen eliminiert mit grosser Wahrscheinlichkeit die Korrelationen zwischen den Störvariablen und der zu untersuchenden Behandlungsmethode. Aufgrund der Randomisierung werden nämlich die unterschiedlichen Ausprägungen der konfundierten Variablen ungefähr gleich auf die verschiedenen Behandlungsbedingungen verteilt. Das Problem ist hierbei, dass man rein zufällig eine ungleiche Verteilung der Merkmale in den Bedingungen erhalten kann und somit die Korrelation substantiell ist (wobei bei grossen Stichproben die Wahrscheinlichkeit einer substantiellen Korrelation praktisch Null ist).

Hier kommt nun die *Parallelisierung* ins Spiel: Falls bekannt ist, dass ein Merkmal (z.B. Geschlecht) einen Einfluss auf den Krankheitsverlauf bzw. – allgemeiner – auf den Outcome hat, so wird man es nicht dem Zufall überlassen, ob die Korrelation dieses Merkmals mit der Behandlung Null wird, sondern man wird die Korrelation *per Design* auf Null fixieren, indem man die verschiedenen Merkmale gleich auf die verschiedenen Behandlungsmethoden verteilt. Diese Methode nennt sich *Parallelisierung* oder *Ausbalancierung*.

Nachdem wir nun die Vorzüge des randomisierten Experiments erkannt haben, stellt sich sofort die folgende Frage:



Frage:

Kann mit Hilfe eines randomisierten Experiments zweifelsfrei der kausale Einfluss einer Variablen auf eine andere festgestellt werden?

Die Antwort auf diese Frage lautet eindeutig: *Nein!* Es gibt hierfür im Wesentlichen zwei Gründe:

1. Wie bereits erwähnt, kann durch Randomisierung nur mit grosser Wahrscheinlichkeit die Elimination aller Korrelationen erreicht werden.
2. Es kann niemals mit absoluter Sicherheit ausgeschlossen werden, dass der Untersucher zusammen mit der Ursachenvariablen noch eine nicht erkannte Störvariable variiert, die zu einer Verzerrung des Ergebnisses führt.

Zusammenfassend lässt sich also sagen. Obwohl das randomisierte Experiment die beste aller bekannten Methoden zur Prüfung kausaler Wirkungen darstellt, bietet es keine perfekte Sicherheit, kausale Fehlschlüsse zu vermeiden.

2.6.2.2 DIE BEDEUTUNG VON ASSOZIATIONEN FÜR DIE PRÜFUNG VON KAUSALSTRUKTUREN

Betrachtet man die bisher gegebene Analyse, sind die Aussichten, die Aufdeckung kausaler Zusammenhänge betreffend ziemlich düster: Aufgrund von Korrelationen kann man nicht auf eine Verursachung schliessen und auch mit Hilfe der besten uns bekannten Methode zur Prüfung von kausalen Einflüssen lassen sich kausale Relationen nicht zweifelsfrei etablieren.

Man mag sich nun fragen, was der Grund für diese »Widerspenstigkeit« kausaler Beziehungen gegen ihre Erforschung ist. Die korrekte Antwort stammt meines Erachtens von dem Philosophen und Wissenschaftstheoretiker Karl Popper (1902–1994). Dieser erkannte, dass das Problem der Erschliessung kausaler Relationen ein Spezialfall eines generelleren Problems darstellt (Popper, 1984): des Induktionsproblems.



Konzept 2-5: Das Induktionsproblem:

Das *Induktionsproblem* besteht in der Frage, ob und wie sich so genannte induktive Schlussfolgerungen rechtfertigen lassen.

Bei induktiven Schlüssen handelt es sich im Wesentlichen um Schlüsse vom Besonderen auf das Allgemeine, d.h. von singulären Beobachtungen auf allgemeine Zusammenhänge.

Im Falle der Erschliessung kausaler Zusammenhänge handelt es sich genau um eine derartige induktive Schlussfolgerung, denn man möchte aufgrund konkreter Beobachtungen auf einen allgemeinen Zusammenhang schliessen (nämlich auf den kausalen Zusammenhang).

Popper (1984) behauptet nun – nicht ganz unbescheiden:

Selbstverständlich kann ich mich irren, aber ich glaube, ein sehr wichtiges philosophisches Problem gelöst zu haben: das Problem der Induktion (Popper, 1984, Seite 1).

Die von ihm vorgeschlagene Lösung ist ebenso elegant wie einfach:

1. Induktive Schlüsse lassen sich nicht rechtfertigen.

2. Es ist jedoch möglich, Aussagen über allgemeine Zusammenhänge aufgrund von empirischen Daten rigoros zu testen. Falls die Aussagen alle Test bestehen, so werden sie als vorläufig akzeptabel anerkannt, bis weitere Befunde deren Gültigkeit widerlegen.

Dies entspricht der Art und Weise wie heute in den (Sozial-) Wissenschaften kausale Zusammenhänge erforscht werden:

1. Aufgrund theoretischer Überlegungen wird ein kausales Modell postuliert.
2. Die Gültigkeit dieses Modell wird anhand konkreter Daten evaluiert.

Im Zuge dieser Prüfung werden nun Abhängigkeiten bzw. Korrelationen zwischen Variablen relevant. Es wird nämlich getestet, ob ein vorgeschlagenes Kausalmodell, die Struktur der Korrelationen zwischen den Variablen vorhersagen kann. Ist dies der Fall und ist das Modell plausibel, d.h. es widerspricht nicht anerkannten Theorien und die zeitlich Ordnung der Variablen ist kompatibel mit der Kausalrichtung der Effekte, so gilt das Modell als nicht verworfen.

Damit ein kausales Modell aber als wissenschaftlich wertvoll betrachtet werden kann, muss noch eine Zusatzbedingung erfüllt sein: Es müssen alle *statistisch ununterscheidbaren Modelle* ausgeschlossen werden können. Hierbei handelt es sich – wie der Name schon sagt – um Modelle, welche mit statistischen Mitteln nicht unterschieden werden können. Daher braucht es inhaltliche Überlegungen, um unplausible Modelle auszuschliessen.



Bsp.2-19: Statistisch ununterscheidbare Modelle:

Das klassische Beispiel für ein Modell mit vielen statistisch ununterscheidbaren Alternativmodellen bildet das so genannte Drei-Variablen Mediatormodell, in welchem einerseits ein direkter Effekt einer Ursachenvariable X auf eine Ergebnisvariable postuliert wird und andererseits ein indirekter Effekt über eine Mediatorvariable M (Abb. 2-7a).

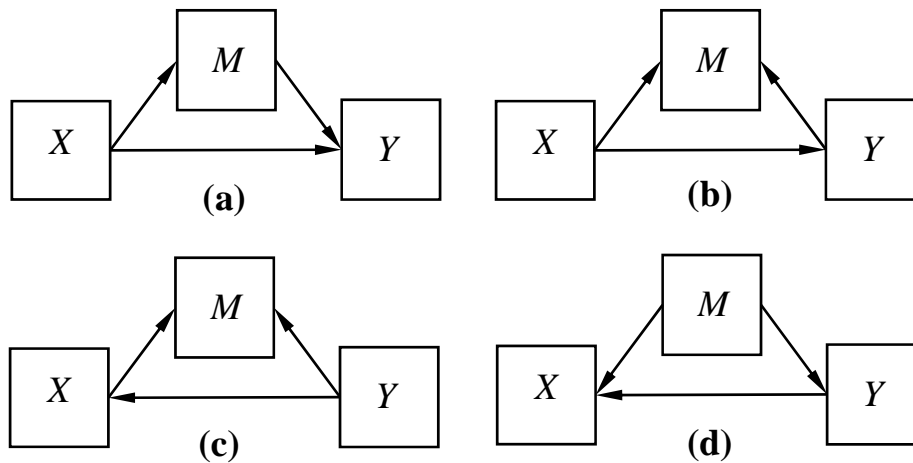


Abb. 2-7: *Drei-Variablen Mediatormodell (a) und .statistische ununterscheidbare Modelle (b) – (d). Die gezeigten Modelle bilden nur eine Auswahl aus der gesamten Menge der statistisch ununterscheidbaren Modelle.*

Die statistisch ununterscheidbaren Modelle ergeben sich indem die Pfeile beliebig verdreht werden, wobei jedoch kein Zyklus entstehen darf.

Zusammenfassend kann man nun das Folgende sagen:

1. Kausale Relationen lassen sich niemals zweifelsfrei etablieren. Damit teilen sie das Schicksal alle naturwissenschaftlichen Aussagen über allgemeine Zusammenhänge.
2. Kausale Strukturen und Modelle lassen sich überprüfen. Hierbei spielen Korrelationen eine wichtige Rolle.

2.6.3 Das Simpson Paradoxon

Die Bedeutung von konfundierenden Variablen für die Interpretation von Zusammenhängen wird in eindrücklicher Weise durch das so genannte *Paradoxon von Simpson* verdeutlicht.



Konzept 2-6: *Simpson Paradoxon:*

Das *Simpson Paradoxon* besteht in der Tatsache, dass die Relation zwischen zwei kategoriellen Variablen durch Einbeziehung einer dritten kategoriellen Variable in der Richtung umgedreht werden kann.

Hier ein reales Beispiel (New York Times Magazine, 11 März, 1979):



Bsp.2-20: Simpson-Pardoxon: Todesurteile in Florida zwischen 1972-1979:

Wir betrachten die Anzahl Todesurteile in Mordprozessen im Bundesstaat Florida, wobei wir der Frage nachgehen, ob Schwarze für Morde eher zum Tode verurteilt werden als Weisse. Tab. 2-4 zeigt, wie viele Todesstrafen für Weisse und Schwarze ausgesprochen wurden:

Gemäss Tab. 2-4 scheint es keine rassistischen Vorurteile bezüglich der Verhängung der Todesstrafe zu geben, denn von den verurteilten Weissen wurden 3.2% zum Tode verurteilt, während für die Schwarzen der Anteil 2.4% betrug. Dementsprechend ist auch Yules Q negativ, was auf einen leicht negativen Zusammenhang zwischen Schwarze und Todesstrafe hinweist.

Tab. 2-4: Todesstrafen für Delinquenten mit weisser und schwarzer Hautfarbe:

Farbe Delinquent	Todesstrafe		Σ	%Ja	Yules Q
	Ja	Nein			
schwarz	59	2448	2507	2.4	-0.16
weiss	72	2185	2257	3.2	
Σ	131	4633	4764		

Wir beziehen nun als weitere Variable die Farbe des Opfers mit ein. Tab. 2-5 zeigt die relevanten Daten.

Tab. 2-5: Todesstrafen für Delinquenten mit weisser und schwarzer Hautfarbe unter Berücksichtigung der Hautfarbe des Opfers:

Farbe Opfer	Farbe Delinquent	Todesstrafe		Σ	%Ja	Yules Q
		Ja	Nein			
schwarz	schwarz	11	2209	2220	0.05	1.00
	weiss	0	111	111	0.00	
weiss	schwarz	48	239	287	16.70	0.71
	weiss	72	2074	2146	3.40	
Σ		131	4633	4764		

Hier ergibt sich erstaunlicher Weise ein völlig anderes Bild:

War das Opfer von schwarzer Hautfarbe, so wurden eher Delinquenten mit schwarzer als mit weisser Hautfarbe zum Tode verurteilt: 0.05% vs. 0%.

War das Opfer von weisser Hautfarbe, so wurden ebenfalls eher Delinquenten mit schwarzer als mit weisser Hautfarbe zum Tode verurteilt: 16.7% vs. 3.4%.

In beiden Teiltabellen (für beide Arten von Opfern) ergibt sich nun ein starker positiver Zusammenhang (siehe Yules Q) zwischen Todesstrafe und Farbe, indem an schwarzen Verurteilten die Todesstrafe eher vollzogen wird.

Wie kommt es zu dieser Diskrepanz? Diese ergibt sich aus den beiden folgenden Tatsachen:

1. Für einen Mord an einer Schwarzen wird seltener die Todesstrafe ausgesprochen als bei einem Mord an einer Weissen.
2. Schwarze töten eher Schwarze als Weisse. Für Weisse ist es umgekehrt.

Ignoriert man daher die Hautfarbe des Opfers, so scheint es, dass Schwarze milder behandelt würden als Weisse (d.h. es werden insgesamt weniger Todesstrafen gegen sie ausgesprochen).

Dieser Eindruck basiert rein auf der Tatsache, dass Schwarze eher Schwarze ermorden und der Mord an einer Schwarzen weniger oft mit dem Tod bestraft wird.

Schlussfolgerung:

Die Entfernung einer Variable (indem die Werte über die verschiedenen Stufen der Variable summiert werden) kann zu einer völlig falschen Einschätzung führen.

Hier ein weiteres Beispiel, welches einer realen Situation nachempfunden wurde:



Bsp.2-21: Simpson-Pardoxon: Wirkung einer neuen Therapie:

Eine PsychologIn hat eine neue Paartherapie entwickelt. Sie testet nun diese gegenüber der herkömmlichen Therapie in den beiden Ortschaften Ziegenwiler und Kuhdorf. Hierbei erhält er die folgenden Ergebnisse:

Ortschaft	Therapie	Erfolgreich		Σ	% Erfolg	Yules Q
		Ja	Nein			
Ziegenwiler	Neu	20	180	200	10%	0.36
	Alt	5	95	100	5%	
Kuhdorf	Neu	90	10	100	90%	0.50
	Alt	150	50	200	75%	
Σ		265	335	600		

Die neue Paartherapie erweist sich also gegenüber der herkömmlichen Therapie sowohl in Ziegenwiler als auch in Kuhdorf als erfolgreicher.

Begeistert sendet unsere PsychologIn einen Bericht an die Zeitungen der beiden Ortschaften. Der Chefredakteur des Ziegenwiler Anzeigers instruiert den Voluntär, der über die Ergebnisse einen Bericht schreiben soll: »Bringen Sie nur eine Tabelle, alles andere verwirrt die LeserInnen nur«.

Gesagt, getan, der Voluntär addiert die Statistiken der beiden Ortschaften und erhält:

Therapie	Erfolgreich		Σ	% Erfolg	Yules Q
	Ja	Nein			
Neu	110	190	300	37%	-0.30
Alt	155	145	300	52%	
Σ	265	335	600		

Offensichtlich ist die alte Therapie eindeutig erfolgreicher als die neue. Der Voluntär schreibt hierauf einen wütenden Artikel mit dem Titel: *Angeblich neue Wundertherapie entdeckt: Die statistischen Tricks der Psychologenlobby*«.

Wie kommt dieses Ergebnis zustande?

Offensichtlich wirken beide Therapien in Kuhdorf viel besser als in Ziegenwiler. Weiters wird in Ziegenwiler vermehrt die neue und in Kuhdorf eher die alte Therapie angewendet.

Summiert man nun über beide Dörfer, so entsteht der *irrig* Effekt, dass die alte Therapie besser wäre als die neue.

Hier ein letztes Beispiel, welches ebenfalls eine reale Situation nachstellt:



Bsp.2-22: Simpson-Pardoxon: Diskriminierung von Frauen:

Eine Pädagogin untersucht in ihrer Bachelorarbeit den Studienerfolg von Männer und Frauen an der Universität Freecastle, für die beiden Studienfächer Sozialarbeit und Psychologie. Sie erhält die folgenden Daten:

Fach	Geschlecht	Erfolgreich		Σ	% Erfolg	Yules Q
		Ja	Nein			
Sozialarbeit	Mann	127	35	162	78%	-0.20
	Frau	27	5	32	84%	
Psychologie	Mann	17	42	59	29%	-0.14
	Frau	92	170	262	35%	
Σ		263	252	515		

Offensichtlich sind Frauen sowohl in der Sozialarbeit als auch in der Psychologie erfolgreicher als Männer.

Die Frauenbeauftragte der Universität möchte die Ergebnisse in der Universitätszeitung *ReflectUni* veröffentlichen. Um die Darstellung zu vereinfachen, lässt sie die Fächer weg und addiert die Werte über die Fächer hinweg. Sie erhält die folgende Tabelle:

Geschlecht	Erfolgreich		Σ	% Erfolg	Yules Q
	Ja	Nein			
Mann	144	77	221	65%	0.47
Frau	119	175	294	40%	
Σ	263	252	515		

Offensichtlich sind Männer insgesamt erfolgreicher als Frauen. Die Frauenbeauftragte verfasst hierauf einen zornigen Artikel mit dem Titel: *Diskriminierung von Frauen an der Universität Freecastle*.

Der Grund für die Unterschied ist ähnlich wie im letzten Beispiel: Das Studium der Sozialarbeit ist einfacher als jenes der Psychologie.

Es studieren aber mehr Frauen Psychologie als Sozialarbeit während für die Männer das Umgekehrte der Fall ist.

Addiert man daher die Werte aus den beiden Teiltabellen über die Fächer hinweg, so ergibt sich die falsche Schlussfolgerung, dass Frauen schlechter abschneiden als Männer.

Die drei Beispiele demonstrieren eindrücklich, dass sich durch Weglassen einer Variable (indem die Daten über diese summiert werden), eine Schlussfolgerung ergeben kann, welche genau entgegengesetzt zu jener ist, welche sich bei Beachtung aller Variablen ergibt. Daher dürfen die Daten aus einer Tabelle nicht ohne weiteres zusammengefasst werden.

Im Folgenden wird sich zeigen, dass das gleiche Phänomen auch im Regressionskontext beobachtet werden kann, auch wenn es hier unter einem anderen Namen fungiert.

2.6.4 Der ökologische Fehlschluss

Der ökologische Fehlschluss stellt das Gegenstück zum Simpson Paradoxon im Regressionskontext dar.



Konzept 2-7: Ökologischer Fehlschluss (falsche Korrelation):

Gegeben:

Eine geschichtete Stichprobe, d.h. die gesamte Stichprobe besteht aus mehreren Teilstichproben, die jeweils aus verschiedenen Teilpopulationen entnommen wurden.

Der *ökologische Fehlschluss* ergibt sich bei Ignorieren der Clusterstruktur der Stichprobe, indem die Regressionskoeffizienten in jeder Teilstichprobe ein anderes Vorzeichen aufweisen als der Regressionskoeffizient, welcher sich aufgrund der Gesamtstichprobe ergibt.

Das folgende Beispiel illustriert das Problem.



Bsp.2-23: Ökologischer Fehlschluss:

Untersucht wird der Zusammenhang zwischen Leistungsfähigkeit und Aggressionsneigung von Schülern. Hierzu werden drei Klassen einer Schule untersucht.

In Abb. 2-8 sind die Populationen, aus denen die Daten gezogen wurden, durch die drei Ellipsen, welche die drei Klassen repräsentieren, dargestellt.

Die grauen Symbole repräsentieren die einzelnen Schüler aus den drei Klassen.

Die rote gepunktet Regressionslinie:

$$Aggression = 4.52 + 0.75 \cdot Leistung$$

täuscht den falschen Eindruck eines (hoch signifikanten) positiven Zusammenhanges zwischen *Aggression* und *Leistung* vor.

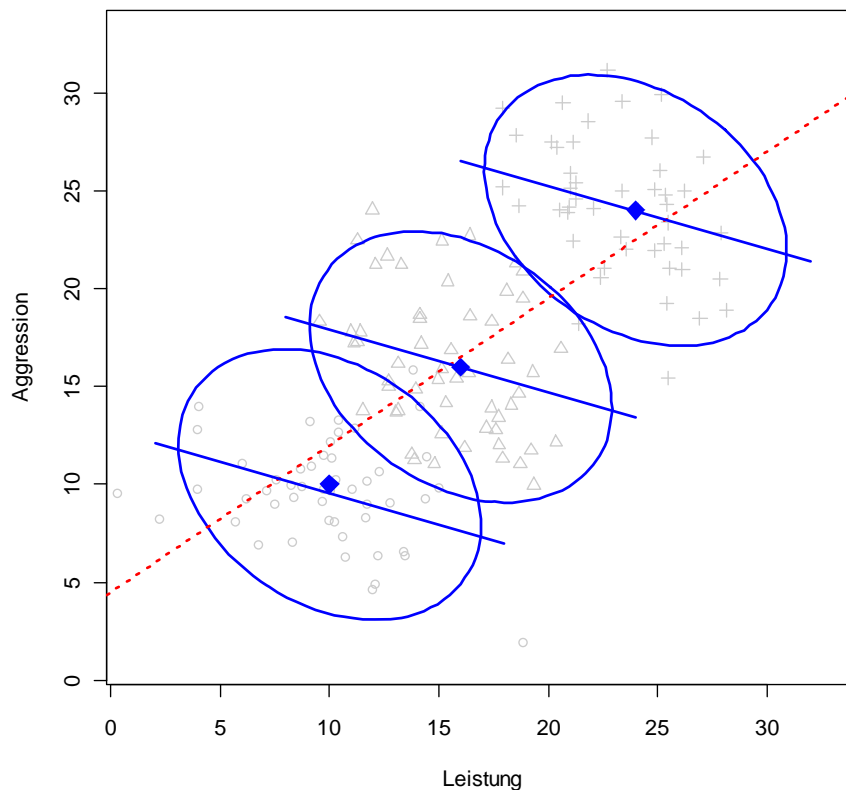


Abb. 2-8: Veranschaulichung des ökologischen Fehlschlusses. Die Steigungsparameter sind in allen drei Gruppen negativ (blaue Linien). Der Steigungsparameter über alle Daten hinweg (unter Ignorierung der Clusterstruktur) ist hingegen positiv (Rote gepunktete Linie).

Bezieht man die Klassen mit ein, so ergibt sich ein signifikanter negativer Zusammenhang.

Interpretation:

Die besondere Datenlage ergibt sich, indem die Aggressivität für Klassen mit der Leistung zunimmt (positiver Effekt der Cluster), während die Aggressivität für die einzelnen Schüler innerhalb der Klassen mit zunehmender Leistung abnimmt:

Der Effekt der Variable *Aggression* ist auf der Ebene der Klassen positiv und auf der Ebene der einzelnen Schüler innerhalb der Klassen negativ.

Bsp.2-23 demonstriert, dass – ähnlich wie beim Paradoxon von Simpson – das Ignorieren einer Variablen zu einer fehlerhaften Interpretation führt.

Nach dieser Darstellung von fehlerhaften Urteilen, welche sich durch Weglassen relevanter Variablen ergeben, kommen wir zu guter Letzt noch einmal zum Problem der Regression zum Mittel.

2.6.5 Regression zum Mittel und Lords Paradoxon

Abb. 2-9 illustriert die statistische Bedeutung der Regression zum Mittel. Die blaue Ellipse zeigt die 95% Konfidenzregion einer Normalverteilung bei einer Korrelation von $\rho_{xy} = 0.5$. Hierbei handelt es sich um jene Punkte gleicher Dichte auf der dreidimensionalen Normalverteilungskurve, innerhalb der sich 95% des Volumens unter der Normalverteilungskurve befindet.

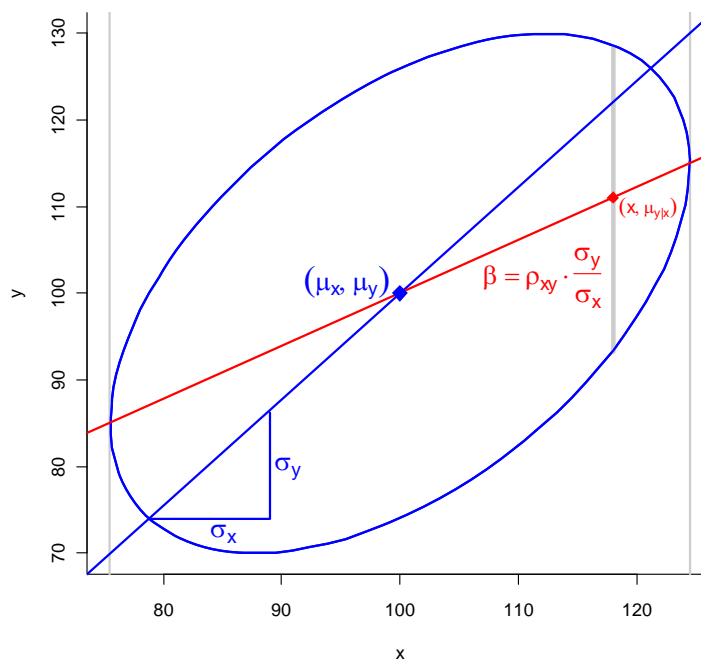


Abb. 2-9: Veranschaulichung der Regression zum Mittel: Die Steigung der Regressionsgerade (rote Linie) ist immer kleiner gleich jener der Hauptachse (blaue Linie) und sie schneidet die ellipsenförmige Region gleicher Dichte (blaue Ellipse) an den Tangenten (graue senkrechte Linien).

Die Regression zum Mittel zeigt sich darin, dass der Betrag der Steigung der Regressionsgerade (rote Linie) immer kleiner ist als jener der Steigung der Hauptachse der Ellipse (blaue Linie).

Dies ergibt sich direkt aufgrund der Formel für die Steigung der Regressionsgeraden (siehe Abb. 2-9): Da die Steigung der Hauptachse dem Quotienten σ_y/σ_x entspricht und der Betrag $|\rho_{xy}|$ des Korrelationskoeffizienten immer kleiner gleich 1 ist, kann der Betrag der Steigung der Regressionsgerade niemals jenen der Hauptachse überschreiten und nur falls gilt $|\rho_{xy}|=1$, so sind die beiden Steigungsparameter identisch.

Abb. 2-9 zeigt eine wichtige Besonderheit: Die Regressionsgerade schneidet die Ellipse genau an den vertikalen Tangenten. Dies kommt

dadurch zustande, dass *jeder Punkt auf der Regressionsgerade dem Mittelwert $\mu_{y|x}$ der y-Werte bei dem gegebenen Wert von x entspricht* (vgl. die dicke graue vertikale Linie in Abb. 2-9). Der (bedingte) Mittelwert $\mu_{y|x}$ liefert – bei Vorliegen einer bivariaten Normalverteilung – die beste Schätzung für den Wert der Variable Y bei gegebenem Wert $X = x$. Die Tatsache, dass $\mu_{y|x}$ geringer ist als der zugehörige Wert auf der Hauptachse, ergibt sich aufgrund der nicht perfekten Korrelation zwischen den beiden Grössen. Falls die Korrelation gleich 0 ist gilt: $\mu_{y|x} = \mu_y$, für alle Werte von X, d.h. die beste Schätzung für Y ist dann der (unbedingte) Mittelwert μ_y (= Gesamtmittelwert der Y), gleichgültig, welcher Wert $X = x$ beobachtet wurde.

Die Regression zum Mittel kann bei Verwendung der Kovarianzanalyse zur Kontrolle von vorgegebenen Unterschieden zu einem fehlerhaften Urteil bezüglich des zu untersuchenden Effekts führen. Dieses Phänomen wurde nach dem bekannten Psychometriker Frederick Lord, der es im Jahre 1967 beschrieb, als Lords Paradoxon bezeichnet.



Konzept 2-8: Lords Paradoxon:

Lords Paradoxon besteht in der Tatsache, dass man bei Verwendung einer Kovarianzanalyse zur Analyse des Effekts einer Intervention, wobei Ausgangsunterschiede der verschiedenen Gruppen mit Hilfe der Kovarianzanalyse kontrolliert werden, zu einem anderen Ergebnis gelangt als mit Hilfe eines *t*-Tests der Differenz aus der Vorher- und Nachhermessung.

Diese etwas abstrakte Darstellung lässt sich anhand eines einfachen Beispiels verdeutlichen.



Bsp.2-24: Lords Paradoxon:

Gegeben:

Zwei Gruppen von Schülern, die sich dadurch unterschieden, dass die Schüler der ersten Gruppe weiterhin die Sekundarschule besuchen und erst später ins Gymnasium überwechseln, während die zweite Gruppe sofort ins Gymnasium eintritt.

Vor dem Zeitpunkt des Eintritts ins Gymnasium (der zweiten Gruppe) wird für beide Gruppen ein Sprachtest [*Test (vorher)*] absolviert.

Weiters absolvieren beide Gruppen ein Jahr später den gleichen Sprachtest [*Test (nachher)*].

Wie aus Abb. 2-10 hervorgeht, unterscheiden sich die beiden Gruppen im Vor- und Nachtest um exakt den gleichen Betrag (50 Punkte).

Dies bedeutet, dass das Jahr im Gymnasium gegenüber der Sekundarschule zu keiner Verbesserung geführt hat. Insgesamt hat sich keiner der beiden Gruppen verbessert.

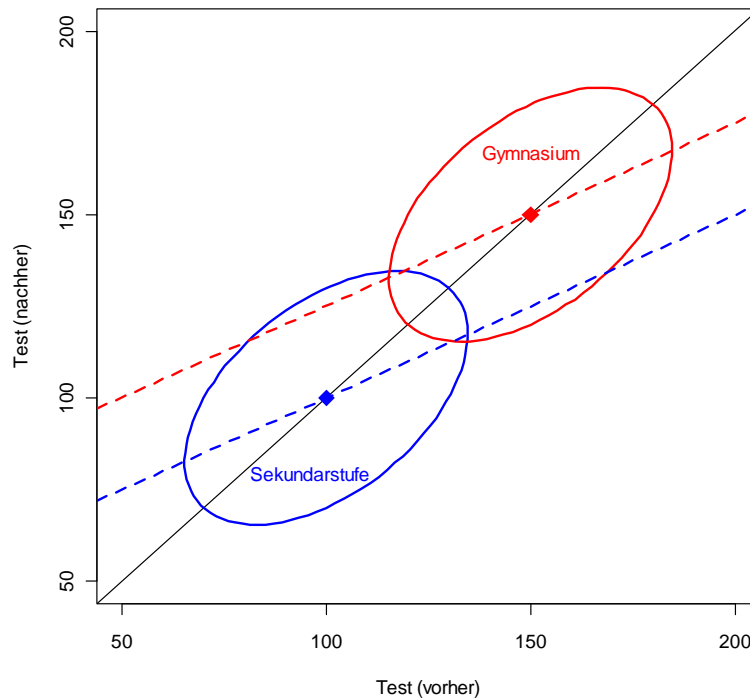


Abb. 2-10: Demonstration des Paradoxons von Lord: Obwohl sich keine der beiden Gruppen in der zweiten Messung (nachher) von der ersten Messung (vorher) unterscheidet, ergibt die Kovarianzanalyse einen signifikanten Effekt der Gruppen.

Ein t -Test welcher die Differenz zwischen Vorher- und Nachhermessung vergleicht, wird daher keinen Unterschied zwischen den Gruppen entdecken.

Führt man jedoch eine Kovarianzanalyse auf die *Nachhermessung* durch, in welche die *Gruppe* (Sekundarstufe vs. Gymnasium) als Faktor und die *Vorhermessung* als Kovariate eingehen, so ergibt sich ein signifikanter Effekt des Faktors *Gruppe* (Übung 2-6): Die Gymnasiastinnen schneiden besser ab als die Sekundarschüler (vergleiche Abb. 2-10).

Der Grund für diesen Unterschied (welcher *Lords Paradoxon* genannt wird) ist die Regression zum Mittel: Aufgrund dieser wird der Unterschied im Vortest nicht vollständig kontrolliert.

2.7 Zusammenfassung

Die folgenden psychologischen Mechanismen, welche zu fehlerhaften Kontingenz- und Kausalurteilen führen, wurden identifiziert:

- Fehlerhafte subjektive Theorien über kausale oder diagnostische Zusammenhänge;

- Eine fehlerhafte Konzeption von Zufallsprozessen und die Unterschätzung der Bedeutung des Zufalls;
- Das Ignorieren bzw. die fehlerhafte Gewichtung von relevanter Information in Kontingenztabelle;
- Das Ignorieren des Phänomens der Regression zum Mittel;
- Der Einfluss von Salienz- und Aufmerksamkeitseffekte.

Folgende methodische Aspekte, welche zentral für die Beurteilung von Kontingenz- und Kausalurteilen sind, wurden diskutiert:

- Ein einfacher Index – *Yules Q* – ermöglicht mit geringem Aufwand eine rasche Beurteilung des Zusammenhanges in 2×2 Kontingenztabelle.
- Für die Beurteilung von Zusammenhängen und kausalen Einflüssen müssen mögliche konfundierende Variablen und gemeinsame Ursachen berücksichtigt werden.
- Das *Paradoxon von Simpson* ist ein im Alltag häufig auftretendes Phänomen. Es lehrt, dass das Summieren über eine (konfundierende) Variable hinweg (=Bildung einer *marginalen* Tabelle ohne die relevante Variable) zu einem fehlerhaften Urteil führen kann. Im Extremfall kann das Urteil auf der Basis der *marginalen* Tabelle dem auf der vollen Tabelle basierenden Urteil entgegengesetzt sein.
- Der *ökologische Fehlschluss* ergibt sich analog zum Paradoxon von Simpson, indem eine relevante Variable, welche gewöhnlich eine Gruppe repräsentiert, ignoriert wird. Auch hier kann das Urteil, welches die Gruppe einbezieht konträr zu jenem sein, welches die Gruppe ignoriert.
- Das Phänomen der *Regression zum Mittel* manifestiert sich in der Tatsache, dass die Regressionsgerade stets unterhalb der Hauptachse der elliptischen Grundfläche, welche der Normalverteilung konstituiert, liegt (ausgenommen im Fall einer perfekten Korrelation). Die Regressionsgerade repräsentiert die bedingten Mittelwerte der abhängigen Variable *Y* bei einem gegebenen Wert der unabhängigen Variable *X* (für alle Werte von *X*). Der bedingten Mittelwert von *Y* gegeben ein fixer Wert von *X* ist besten (lineare) Schätzer von *Y*.
- Das *Paradoxon von Lord* ist eine Konsequenz des Phänomens der Regression zum Mittel. Es zeigt die Grenzen der Verwendung einer Kovarianzanalyse zur Kontrolle von Anfangsunterschieden zwischen den zu vergleichenden Gruppen auf.

2.8 Übungen



Übung 2-1:

Gegeben:

Ein akademischer Leistungstest, der *moderat* mit der GPA (Grade Point Average) korreliert ist: $\rho = 0.1$.

Tab. 2-6 zeigt die Perzentilwerte für die beiden Tests. Die Einträge sind wie folgt zu lesen:

Die besten 10% aller Personen, welche den Leistungstest absolvieren erreichen einen Wert von >750 .

Die besten 10% aller Personen, für welche der GPA vorliegt, haben einen Wert von > 3.7 .

Studenten	Leistungstest	GPA
Obersten 10%	>750	>3.7
Obersten 20%	>700	>3.5
Obersten 30%	>650	>3.2
Obersten 40%	>600	>2.9
Obersten 50%	>500	>2.5

Tab. 2-6: Perzentile der Scores für Leistungstest und GPA.

1. Angenommen eine ausgewählte Person hat beim Leistungstest einen Wert von 725 erreicht. Welchen Wert würden Sie für diese Person beim GPA vorhersagen.

Begründen Sie Ihr Urteil.

2. Versuchen Sie mittels linearer Regressionsgleichung die Werte des GPA für den beobachteten Wert des Leistungstests vorherzusagen, für die Korrelationskoeffizienten:

$$\rho = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 .$$

Nehmen Sie hierzu an, dass die Werte normal verteilt sind.

Hinweise:

(i) Aus den Daten müssen die Standardabweichungen für den Leistungstest sowie für den GPA ermittelt werden.

(ii) Die Regressionsgleichung zur Vorhersage von y aufgrund von x lautet:

$$y = \mu_y + b \cdot (x - \mu_x)$$

Hierbei gilt:

μ_y ist der Mittelwert von y .

μ_x ist der Mittelwert von x .

b ist der (unstandardisierte) Regressionskoeffizient.



Übung 2-2:

Im Jahre 1976 wurde der Polizeichef von Denver mit folgender Begründung entlassen:

Seit dem Amtsantritt des Polizeichefs im Jahre 1971 sei die Verbrechensrate um 14% gestiegen.

Was ist an dieser Argumentation problematisch?

**Übung 2-3:**

Geben Sie jeweils ein Beispiel zu:

- (a) Scheineffekten
- (b) Konfundierung

**Übung 2-4:**

Gegeben:

Untersuchung des Zusammenhangs zwischen Blutdruck und Geburtsgewicht (Tu, Gunnell & Gilthorpe, 2008):

Aktuelles Gewicht	Geburts- gewicht	Blutdruck		Σ	% normal	Yules Q
		normal	hoch			
≤ 90 kg	gering	329	99	428	77%	-0.09
	hoch	221	5	276	80%	
> 90 kg	gering	25	33	58	43%	-0.04
	hoch	101	131	238	45%	
Σ		682	318			

Tab. 2-7: Kreuztabellierung der drei Faktoren Blutdruck (normal vs. hoch), Geburtsgewicht (gering, vs. hoch) und aktuellem Gewicht (≤ 90 kg vs. > 90 kg).

Gemäss Tab. 2-7 gibt es einen leicht negativen Zusammenhang zwischen Geburtsgewicht und Blutdruck: Personen mit geringem Geburtsgewicht haben einen leicht erhöhten Blutdruck (wobei der Zusammenhang allerdings nur gering ausgeprägt ist). Dies gilt sowohl für die Personen mit aktuellem Gewicht ≤ 90 kg als auch für solche mit Gewicht > 90 kg.

Für die zusammengefassten Daten, summiert über das aktuelle Gewicht, ergibt sich:

Geburts- gewicht	Blutdruck		Σ	% normal	Yules Q
	normal	hoch			
gering	354	132	486	73%	0.207
hoch	326	186	514	64%	
Σ	682	318			

Tab. 2-8: Kreuztabellierung der drei Faktoren Blutdruck (normal vs. hoch) und Geburtsgewicht (gering, vs. hoch). Die Werte ergeben sich durch aus den Daten von Tab. 2-7 durch Summation über den Faktor aktuelles Gewicht.

Für die zusammengefassten Daten, summiert über das aktuelle Gewicht, ergibt sich nun ein positiver Zusammenhang zwischen geringem Blutdruck und Geburtsgewicht: Personen mit geringerem Geburtsgewicht haben eher einen normalen Blutdruck (73% vs. 64%).

Erklären Sie möglichst genau die Gründe für die beobachtete Diskrepanz in der Beziehung zwischen Geburtsgewicht und Blutdruck zwischen Tab. 2-7 und Tab. 2-8.



Übung 2-5:

Gegeben:

Die Daten zu Bsp.2-23 (Abb. 2-8).

Berechne:

- Die falsche Regression, welche die Clusterstruktur ignoriert.
- Die korrekte Regression, welche die Clusterstruktur berücksichtigt.

Berichte:

- Das Ergebnis des Signifikanztests für den Regressionsparameter für (a).
- Die drei Regressionsgleichungen für (b).
- Das Ergebnis des Signifikanztests für den Regressionsparameter für (b).



Übung 2-6: Lords Paradoxon (Lord, 1967):

Gegeben: Die Daten zweier Gruppen (vgl. Abb. 2-11)

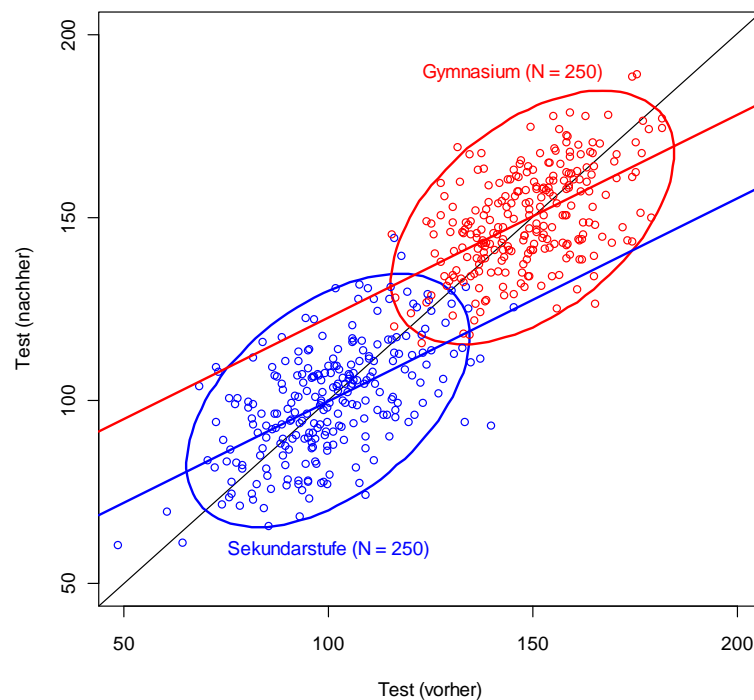


Abb. 2-11: Ergebnisse eines Test für Sekundarschüler und Gymnasiasten vor Eintritt ins Gymnasium und ein Jahr danach (Simulierte Daten). Die Ellipsen repräsentieren die 95%-Konfidenzbereiche der bivariaten Normalverteilungen für die beiden Populationen, aus denen die beiden Stichproben gezogen wurden. Die kleinen offenen Kreise repräsentieren die einzelnen Datenpunkte. Eingezeichnet sind auch die beiden Regressionsgeraden, welche sich aufgrund der gezogenen Stichproben ergeben.

Demonstrieren Sie aufgrund der Daten Lords Paradoxon:

- (a) Führen Sie eine Kovarianzanalyse durch mit Test (*nachher*) als abhängige und *Schule* als unabhängige Variable, sowie mit Test (*vorher*) als Kovariate.

Diese Analyse sollte einen signifikanten Effekt des Faktors *Schule* ergeben.

- (b) Nun führen Sie einen *t*-Test für unabhängige Stichproben (die beiden Schultypen) auf die Differenz *nachher* – *vorher* durch.

Der *t*-Test sollte keinen signifikanten Unterschied zwischen den beiden Gruppen ergeben.

Bemerkung:

Anstelle der Kovarianzanalyse kann auch eine Regressionsanalyse durchgeführt werden.

3. Gedächtnisurteile

In diesem Kapitel werden verschiedene Formen von Urteilsfehlern im Zusammenhang mit Gedächtnisurteilen behandelt. Wir beginnen mit einem kleinen Überblick über klassische Gedächtnisfehler. Dem folgt die Behandlung von Urteilsfehlern im Zusammenhang mit der Einschätzung von Veränderungen und Konstanten im persönlichen Leben. Ein klassischer Fehler ist der so genannte Rückschaufehler, welcher als nächstes erklärt wird. Zum Schluss behandeln wir ein interessantes Phänomen im Zusammenhang mit der retrospektiven Beurteilung von schmerzhaften Erlebnissen.

3.1 Klassische Studien zu Gedächtnisfehlern

Die ersten wirklich bekannten Studien, welche klar demonstrieren, dass Personen ihre Gedächtnisinhalte systematisch anpassen um ein Geschehen plausibel zu machen, stammen von Bartlett (1932). Diese Untersuchungen enthüllen folgende zentrale Mechanismen unseres Gedächtnisses.



Kognitiver Mechanismus 3-1: Speicherung, Abruf und Vergessen von Gedächtnisinhalten

1. Speicherung / Enkodierung:

Der Prozess der Speicherung darf nicht als ein einfaches Ablegen von Information in einem Speicher betrachtet werden. Vielmehr wird die zu speichernde Information *in sinnvolle Einheiten zerlegt und an die bestehende Wissensstruktur angepasst*.

Es werden meist auch nicht alle Details gespeichert, sondern die gespeicherte Information ist eher *schematisch*, wobei Faktoren wie Aufmerksamkeit, Tiefe der Verarbeitung, Vorwissen, Interesse oder emotionale Bedeutung zentral dafür sind, welche Information gespeichert wird.

2. Abruf von Information:

Der Abruf von Information aus dem Gedächtnis ist als ein *Prozess der (Re-) Konstruktion* zu betrachten, in dem Informationen aus verschiedenen Informationsquellen integriert werden.

Hierbei werden Lücken gefüllt, wobei subjektive Theorien, sowie durch Überlegungen zur Plausibilität und Stimmigkeit herangezogen werden.

3. Vergessen / Interferenz:

Gemäss der Interferenz-Theorie des Vergessens können neue Informationen den Abruf von zuvor enkodierter Information blockieren (retroaktive Interferenz) [Auch das Umgekehrte ist möglich: proaktive Interferenz].

Hierbei kann es zu so genannten *Intrusionen* kommen, d.h. irrelevante interferierende Informationseinheiten werden abgerufen und verhindern so den Abruf der relevanten Information.

Wir werden im Folgenden diese allgemein gehaltene Spezifikation durch die Beschreibung konkreterer Mechanismen noch weiter anreichern. Hierbei konzentrieren wir uns auf zwei Forschungsrichtungen (die jedoch nicht unabhängig sind), welche einen grossen Einfluss auf die Gedächtnisforschung hatten:

1. Die Studien von Loftus und Kollegen zum Einfluss von Fehlinformation auf Zeugenaussagen. Einen Überblick über diese Studien findet man in Pickrell, Bernstein und Loftus, (2004).
2. Die Studien von Deese (1959) und Roediger und McDermott (1995) über so genannte assoziative Gedächtnisillusionen. Einen Überblick hierzu findet man in Roediger und Gallo (2004).

3.1.1 Fehlinformationen und Gedächtnisfehler

In den 70-iger Jahren wurde von Elisabeth Loftus das Fehlinformationsparadigma entwickelt.



Konzept 3-1: Das Fehlinformationsparadigma:

Beim *Fehlinformationsparadigma* handelt es sich um eine (experimentelle) Prozedur, bei der durch gezielte Fehlinformation nach dem eigentlichen Ereignis (= *Post-event* Information) die Erinnerung an das ursprüngliche Ereignis verfälscht wird.

Loftus und Kollegen haben in mehreren Experimenten demonstriert, wie durch gezielte Fehlinformation falsche Zeugenaussagen über ein Ereignis provoziert werden können (z.B. Loftus & Palmer, 1974). Die beiden folgenden Beispiele aus dem Alltag demonstrieren eindrucksvoll die Wirkung von Fehlinformation auf Erinnerungen, wobei jedoch – im Gegensatz zu den Experimenten von Loftus – (höchstwahrscheinlich) überhaupt kein vorangehendes Ereignis stattgefunden hat.



Bsp.3-1: Erinnerung und Suggestion (Loftus, 1980, Seite 119f):

Jean Piaget berichtete, dass er sich lebhaft an den Versuch, ihn auf dem Champs-Élysée aus dem Kinderwagen zu entführen, erinnerte. Er »erinnerte« sich an die Menschenansammlung, an die Kratzer im Gesicht des mutigen Kindermädchens, das ihn gerettet hatte, an den weissen Knüppel des Polizisten und an den weglaufernden Angreifer.

So lebhaft Piagets Erinnerungen auch waren, sie waren dennoch falsch. Jahre später gestand das Kindermädchen, dass sie die ganze Geschichte erfunden hatte. Scheinbar hatte Piaget die Geschichte unter dem Einfluss der Autoritätsperson internalisiert.



Bsp.3-2: Erinnerungen eines Hilfssheriffs (Schacter, 1999, Seite 214ff):

Der 43-jährige Hilfssheriff Paul Ingram wurde von seinen beiden Töchtern beschuldigt, sie als Kinder sexuell missbraucht zu haben.

Zuerst leugnete er alle Vorwürfe, da er sich absolut nicht an die Vorfälle erinnern konnte. Seine Kollegen, die Beamten und der Pfarrer (Ingram gehörte einer fundamentalistischen Kirche an) versicherten ihm, dass er, sobald er ein Geständnis abgelegt habe, sich auch wieder an die Vorfälle erinnern werde.

Nach stundenlangen Gesprächen und Verhören gestand Ingram und meinte, er habe die Vorfälle wahrscheinlich verdrängt.

Im Fortgang der Verhöre zog die Untersuchung immer weitere Kreise, da die Beamten vermuteten, die sexuellen Missbräuche haben im Rahmen eines Satanskults stattgefunden. Ingram konnte sich immer besser an die Ereignisse »erinnern« und beschuldigte weitere Mittäter, die die Taten vehement bestritten.

Gleichzeitig »erinnerten« sich die beiden Mädchen an weitere Details des Satanskults, wie Babymorde, Massenorgien und Tieropfer.

Als Ingram im Gefängnis auf seinen Prozess wartete, wurde er von dem Sozialpsychologen Richard Ofshe befragt. Dieser forderte Ingram dazu auf, sich daran zu erinnern, wie er seinen Sohn gezwungen hatte, vor seinen Augen mit seiner Tochter zu schlafen.

Dieses Ereignis ähnelte in der Logik den anderen, mit dem einen Unterschied, dass keine seiner Töchter je behauptet hatte, dass es stattgefunden habe und auch der Sohn bestritt es vehement.

Ingrams Reaktion auf die Frage des Psychologen folgte einem vorhersehbaren Muster. Zuerst konnte er sich nicht daran erinnern aber nach eingehendem Visualisieren und Beten entdeckte er eine lebhaft »Erinnerung« an den Vorfall.

Ingram wurde schliesslich zu einer Freiheitsstrafe von 20 Jahren verurteilt. Zum Zeitpunkt des Berichts von Schacter (Mehr als 6 Jahre nach dem Urteil) sass er noch im Gefängnis.



Bemerkungen:

1. Der Fall Ingram zeigt, dass Geständnisse noch lange nicht die Schuld des Angeklagten implizieren. So gab es z.B. mehr als 200 Personen, welche gestanden, an der Entführung des Lindbergh-Babies (am 1. März 1932) beteiligt gewesen zu sein (Lilienfeld et al., 2010).
2. Kinder sind als Zeugen vor Gericht ungeeignet, da sie in extremem Ausmass zu Konfabulationen neigen (siehe z.B. Schacter, 1999).

Obwohl der Effekt von Fehlinformation auf das Gedächtnis zweifelsfrei belegt ist, gibt es hinsichtlich der Erklärung des Phänomens keine Einigkeit. Loftus hatte ursprünglich die Ansicht vertreten, dass der ursprüngliche Gedächtnisinhalt durch die Fehlinformation abgeändert wird.

Experimente von McCloskey & Zaragoza (1985) liessen jedoch Zweifel an dieser Erklärung aufkommen: Gibt man nämlich den Vpn die Möglichkeit, im Nachtest zwischen dem ursprünglich gezeigten Bild und einem neuen Bild zu wählen, so wählen Personen mit und ohne Fehlinformation in etwa gleich oft das korrekte Bild. Falls der ursprüngliche Gedächtnisinhalt abgeändert worden wäre, so sollte bei der Gruppe mit Fehlinformation in grösserem Ausmass das falsche Bild gewählt werden als bei der Gruppe ohne Fehlinformation.

Eine bessere Erklärung ergibt sich aus den Untersuchungen zum so genannten *Reality Monitoring* und zum *Quellengedächtnis*.



Kognitiver Mechanismus 3-2: Reality Monitoring (Johnson & Raye, 1981):

Reality Monitoring beschäftigt sich mit der Frage, wie wir reale Ereignisse von nicht realen (gedachten, geträumten, etc.) unterscheiden.

Zur Beurteilung der Korrektheit von Erinnerungen orientieren wir uns an bestimmte Qualitäten der Erinnerung, wie Lebhaftigkeit, Detailliertheit, Vertrautheit und Plausibilität.



Kognitiver Mechanismus 3-3: Quellengedächtnis und Source Monitoring (Johnson, Hashtroudi & Lindsay, 1993; Mitchell & Johnson, 2000, 2009):

Das *Quellengedächtnis* beinhaltet Information über die Quellen bestimmter Informationen. Dies betrifft Probleme wie: »Woher kenne ich eine vertraute Person?« oder »Woher habe ich diese Information?«.

Das Quellengedächtnis ist viel stärker durch Alterungsprozesse betroffen als das Gedächtnis für Faktenwissen. Weiters werden Defizite im Quellengedächtnis durch Verwendung von Schlussfolgerungen auf der Basis des allgemeinen Weltwissens kompensiert (Mather, Johnson, & DeLeonardis, 1999).



Frage:

Wie können die Erkenntnisse zum Quellengedächtnis und Reality Monitoring zur Erklärung der Wirkung von Fehlinformation herangezogen werden?

Zahlreiche Studien belegen, dass die Instruktion, sich irgendwelche vergangene Ereignisse vorzustellen und sich im Detail auszumalen, dazu führen, dass die Ereignisse eher als tatsächlich stattgefunden empfunden werden (z.B. Gerry, Manning, Loftus, & Sherman, 1996). Dies kann erklären, warum Paul Ingram nach »eingehendem Visualisieren« plötzlich lebhaftere Erinnerungen an die Ereignisse aufwies.

Ein weiterer Faktor, der hier relevant ist, bildet die Tatsache, dass sich die Aussagen von Ingram und seinen Töchtern über stattgefundene Satanskulte sich gegenseitig zu bestätigen schienen. Dies erhöhte ihre Plausibilität und damit die Wahrscheinlichkeit ihrer Realität.

Das Quellengedächtnis ist hier ebenfalls von Bedeutung, denn es scheint, dass – speziell im Fall von Piaget – die Personen sich zwar noch an die Inhalte aber nicht mehr an die Quellen (woher die Inhalte stammen) erinnern konnten. So glaubten die Personen, dass sie bestimmte Erlebnisse, die sie nur von anderen gehört hatten, plötzlich als selbst erlebt erinnern.



Schlussbemerkung:

Man hüte sich vor Erinnerungen, welche durch spezielle imaginative Techniken, wie sie in verschiedenen Therapien eingesetzt werden, oder durch Hypnose »wieder gefunden« werden. Derartige wieder aufgedeckte Erinnerungen haben in der (amerikanischen) Rechtssprechung ziemlichen Schaden angerichtet.

3.1.2 Assoziative Gedächtnisillusionen: Das DRM-Paradigma

Eine andere, subtilere Art von Fehlinformation wird im so genannten *DRM-Paradigma* (Deese-Roediger-McDermott Paradigma) präsentiert.



Konzept 3-2: *DRM-Paradigma (Deese, 1959; Roediger & McDermott, 1995, Experiment 2)*

Das *DRM-Paradigma* ist eine experimentelle Prozedur zur Erzeugung assoziativer Gedächtnisillusionen.

(i) *Versuchsmaterial:*

Das Material besteht aus mehreren Listen (z.B. 16 Listen) von Worten, wobei alle Worte mit einem bestimmten kritischen Wort assoziiert sind: z.B. 24 Listen, mit 15 Worten pro Liste. Die kritischen Wörter befinden sich jedoch nicht auf den Listen.

Beispiel:

Kritische Wort: *Schlaf*

Wortliste mit 15 Items:

Bett, Ruhe, Aufwachen, Müde, Traum, Wach, Nickerchen, Bettdecke, Dösen, Schlummern, Schnarchen, Schläfchen, Friede, Gähnen, Schläfrig.

(ii) *Ablauf:*

Von den 24 Listen werden 16 präsentiert, eine Liste nach der anderen (Präsentationsrate: 1-2 sec pro Wort).

Bei der Hälfte der Listen erfolgt nach der Präsentation der Liste ein Test in Form eines freien Abrufs. Hierbei sollten sie Vpn so viele Worte wie möglich aus der Liste wiedergeben, *aber ohne zu raten*.

Bei den anderen 8 Listen erfolgt nach Präsentation einer Liste anstelle des Abrufs eine kurze Distraktoraufgabe.

Nach Abarbeitung aller 16 Listen erfolgt ein Wiedererkennungstest. In diesem werden 96 Worte aus den folgenden Kategorien präsentiert:

- 48 neue Worte, unter denen sich auch die 16 kritischen Worte zu den 16 präsentierten Listen befinden.
- 48 alte Worte, davon 24 aus den 8 Listen mit Abruf und 24 aus den anderen 8 Listen.

Im Wiedererkennungstest müssen die Vpn für jedes Wort, das sie wieder erkannten, angeben, ob sie sich genau an den Moment der Präsentation erinnerten (=Erinnern) oder nicht (=Wissen).

(iii) *Die wichtigsten Ergebnisse:*

- Die kritischen Worte werden im freien Abruf ungefähr gleich oft produziert (in manchen Untersuchungen sogar öfter) wie die Worte auf den Listen.
- Im Wiedererkennungstest werden etwas mehr kritische Worte wieder erkannt als Listenworte.
- Falls es sich um eine Liste handelt, für die zuvor eine Abrufaufgabe durchgeführt wurde, so wurde bei 70% der wieder erkannten kritischen Worte von den Vpn angegeben, dass sie sich an den Moment der Präsentation erinnerten. Insgesamt wurden 80% der kritischen Worte fälschlicher Weise wieder erkannt.

- Bei den Listen, für die kein Abruf erfolgte war der Anteil der unter der Kategorie Erinnerung genannten kritischen Wörter geringer (etwa 45-50% der wieder erkannten kritischen Worte). Insgesamt wurden in dieser Bedingung 70% der kritischen Worte fälschlicher Weise wieder erkannt.
- Der Anteil falsch wieder erkannter kritischer Worte sinkt mit der Anzahl der mit dem kritischen Wort assoziierten Items in den Listen (bei Konstanthaltung der Anzahl zu lernender Worte).
- Kritische Worte werden auch nach einem längeren Zeitraum von 3 Wochen noch genauso gut »wieder erkannt« wie unmittelbar nach Präsentation, während für die präsentierten Items ein deutlicher Abfall zu beobachten ist.

Zur Erklärung des Phänomens wird der folgende kognitive Mechanismus herangezogen:



Kognitiver Mechanismus 3-4: Implizite Assoziative Response

Gemäss dieser Erklärung führt die Präsentation der assoziierten Items zur Aktivierung des kritischen Wortes.

Das Auftreten impliziter assoziativer Aktivierungen wird durch folgende Fakten gestützt:

- Der Effekt des falschen Erinnerns und Wiedererkennens zeigt sich auch bei sehr schneller Präsentation (40 msec) der Items.
- In einer nachfolgenden Aufgabe konnten Priming-Effekte auf mehreren impliziten Gedächtnistests gemessen werden.

Jede dieser impliziten Aktivierungen erhöht die *Fluidität* des Abrufs des Wortes und macht so das Wort für die Person vertraut ohne dass sich die Person dessen bewusst sein muss.

Die implizite Aktivierung führt jedoch oft auch dazu, dass die Personen ganz bewusst an das kritische Wort denken. Auch dieses bewusste Denken an das Wort hat einen Effekt auf das falsche Wiedererkennen des kritischen Wortes.



Bemerkungen:

Bei *Primingeffekte* handelt es sich um Erleichterungseffekte im weitesten Sinne, d.h. aufgrund der vorgängigen Verarbeitung von Information wird die nachfolgende Verarbeitung von damit in Verbindung stehender Information erleichtert.

So wird z.B. beim so genannten *semantischen Priming* die Verarbeitung eines Wortes (z.B. eine lexikalische Entscheidung darüber, ob es sich um ein Wort oder Non-Wort handelt) durch die vorangegangene Präsentation erleichtert.

Bei *Impliziten Gedächtnistests* handelt es sich um Gedächtnistests, die nicht als solche erkennbar sind. Sie dienen der Messung von Gedächtniseffekten, welche mit expliziten Gedächtnistests (wie Abruf und Wiedererkennen) meist nicht messbar sind.

Ein Beispiel für einen impliziten Gedächtnistest bildet eine Wortvervollständigungs-aufgabe, wobei untersucht wird, ob ein Wortfragment eher zu einem zuvor präsentierten Wort bzw. zu einem dazu verwandten vervollständigt wird.

Der Prozesse der impliziten assoziativen Response reicht alleine nicht aus, um zu erklären, warum bei einem Grossteil der kritischen Worte die Illusion entsteht, das Wort tatsächlich wahrgenommen zu haben. Zur Erklärung müssen auch die beiden Mechanismen der *Realitätsüberwachung* (vgl. Kognitiver Mechanismus 3-2) und der *Quellenüberwachung* (vgl. Kognitiver Mechanismus 3-3) herangezogen werden.

Der Prozess der *Realitätsüberwachung* ist im aktuellen Kontext insofern von Bedeutung als durch die wiederholte Aktivierung des kritischen Wortes eine *erhöhte Vertrautheit und Fluidität des Abrufs* induziert wird. Diese erhöhte Vertrautheit und Fluidität wird zur Beurteilung der Realität der Erinnerung mit herangezogen und kann daher zur Illusion führen, dass einer konkreten Erinnerung vorliegt.



Bemerkungen:

Die Bedeutung von derartigen Überwachungsprozessen wird durch neurophysiologische Daten gestützt: Personen mit Läsionen im Frontalhirnbereich zeigen einen erhöhten DRM-Effekt. Bekanntlich ist dieser Hirnbereich auf Kontroll- und Steuerungsaufgaben spezialisiert.

Dies erklärt auch den Befund, warum Gedächtnisillusionen bei Kindern relativ einfach hervorgerufen werden können. Bei diesen sind die relevanten Kontroll- und Steuerungsfunktionen noch nicht ausreichen ausgeprägt.

Eine noch wichtigere Rolle spielt das *Quellengedächtnis*: Köderworte wurden vor allem dann als »mit einer konkreten Erinnerung verbunden« beurteilt, wenn sie im zuvor durchgeführten freien Erinnerungstest irrtümlich genannt worden waren. Scheinbar können sich die Personen zwar erinnern, das Item zuvor gesehen zu haben, es ist jedoch das Wissen über die Quelle, d.h. ob sie dem Wort beim Lernen oder beim Abruftest begegnet sind, verloren gegangen.

3.2 Beurteilung von Stabilität und Veränderung

Die im letzten Abschnitt präsentierten Befunde belegen die Bedeutung von Kontroll- und Überwachungsprozessen für das Gedächtnis. Es wurde aber auch der Einfluss von schlussfolgernden Prozessen auf die Erinnerung erkennbar (vgl. Bsp.3-2). Die Bedeutung von plausiblen Schlussfolgerungen auf die Erinnerung zeigt sich auch vor allem in

Studien, welche sich damit befassen, wie Personen mögliche Veränderungen bzw. Konstanten in ihrem persönlichen Leben beurteilen.



Bsp.3-3: Verzerrungen in der Beurteilung vergangener Einstellungen (Marcus, 1986):

Ziel der Studie:

Untersuchung der politischen Einstellung bzw. Einstellungsänderung zwischen 1965-1982.

Stichprobe:

1669 Vpn (Studenten und ihre Eltern), wobei 64% der Eltern an allen 3 Befragungszeitpunkten teilnahmen, sowie 68% der Studenten.

Durchführung:

Die Vpn wurden zu 3 Zeitpunkten (1965, 1973, 1982) über ihre politische Einstellung befragt. Speziell mussten sie auf einer 7-Punkteskala ihre Einstellung zu 5 Punkten darlegen:

- Garantierter Job,
- Rechte von Angeklagten,
- Hilfe für Minderheiten,
- Legalisierung von Marihuana,
- Gleichstellung der Frau.

Zusätzlich sollten sie ihre eigene politische Einstellung als liberal oder konservativ klassifizieren.

Ergebnis:

Die 1982 erhobene Einschätzung der Einstellung von 1973 zeigte, dass sich die Vpn nahezu identische Einstellungen wie zum aktuellen Zeitpunkt zuschrieben: Eine Regression, in der die retrospektiv beurteilten Einstellung (von 1973) (= RE) mittels der Einstellung von 1982 (= E82) und der wahren Einstellung (von 1973) (= E73) geschätzt wurde:

$$RE = a + b_{E82} \cdot E82 + b_{E73} \cdot E73$$

ergab, dass E82 ein weitaus grösseres Gewicht hatten als E73. Dies bedeutet, dass sich die retrospektive Einschätzung der vergangenen Einstellung viel besser mittels der aktuellen Einstellung vorhersagen lässt, als mittels der Einstellungen, welche die Personen tatsächlich im Jahre 1973 aufwiesen.

Die Erklärung dieser Ergebnisse basiert auf folgendem Mechanismus:



Kognitiver Mechanismus 3-5: Verankerung und Anpassung
(»Anchoring and adjustment«)

Verankerung und Anpassung wird vor allem im Kontext der Schätzung von Quantitäten verwendet.

Der Mechanismus der *Verankerung und Anpassung* besteht aus zwei Schritten:

1. In einem ersten Schritt wird aufgrund der verfügbaren Information eine Anfangsschätzung gebildet. Die verfügbare Information bildet hierbei den Anker.
2. In einem zweiten Schritt wird sodann die Anfangsschätzung aufgrund von zusätzlichen Überlegungen angepasst.

Das Problem dieser Methode zur Urteilsbildung besteht darin, dass einerseits für die Schätzung oft irrelevante Information verwendet wird und andererseits der Prozess der Anpassung nur ungenügend ist.



Bsp.3-4: Verankerung und Anpassung (Tversky & Kahneman, 1974):

Im Experiment wurde zuerst ein Glücksrad gedreht. Anschliessend wurde den Vpn folgende Frage vorgelegt:

»Wie hoch ist der Prozentsatz afrikanischer Nationen in den vereinten Nationen?«

Innerhalb der Gruppe von Vpn, bei welcher die Nadel des Glücksrades die Zahl 65 indizierte, betrug der mediane Schätzwert 45 (Median bedeutet, dass 50% der Schätzungen darunter und 50% darüber lagen).

In der Gruppe, bei der die Nadel auf den Wert 10 zeigte, betrug der Median der Schätzungen 25.

Interpretation:

Die Personen nehmen den vom Glücksrad angezeigten Wert als Anker für ihre Schätzung und passen ihr Urteil dann an, in einem Fall nach unten und im anderen Fall nach oben.

Der Prozess der Anpassung ist nicht ausreichend.

Kehren wir nun zur Untersuchung von Marcus (1986) zurück: Offensichtlich verwendeten die Personen zur Beurteilung ihrer Einstellungen von 1973 ihre aktuellen Einstellungen als Anker. Sodann führten Sie einen Prozess der Anpassung durch. Die Ergebnisse der Regressionsanalyse zeigen eindeutig, dass dieser Prozess der Anpassung nur in unzureichender Weise durchgeführt wurde.

Die Untersuchung zeigt auch, dass der Prozess der Anpassung durch naive subjektive Theorien geleitet wurde: So schätzten die Personen ihre Meinung von 1973 als eher liberaler ein, oder die Eltern schätzten ihre Einstellung als stabiler ein als jene der Kinder (in Wirklichkeit war genau das Gegenteil der Fall). Dies beruht auf plausiblen Annahmen, wie: »Personen werden mit zunehmenden Lebensalter konservativer«, oder »Ältere Personen haben eine stabilere politische Einstellung«.

Die Ergebnisse dieser Untersuchung werden durch weitere Untersuchung gestützt, die zeigten, dass die oben dargestellten Verzerrungen nicht nur politische Einstellungen sondern auch andere persönliche

Einschätzungen betreffen (Ross, 1989). Interessant ist hier noch eine Studie von Conway & Ross (1984), welche zeigt, wie durch Verzerrung der Vergangenheit ein zufrieden stellendes Ergebnis für die Gegenwart resultiert.



Bsp.3-5: Verzerrung hinsichtlich der Beurteilung der eigenen Leistung (Conway & Ross, 1984):

Ziel der Studie:

Untersuchung der Wirkung eines Universitätsprogramms zur Förderung der Lernfähigkeiten der Studenten.

Durchführung:

Vpn wurden zufällig in 2 Gruppen geteilt, in eine Kontrollgruppe (keine Teilnahme am Programm, sondern Platz auf Warteliste) und in eine Experimentalgruppe. In einem Vor- und Nachtest wurde die Wirkung des Programms erhoben.

Ergebnis:

- Das Programm hatte keine Wirkung.
- Befragt über Dauer ihrer Studienzeiten, sowie ihre Studienleistung im Allgemeinen gab es Unterschiede zwischen den Gruppen:
Die Vpn der Experimentalgruppe schätzten ihre frühere Leistung als schlechter ein, nicht so die Vpn der Kontrollgruppe.

Diese Ergebnisse belegen, dass die Beurteilung von Veränderungen und Stabilität bezüglich der eigenen Vergangenheit sehr stark durch subjektive Theorien über Veränderungen mit bestimmt wird.

3.3 Der Rückschaufehler (*Hindsight Bias*)

Wir beginnen mit einer Spezifikation des Konzepts des Rückschaufehlers.



Konzept 3-3: Rückschaufehler (*Hindsight Bias*)

Der *Rückschaufehler* besteht in der Tatsache, dass Personen in der Rückschau ihr früheres Wissen (vor Eintreten des relevanten Ereignisses) falsch beurteilen: Im Spezifischen sehen sie im Nachhinein eine grössere Übereinstimmung zwischen ihrem früheren Wissen und dem aktuellen Wissen (nach Eintreten des relevanten Ereignisses) [»Ich wusste es schon immer«].

Die Fälle von Rückschaufehlern sind zahlreich. Hier eine klassische Studie.



Bsp.3-6: Rückschaufehler (Fischhoff & Beyth, 1975):

Die Untersuchung bestand aus 2 Phase:

In der ersten Phase wurden mehrere Gruppen von Studenten gebeten, die Wahrscheinlichkeit von 15 Ergebnissen zur Reise von Nixon entweder in die UdSSR oder nach China im Jahre 1972 zu schätzen. Diese Einschätzungen erfolgten vor der Reise.

Beispiele: Würde die USA eine diplomatische Niederlassung in China gründen? oder Würde Nixon ein gemeinsames Weltraumprogramm mit der UdSSR initiieren?

In der zweiten Phase, welche zwischen 2 Wochen bis 6 Monate nach der Reise erfolgte, sollten die Vpn sich daran erinnern, welche Schätzungen sie vor der Reise abgegeben hatten. Weiters sollten sie angeben, ob nach ihrem Wissen das Ereignis tatsächlich stattgefunden hat oder nicht.

Ergebnisse:

1. 3/4 der Studenten nannten für Ereignisse, von denen sie glaubten, dass sie tatsächlich stattgefunden hatten, im Nachhinein höhere Schätzungen als sie ursprünglich gegeben hatten.
2. Die Mehrheit der Studenten gab hingegen für Ereignisse, von denen sie glaubten, dass sie nicht stattgefunden hatten, im Nachhinein eine niedrigere Schätzung als die ursprünglich gegebene.
3. Der Rückschaufehler verstärkte sich mit dem Abstand zum Ereignis. Nach 3-6 Monaten zeigten 84% der Vpn einen Rückschaufehler.

Der Rückschaufehler lässt sich ebenfalls mit dem Mechanismus der *Verankerung und unvollständigen Anpassung* erklären: Der aktuelle Outcome dient als Anker und die Personen sind nicht in der Lage ihr Urteil adäquat anzupassen, um ihre früheren Vorhersagen korrekt abzuschätzen.

Der Rückschaufehler lässt sich verkleinern, wenn die Personen gebeten werden nach Gründen zu suchen, warum die Ereignisse sich auch anders hätten entwickeln können. Unter diesen Bedingungen sind die Personen scheinbar besser in der Lage, sich vom Anker zu lösen und eine adäquatere Anpassung durchzuführen.

3.4 Retrospektive Evaluation affektiver Episoden

Eine Reihe von Studien zeigt, dass Personen bei der retrospektiven Evaluation eines länger andauernden Ereignisses, sei es nun positiv oder negativ, die Dauer des Ereignisses nicht ausreichend berücksichtigen. Das folgende Experiment verdeutlicht dies.



Bsp.3-7: Evaluation einer schmerzlichen Erfahrung
(Kahneman, Fredrickson, Schreiber, & Redelmeier, 1993)

Die Vpn mussten sich zwei Erfahrungen mit kaltem Druckwasser unterziehen, eine pro Hand:

In der *kurzen Episode* wurde die Hand für 60 sec in 14°C kaltes Wasser getaucht.

In der *langen Episode* wurde die Hand zuerst für 60 sec in 14°C kaltes Wasser getaucht. Hierauf folgten 30 sec, in der das Wasser schrittweise auf 15°C erwärmt wurde, wobei die Hand die ganze Zeit im Wasser verblieb.

Die Personen wurden einige Zeit nach dem Experiment gefragt, welche der beiden Bedingungen sie wiederholen möchten.

Ergebnis:

Paradoxerweise wollte eine signifikante Mehrheit der Vpn die *lange Episode* wiederholen, obwohl diese die kürzere Episode beinhaltet und dies für die Vpn völlig transparent war.

Das Ergebnis dieser Studie kann für medizinische Zwecke ausgenutzt werden.



Bsp.3-8: Evaluation der Erfahrung bei einer Darmspiegelung
(Redelmeier, Katz, & Kahneman, 2003)

Bei den Personen ($N = 682$) handelte es sich um Klienten, die sich einer Darmspiegelung unterziehen mussten.

Bei der Hälfte der Personen wurde ohne deren Wissen die Darmsonde nicht sofort nach der eigentlichen Untersuchung entfernt, sondern noch für ca. 1 Minute bewegungslos im Darm behalten (= verlängerte Untersuchung).

Ergebnis:

- In der Gruppe mit verlängerter Untersuchungszeit wurde die Untersuchung retrospektiv positiver beurteilt als in der anderen Gruppe.
- In einer Befragung 5 Jahre nach der Untersuchung waren mehr Personen der Gruppe mit verlängerter Zeit bereit, sich der Prozedur nochmals zu unterziehen als in der anderen Gruppe.

Wie lässt sich nun dieses Phänomen erklären? Hierzu wurde der folgende Mechanismus vorgeschlagen.



Kognitiver Mechanismus 3-6: Das »Schnappschussmodell« der retrospektiven Evaluation von affektiven Episoden (Fredrickson & Kahneman, 1993)

Grundannahme des Modells:

Die retrospektiv evaluierte Episode wird durch prototypische Momente repräsentiert.

Im Falle affektiver Episoden werden die affektiven Spitzenmomente und der Affekt am Ende der Episode repräsentiert.

Die retrospektive Evaluation der Gesamtepisode basiert auf einem Mittelwert aus der Evaluation der Momente mit affektiver Spitze und dem Affekt am Ende der Episode.

Im Falle von Bsp.3-7 & Bsp.3-8 ergibt sich die Bevorzugung der längeren Episode, indem hier die Episode jeweils durch ein weniger unangenehmes Ereignis am Ende verlängert wird. Dieses weniger unangenehme Ende treibt die Evaluation der Gesamtepisode nach oben.

Für die retrospektive Evaluation ist von Bedeutung, dass das angenehmere Ende noch als Teil der Episode gesehen wird. Folgt auf eine negative Episode ein »belohnendes« Ereignis, so muss dieses nicht zu einer erhöhten Evaluation der negativen Episode führen.

3.5 Zusammenfassung

Die folgenden psychologischen Phänomene und Mechanismen im Zusammenhang mit Gedächtnisurteilen wurden in diesem Kapitel besprochen:

- Die gezielte (und permanente) Präsentation von Fehlinformation kann zu falschen Erinnerungen führen, die subjektiv als sehr lebhaft und realistisch empfunden werden können.
- Suggestive und imaginative Techniken zur »Wiedergewinnung« von Erinnerungen, wie sie in manchen Therapien angewendet werden, sind besonders geeignet, lebhaft aber falsche Erinnerungen zu evozieren. Der Grund für den »Erfolg« derartiger Techniken liegt in ihrem Einfluss auf Prozesse der Realitäts- und Quellenüberwachung: Sie erzeugen Hinweisreize welche irrtümlicherweise Realität vorspiegeln. Bei permanenter Anwendung wird zudem die Beurteilung des Ursprungs der falschen Erinnerung erschwert.
- Eine Art der Präsentation von irreführender Information besteht in Verwendung von semantisch verwandter Information. Diese führt zu einer impliziten Aktivierung von Information bzw. einer quasi-automatischen Schlussfolgerung auf die relevante, aber nicht präsentierte Information. Hieraus resultiert das Phänomen der assoziativen Gedächtnisillusion.
- Im Prozess des Wiedererinnerns werden Erinnerungslücken durch Schlussfolgerung ausgeglichen, so dass die Stimmigkeit der Erinnerung erhalten bleibt. Dies kann zu Umdeutungen von vergangenen Ereignissen und Zuständen führen.
- Für die Beurteilung vergangener Ereignisse ist der Prozess der Verankerung und Anpassung von Bedeutung, wobei das aktuell verfügbare Wissen als Anker dient und hieraus der frühere Zustand

mittels plausibler Schlussfolgerungen erschlossen wird. Eine typische Erscheinung, die sich gut mit Hilfe dieses Prozesses erklären lässt, ist der Rückschaufehler.

- Bei der Evaluation vergangener affektiver Episoden wird die Dauer der Episode nur ungenügend einbezogen. Dies führt dazu, dass eine längere negative Episoden, welche eine kürzere negative Episode beinhaltet, positiver beurteilt als die eingebettete kürzere Episode.
- Das »Schnappschussmodell« der retrospektiven Evaluation affektiver Episoden kann diese Befunde erklären.

4. Wahrscheinlichkeitsurteile und statistische Schlussfolgerungen

Das folgende Kapitel über Wahrscheinlichkeitsurteile behandelt ein Gebiet, welches ausgiebig beforscht wurde und noch beforscht wird. Zudem gibt es zahlreiche Kontroversen bezüglich der Gültigkeit und der Interpretation der Ergebnisse.

Wir beginnen mit einer Darstellung des *Heuristics and Biases Ansatz*, der die Forschung in diesem Bereich wesentlich beeinflusst hat und bis heute als der wichtigste Ansatz gilt.

Es folgt ein Überblick über die Forschung zu Wahrscheinlichkeits-schlüssen. Im Zentrum dieser Untersuchungen befinden sich Schlussfolgerungen im Kontext des *Bayes-Theorems*. Wir behandeln hier auch Themen, die in engem Zusammenhang stehen, wie *Basis-Raten-Neglect* oder die fehlerhafte Interpretation bedingter Wahrscheinlichkeiten. Als nächstes wenden wir uns der Forschung zu, welche alternative Interpretationen der gefundenen Ergebnisse anbietet. Hierbei handelt es sich vor allem um evolutionspsychologische Argumente, welche die Ursachen für die Ergebnisse liefern sollen. Zudem werden zahlreiche Kritikpunkte am traditionellen Ansatz präsentiert und die Ergebnisse desselben als teilweise irrelevant oder artifiziell diskreditiert. Wir werden genau untersuchen, inwieweit diese Kritikpunkte zutreffend sind.

Als methodisches Zusatzkapitel folgt eine Erklärung der Logik statistischer Schlussfolgerungen. Im Speziellen wird die Debatte über Sinn oder Unsinn von *Nullhypothesentests*, die durchgängig in den Sozialwissenschaften angewendet werden, dargestellt.

In einem Anhang werden die wichtigsten Konzepte und Methoden der Wahrscheinlichkeitsrechnung zusammengefasst.

4.1 *Heuristics and Biases Ansatz von Tversky and Kahnemann*

Der Heuristics and Biases Ansatz (im Folgenden *H&B Ansatz* genannt) wurde in den 70er Jahren durch die Forschungen von Amos Tversky (1937-1996) und Daniel Kahneman (1934-) ins Leben gerufen. Für diese Forschung erhielt Kahneman im Jahre 2002 den Nobelpreis für Ökonomie.

Die grundlegende Idee des H&B Ansatzes besteht in der Annahme, dass viele unserer Wahrscheinlichkeitsurteile auf der Verwendung von *Heuristiken* beruhen. Hierbei handelt es sich um so genannte Daumenregeln, welche – je nach Bedingung – mit geringem Aufwand zu zufrieden stellenden Urteilen führen können oder aber auch zu sehr starken Verzerrungen (Biases).

Die Verwendung von Heuristiken kann als eine Realisierung des Konzepts der *beschränkten Rationalität* (»bounded rationality«) von Herbert Simon (1916-2001) betrachtet werden. Gemäss dieser Konzeption

sind wir aufgrund unserer beschränkten Informationsverarbeitungskapazität und aufgrund unvollständiger Information meist nicht in der Lage, optimale Entscheidungen und Urteile zu treffen. Aus diesem Grunde versuchen wir zufrieden stellende anstelle von optimalen Entscheidungen und Urteile zu treffen: Herbert nennt diesen Prozess *Satisficing* anstelle von *Optimizing*. Die Einsicht, dass wir in vielen Situationen nicht optimales Verhalten zeigen, hat in der Zwischenzeit auch in die Ökonomie Eingang gefunden, die sich dieser Erkenntnis lange verschlossen hatte.

Die Forschung von Tversky und Kahneman über Heuristiken und Biases befasst sich vor allem mit Urteilsfehlern, welche sich durch die Verwendung von Heuristiken ergeben (Obwohl sie auch Beispiele präsentieren, wo Heuristiken gute Ergebnisse liefern).

Die wichtigsten Heuristiken haben wir im Grunde bereits kennen gelernt. Wir fassen sie hier nochmals zusammen



Kognitiver Mechanismus 4-1: Heuristiken des Heuristics und Biases Ansatzes

Die klassische Heuristics und Biases Forschung umfasst die Wirkungsweise der folgenden drei zentralen Heuristiken:

1. Die Verfügbarkeitsheuristik (Availability heuristic):

Gemäss der Verfügbarkeitsheuristik basieren Wahrscheinlichkeits- bzw. Häufigkeitsurteile darauf inwieweit positive Instanzen für die zu beurteilenden Ereignisse generiert werden können: Je einfacher eine Instanz aus einer Ereignisklasse zu generieren ist, als desto häufiger/wahrscheinlicher wird die Ereignisklasse beurteilt.

2. Die Repräsentativitätsheuristik (Representativity heuristic):

Gemäss der Repräsentativitäts-Heuristik basiert die Beurteilung der Häufigkeit/Wahrscheinlichkeit eines Ereignisses auf der Ähnlichkeit oder Typikalität des Ereignisses für die gesamte Ereignisklasse: Je typischer eine Instanz für die Ereignisklasse, als desto höher wird die Wahrscheinlichkeit der Zugehörigkeit der Instanz zu dieser Ereignisklasse beurteilt.

3. Verankerung und Anpassung (Anchoring and Adjustment)

Die Methode der Verankerung und Anpassung wurde bereits im Zusammenhang mit Gedächtnisurteilen behandelt (vgl. Kognitiver Mechanismus 3-5). Bsp.3-4 präsentiert ein konkretes Beispiel der Anwendung der Prozedur im Rahmen der Beurteilung von Häufigkeiten.

Wie oben gezeigt, spielt die *Verfügbarkeit* bestimmter Informationen eine wichtige Rolle in der Anwendung dieser Prozedur.

Wir werden im Folgenden einige Beispiele für die Wirkungsweise der Verfügbarkeits- und der Repräsentativitätsheuristik betrachten.

4.1.1 Die Wirkungsweise der Verfügbarkeitsheuristik

Die Funktionsweise der Verfügbarkeitsheuristik als eine Methode zur Gewinnung von subjektiven Wahrscheinlichkeiten wurde von Tversky and Kahneman (1973) in einer Reihe von Experimenten demonstriert. Wir bringen zwei Beispiele:



Bsp.4-1: Wirkung der Verfügbarkeitsheuristik I (Tversky & Kahneman, 1973):

Betrachte die beiden Strukturen in Abb. 4-1. Durch welche gibt es nach Deiner Schätzung mehr unterschiedliche Wege? [Ein Weg ist eine Verbindung zwischen einer Raute in der obersten Zeile mit einer Raute in der untersten Zeile, wobei der Weg jeweils durch genau eine der Raute in jeder der dazwischen liegenden Zeilen führen muss (vgl. Abb. 4-1).]

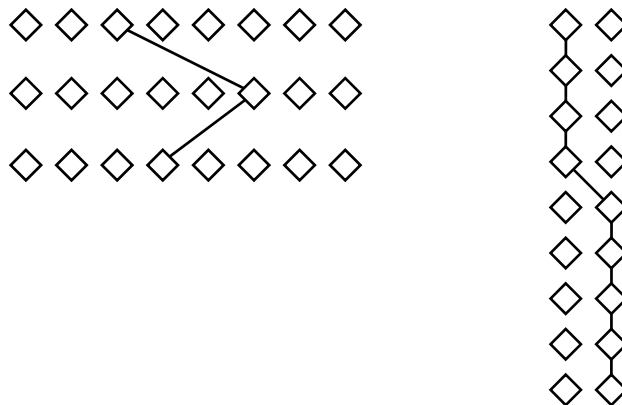


Abb. 4-1: Möglicher Wege durch zwei unterschiedliche Strukturen.

Die meisten Personen geben an, dass es vermutlich mehr unterschiedliche Wege in der linken Figur gibt (tatsächlich ist die Anzahl möglicher Wege für beide Figuren identisch: $8^3 = 2^9$).

Es scheint auf den ersten Blick ziemlich irrelevant zu sein, ob jemand abschätzen kann, wie viele verschiedene Wege es durch die Figuren gibt. Das Beispiel zeigt aber auf, wie wir bei der Lösung derartiger Beispiele vorgehen: Da es unmöglich ist, alle Wege zu testen, werden versuchsweise ein paar getestet. In der linken Figur ist es einfacher verschiedene Wege zu finden als in der rechten. Dies führt zu der Fehleinschätzung.

Ein interessanteres Beispiel ist das folgende:



Bsp.4-2: Wirkung der Verfügbarkeitsheuristik II (Tversky & Kahneman, 1973):

Versuchsmaterial:

2 Listen, A und B:

Liste A enthält 19 berühmte Frauen- und 20 alltägliche Männernamen.

Liste B enthält 19 berühmte Männer- und 20 alltägliche Frauennamen.

Durchführung:

Das Experiment wird als ein Gedächtnisexperiment deklariert in der die Aufgabe der Vp darin besteht, sich möglichst viele Namen zu merken.

Die Hälfte der Vpn erhält Liste A, die andere Hälfte Liste B. Die Namen werden in Zufallsreihenfolge präsentiert (jede Vpn enthält eine andere Reihung), wobei darauf geachtet wird, dass im Beginn oder Ende keine der Kategorien von Namen (berühmt oder nicht berühmt) gehäuft vorkommen.

Die Präsentation erfolgt, indem die einzelnen Namen in normaler Geschwindigkeit vorgelesen werden.

Am Ende muss die Person die Anzahl der Frauen- und Männernamen auf der Liste einschätzen.

Ergebnisse:

- Das Geschlecht mit den berühmten Namen wird als häufiger eingeschätzt (unabhängig davon, ob es sich um Männer oder Frauen handelt).
- Die Häufigkeit der unbekannteren Namen wird meist deutlich unterschätzt, während die Häufigkeit der berühmten Namen meist sehr nahe der korrekten Anzahl liegt.

Interpretation:

Die Bedeutung der Verfügbarkeitsheuristik für das Zustandekommen des Hauptergebnisses ist evident: Die berühmten Namen sind besser verfügbar und daher werden diese als häufiger vorkommend eingeschätzt.

Die Verfügbarkeit von Information wird durch die öffentlichen Medien deutlich mitbestimmt. Aufgrund der Selektivität in der Berichterstattung kann es zu Urteilsverzerrungen hinsichtlich der Auftretenshäufigkeit verschiedener Ereignisse kommen.



Bsp.4-3: Wirkung der Verfügbarkeitsheuristik: Beurteilung von Todesursachen

Entscheide für jedes Paar, welche der genannten Todesarten in den USA häufiger ist:

- Von herabfallenden Flugzeugteilen erschlagen oder von einem Hai angegriffen und getötet zu werden.

- Diabetis vs. Mord
- Autounfall vs. Magenkrebs

Antworten (Nach Death Odds, September 1990):

- Von fallenden Flugzeugteilen erschlagen zu werden, ist 30× wahrscheinlicher als von einem Hai getötet zu werden.
- An Diabetis oder Magenkrebs sterben ungefähr 2 × so viele Menschen wie an Mord oder Autounfällen.

Die realen Verhältnisse der Auftretenshäufigkeiten überraschen viele Menschen. Dies ist einfach mittels der Verfügbarkeitsheuristik zu erklären: Die »stillen Killer«, welche die meisten Tode verursachen, wie Herzinfarkt und Schlaganfall sind einfach keinen Zeitungsbericht Wert, während Autounfälle und Angriffe von Haien einen Platz in dem Medien finden.

4.1.1.1 IMAGINATION UND VERFÜGBARKEIT

In Abschnitt 3.1.1 wurde argumentiert, dass imaginative Techniken zu Verfälschungen der Erinnerungen führen können. Es liegt nahe, dass eine intensive Vorstellung des Eintretens eines Ereignisses die subjektive Wahrscheinlichkeit des Ereignisses erhöht, indem sie die Verfügbarkeit des Ereignisses erhöht. Diese Hypothese wurde in dem folgenden Experiment von Carroll (1978) getestet.



Bsp.4-4: Vorstellung und Verfügbarkeit (Carroll, 1978):

Das Experiment fand einen Tag vor der Amerikanischen Präsidentenwahl 1976 (Jimmy Carter gegen Gerald Ford) statt.

Es gab zwei Bedingungen, je nach Inhalt des sich vorzustellenden Szenarios:

1. Ford gewinnt die Wahlen, da Carter bestimmte Schlüsselstaaten nicht halten kann und Ford den mittleren Westen und den Westen gewinnen. Er gewinnt 316 Wahlstimmen und Carter 222. Weiters gewinnt Ford 32 der Staaten und Carter nur 18 plus den District of Columbia.
2. Carter gewinnt die Wahl da ihm seine Stärke im Süden und Westen einen uneinholbaren Vorsprung sichert, trotz des Vorsprungs von Ford im Westen. Carter gewinnt 342 Wahlmänner und Ford 196. Weiters gewinnt Carter in 28 Staaten plus den District of Columbia währende Ford 22 Staaten gewinnt.

Diese Szenarien basierten auf den letzten Umfragen. Die Vpn sollten sich die Szenarien möglichst intensiv vorstellen und sollten sich auch die Rede Siegers vorstellen sowie die Rede des Verlierers, der seine Niederlage eingesteht.

Anschliessend sollten die Vpn den Sieger der Wahlen vorhersagen.

Ergebnis:

Die Vorhersagen entsprachen den vorgestellten Szenarien: Personen, die sich Ford als Sieger vorstellten sagten mehrheitlich diesen als Sieger vorher und umgekehrt.

Das Ergebnis dieser Studie ist schwierig zu interpretieren, da zwischen den Bedingungen nicht nur die unterschiedlichen Vorstellungen variierten sondern den Personen auch alternative plausible Szenarien präsentiert wurden. Damit lässt sich die Bildung einer lebhaften Vorstellung nicht notwendig als Ursache identifizieren (vgl. die Ausführungen zur Konfundierung in Kapitel 2.6.2).

Eine andere Studie zeigt den Zusammenhang zwischen lebhafter Vorstellung und Verfügbarkeit besser auf, indem sie eine Zu- oder Abnahme der subjektiven Wahrscheinlichkeit demonstriert, je nach Schwierigkeit der Bildung einer Vorstellung.



Bsp.4-5: Schwierigkeit der Bildung einer Vorstellung und Verfügbarkeit (Sherman, Cialdini, Schwartzman, & Reynolds (1985):

In dem Experiment, an dem 120 weibliche Studentinnen teilnahmen, ging es um die Möglichkeit, sich mit einer (hypothetischen) Krankheit *Hyposcemia-B*, welche auf dem Campus der Arizona State University grassiert, anzustecken.

Es gab zwei Faktoren:

1. *Mit oder Ohne Vorstellungsinstruktion:*

In der Bedingung mit Vorstellung sollten die Vpn die Beschreibung der Symptome der Krankheit lesen und sich vorstellen, dass sie sich mit der Krankheit angesteckt hatten und während 3 Wochen an den geschilderten Symptomen leiden.

In der Bedingung ohne Vorstellung sollten die Vpn nur die Beschreibung lesen.

2. *Konkrete, einfach vorzustellende vs. abstrakte schwierig vorzustellende Symptome:*

In der Bedingung mit den einfach vorzustellenden Symptomen waren letztere konkret: Muskelschmerzen, Antriebslosigkeit, Anfälle von Kopfschmerzen.

Die schwierig vorzustellenden Symptome waren abstrakt und diffus: Orientierungslosigkeit, Fehlfunktion des Nervensystems, entzündete Leber.

Die Vpn wurden zufällig den 4 Bedingungen – gemäss der Kombination der beiden Faktoren – zugeteilt.

Nach Lesen und – je nach Bedingung – Vorstellung der Symptome mussten die Personen auf einer Skala von 1 (sehr wahrscheinlich) bis 10 (sehr unwahrscheinlich) ihre subjektive Wahrscheinlichkeit, mit der Krankheit angesteckt zu werden, angeben

Ergebnisse:

- Vpn in der Vorstellungsbedingung mit einfach vorstellbaren Symptomen schätzten die Wahrscheinlichkeit der Ansteckung als höher ein als die Vpn ohne Vorstellungsinstruktion (mit einfach vorstellbaren Symptomen): 5.25 vs. 6.20 (beachte die umgekehrte Kodierung der Skala)
- Vpn in der Vorstellungsbedingung mit schwierig vorstellbaren Symptomen schätzten die Wahrscheinlichkeit der Ansteckung als geringer ein gegenüber den Vpn ohne Vorstellungsinstruktion (mit schwierig vorstellbaren Symptomen): 7.70 vs. 6.55
- Die Leichtigkeit der Vorstellung der Symptome hatte in den Bedingung ohne Vorstellungsbildung nur einen ganz geringen Effekt: 6.20 vs. 6.55.

Interpretation:

Dieses Experiment zeigt sehr schön die Wirkung der Verfügbarkeit auf die Wahrscheinlichkeitsschätzung: Bei einfach vorstellbaren Symptome führt die konkrete Vorstellung zu einer Erhöhung der subjektiven Wahrscheinlichkeit, die Symptome zu bekommen, während bei schwer vorstellbaren Symptomen der umgekehrte Effekt zu beobachten ist.

Der gezeigte Einfluss der Vorstellungsbildung auf die Einschätzung der Wahrscheinlichkeit eines Ereignisses kann in Praxis fatale Folgen haben, wenn sich Personen in ihrer Vorstellung irgendwelche unrealistische Szenarien ausmalen und danach handeln. Dies kann vielleicht auch die vielen jährlichen Insolvenzen erklären, indem sich einige Personen schon als die neuen »Bill Gates«, »Steve Jobs« oder – meinetwegen – »Frank Sinatras« fühlen.

Ein anderer wichtiger Aspekt im Zusammenhang mit der Verfügbarkeit, der auch in der Praxis von grosser Bedeutung ist, betrifft die Tatsache, dass persönlich Erlebtem zu viel Stellenwert gegeben wird.

4.1.1.2 LEBHAFTIGKEIT, PERSÖNLICHES ERLEBEN UND VERFÜGBARKEIT

Die Verfügbarkeit selbst ist wiederum von bestimmten Faktoren abhängig, wie der Lebhaftigkeit, mit der ein Ereignis erlebt wurde, bzw. der Salienz bestimmter Ereignisse. Hierzu einige Beispiele:



Bsp.4-6: Eine Anekdote, »die das Leben schrieb« (nach Nisbett & Ross, 1980):

Herr Bürli möchte sich ein Auto kaufen und entscheidet sich für einen schwedischen Wagen, eine Saab oder Volvo. Da er nicht genau weiss, welche der beiden Marken solider ist, studiert er die einschlägigen Fachzeitschriften und Reparaturstatistiken. Aufgrund dieser Informationen, die auf Hunderten von Erfahrungen beruhen, entscheidet er sich für einen Saab.

Am Sonntagabend vor dem Kauf trifft er am Stammtisch seinen Freund Willi und erzählt ihm von seiner Entscheidung. Dieser reagiert mit Erstaunen. »Ich habe einen Bekannten, der hatte einen Saab. Zuerst waren die Bremsen defekt, dann ging die Einspritzpumpe kaputt und ständig hatte er Probleme mit der Elektronik. Nach 3 Jahren verkaufte er den Wagen für ein Butterbrot«. Aufgrund dieser Auskunft kauft Herr Bürli schliesslich einen Volvo.

Diese Geschichte zeigt in typischer Weise, wie Menschen selektiv Informationen erwerben und verarbeiten: Aufgrund einer einzigen lebhaften Schilderung wird Information, die auf Hunderten von Vorerfahrungen beruht, einfach ignoriert.

Menschen räumen lebhafter und persönlich erlebter Information einen höheren Stellenwert ein, als abstrakter, statistischer Information. So hat Spielbergs Film „Schindlers Liste“ keine grundlegend neuen Erkenntnisse über die Nazi-Greuel gebracht, trotzdem hat er die Menschen viel mehr bewegt als Statistiken, die Auskunft geben, wie viele Millionen Juden während der Nazi-Zeit ermordet wurden. Der Grund liegt darin, dass der Film konkrete Schicksale lebensnah miterleben lässt.

Ein typischer Ausdruck der Anfälligkeit für konkrete Information (gepaart mit Übergeneralisierung) ist die – auch unter Akademikern grassierende – Untugend, mittels Beispielen und Anekdoten zu argumentieren und sich durch Einzelbeispiele überzeugen zu lassen. Wer kennt nicht Argumente wie: »Mein Grossvater hat geraucht ‘wie ein Schlot’ und ist 90 Jahre alt geworden«. Und leider zu selten hört man die für derartige Argumentationen wohl einzig sinnvolle Antwort: *»Sicher, aber was willst du damit beweisen. Dies kann doch nicht den statistisch gesicherten Zusammenhang zwischen Rauchen und Krebs, sowie zwischen Rauchen und einer verkürzten Lebenserwartung widerlegen«.*

Das folgende abschliessende Beispiel demonstriert, wie die persönliche lebhafteste Anteilnahme zu irrigen Urteilen über die tatsächlichen politischen Kräfteverhältnisse führen kann.



Bsp.4-7: Lebhaftigkeit und die fehlerhafte Einschätzung von politischen Verhältnissen (Nisbett & Ross, 1980):

Bei Wahlen von 1972 behauptete eine grosse Anzahl von Reportern, die an den Wahlveranstaltungen von McGovern teilgenommen hatten, steif und fest (und natürlich irrigerweise), dass letzterer im Endergebnis nicht mehr als 10% hinter dem Gegenkandidaten rangieren würde, obwohl alle Umfragen anzeigten, dass McGovern mehr als 20%-Punkte im Rückstand war. Weiters wussten sie, dass die Umfragen in den letzten 24 Jahren noch nie mehr als 3% vom realen Ergebnis abgewichen sind.

Der Grund war, dass sie mit eigenen Augen 10000e von enthusiastischen McGowern Anhänger gesehen hatten

Derartige Überschätzungen findet man häufig bei Gruppen, die sich in der Öffentlichkeit besonders artikulieren.

Betrachten wir nun einige Beispiele, welche die zweite der beiden wichtigen Heuristiken betreffen.

4.1.2 Die Wirkungsweise der Repräsentativitätsheuristik

Einige Wirkungen der Repräsentativitätsheuristik, wie die fehlerhafte Beurteilung von Zufallssequenzen, wurde bereits im Abschnitt 2.3 behandelt).

Die klassische Studie, welche die Wirkung der Repräsentativitätsheuristik demonstriert, ist die Untersuchung zum so genannten Konjunktionsfehler.



Bsp.4-8: Der Konjunktionsfehler (Tversky & Kahneman, 1983):

Vpn: rekrutierten sich aus 3 Gruppen:

1. Undergraduates ohne statistische Kenntnisse
2. Graduates mit mittleren statistischen Kenntnissen
3. Graduates mit guten statistischen Kenntnissen

Stimuli und Durchführung:

1. Bill:

Die Vpn erhielten folgende Personen-Beschreibung und Instruktion:

Bill ist 34 Jahre alt. Er ist intelligent, aber phantasielos, zwanghaft und im Allgemeinen versteinert. In der Schule war er gut in Mathematik aber schlecht in den Human- & Sozialwissenschaften.

Bitte reihe die folgenden Aussagen nach ihrer Wahrscheinlichkeit, wobei »1« für »am wahrscheinlichsten« und »8« für »am unwahrscheinlichsten« steht.

Ergebnis:

Reihung	Aussage
(4.1)	Bill ist Physiker der als Hobby Pocker spielt
(4.8)	Bill ist Architekt
(1.1)	Bill ist Buchhalter (B)
(6.2)	Bill spielt in seiner Freizeit Jazz (J)
(5.7)	Bill surft als Hobby
(5.3)	Bill ist Reporter
(3.6)	Bill ist ein Buchhalter, der in seiner Freizeit Jazz spielt (B ∧ J)
(5.4)	Bill hat Bergsteigen als Hobby

2. Linda:

Personen-Beschreibung und Instruktion:

Linda ist 31 Jahre alt. Sie ist Single, verbal versiert und sehr geschickt. Als Studentin befasste sie sich mit Fragen der Diskriminierung und der sozialen Gerechtigkeit und nahm auch an Antikernkraft-Demonstrationen teil.

Bitte reihe die folgenden Aussagen nach ihrer Wahrscheinlichkeit, wobei »1« für »am wahrscheinlichsten« und »8« für »am unwahrscheinlichsten« steht.

Ergebnis:

Reihung	Aussage
(5.2)	Linda ist Lehrerin in der Elementarschule
(3.3)	Linda arbeitet in einem Buchgeschäft und nimmt Jogakurse
(2.1)	Linda ist aktiv in der Frauenbewegung (F)
(3.1)	Linda ist psychiatrische Sozialarbeiterin
(5.4)	Linda ist Mitglied in der Liga für weibliche Wähler
(6.2)	Linda ist Bankangestellte (B)
(6.4)	Linda verkauft Versicherungen
(4.1)	Linda ist Bankangestellte und aktiv in der Frauenbewegung (B ∧ F)

In beiden Fällen begehen die Vpn den so genannten *Konjunktionsfehler*: Sie stufen das zusammengesetzte Ereignis als wahrscheinlicher ein als eines der Komponenten. Im Falle von Bill wird **B ∧ J** als wahrscheinlicher eingeschätzt als **J** und im Falle von Linda wird **B ∧ F** als wahrscheinlicher eingeschätzt als **B**.

Wie folgende Tabelle zeigt, hat der Grad an statistischer Ausbildung hatte wenig Einfluss auf den Fehler:

Tab. 4-1: *Prozentsatz der Reihungen des konjunktiven Ereignisses vor Elementarereignissen für die verschiedenen Versuchsgruppen und Szenarien*

	Naive Vpn		Mittlerer Level		Hoher Level	
	Linda	Bill	Linda	Bill	Linda	Bill
	89%	93%	90%	86%	85%	83%
<i>N</i>	88	94	53	56	32	32

Interpretation:

Die präsentierten Personen-Beschreibungen sind repräsentativ für einen Buchhalter bzw. eine aktive Feministin aber nicht repräsentativ für Jazzspieler bzw. Bankangestellte. Die Vpn reihen das zusammengesetzte Ereignis in der Mitte zwischen den beiden Einzelereignissen.

Die Wahrscheinlichkeit wird also aufgrund der Ähnlichkeit bestimmt und nicht auf Grund der logischen Beziehung zwischen den Mengen, wie es die Wahrscheinlichkeitsaxiome (vgl. Anhang) vorschreiben.

Eine Befragung der Vpn am Ende ergab, dass sie das statistische Prinzip kennen und akzeptieren (mit Ausnahme von 2 Vpn), und dass sie ihr Urteil revidierten.

Möglicher Einwand:

Gegen die Ergebnisse könnte der folgenden – m. E. ziemlich weit hergeholt – Einwand vorgebracht werden:

Die Personen habe aufgrund des präsentierten Materials die Aussagen:

Linda ist Bankangestellte bzw. Bill spielt in seiner Freizeit Jazz

Wie folgt interpretiert:

Linda ist Bankangestellte, die nicht aktiv in der Frauenbewegung ist bzw.

Bill spielt in seiner Freizeit Jazz und ist kein Buchhalter

Um diesen Einwand auszuschliessen, wurde ein Experiment durchgeführt, bei dem nur eine der Items (entweder das konjunktive oder das Einzelereignis) vorhanden war.

In allen Fällen wurde das konjunktive Ereignis als wahrscheinlicher gereiht.

In einer weiteren Untersuchung testeten Tversky und Kahneman die Wirkung der Repräsentativität auf Vorhersagen. Hierbei handelte es sich um Repräsentativität im Sinne von »Konsequenz *X* ist repräsentativ für Ursache *M*«.



Bsp.4-9: Der Konjunktionsfehler und die Vorhersage von Ereignissen (Tversky & Kahneman, 1983):

Die untenstehenden Aussagen sind nach ihrer Wahrscheinlichkeit zu reihen (Vorhersage für das Jahr 1981):

Ergebnis:

Reihung	Aussage
(1.5)	Reagan wird die staatlichen Zuschüsse für lokale Verwaltungen kürzen (A)
(3.3)	Reagen wird eine staatliche Unterstützung für unverheiratete Mütter einführen (B)
(2.7)	Reagan wird das Verteidigungsbudget um weniger als 5% erhöhen. Reagan wird die staatlichen Zuschüsse für lokale Verwaltungen kürzen und eine staatliche Unterstützung für unverheiratete Mütter einführen
(2.9)	(A∧B)

Der Konjunktionsfehler wurde von 68% der Vpn gemacht.

Diese Beispiele geben Anlass zu folgender Warnung:

Hüte dich vor detaillierten, intern kohärenten und plausiblen Szenarios (sowohl was die Zukunft als auch was die Vergangenheit betrifft).

Je detaillierter ein Szenario, desto plausibler erscheint es uns, aber desto unwahrscheinlicher ist es. Daher ist ein bruchstückhaftes Szenario oft wahrscheinlicher.

Zusammenfassend kann man also sagen, dass Personen ihre Wahrscheinlichkeitsurteile auf der Basis von Ähnlichkeit oder Typikalität (Überbegriff: *Repräsentativität*) bilden, wobei sie die Axiome der Wahrscheinlichkeitstheorie verletzen.

Nach diesem Überblick über die Wirkungen der beiden zentralen Heuristiken wenden wir und einem anderen Problem zu, nämlich des fehlerhaften Umgangs mit Wahrscheinlichkeiten.

4.2 Fehlerhafter Umgang mit Wahrscheinlichkeiten

Im Folgenden behandeln wir vier Aspekte des fehlerhaften Umgangs mit Wahrscheinlichkeiten: (1) Überbewertung/Unterbewertung kleiner/grosser Wahrscheinlichkeiten, (2) das Phänomen des so genannten Probability Matchings, (3) die fehlerhafte Beurteilung bedingter Wahrscheinlichkeiten und (4) das Ignorieren von Basisraten.

4.2.1 Subjektiver Gewichtung von Wahrscheinlichkeiten

Die meisten Personen überschätzen kleine und überschätzen hohe Wahrscheinlichkeiten. Die ist in Abb. 4-2 illustriert, welche die sub-

jektive Wahrscheinlichkeiten in Abhängigkeit von der objektiven darstellt.

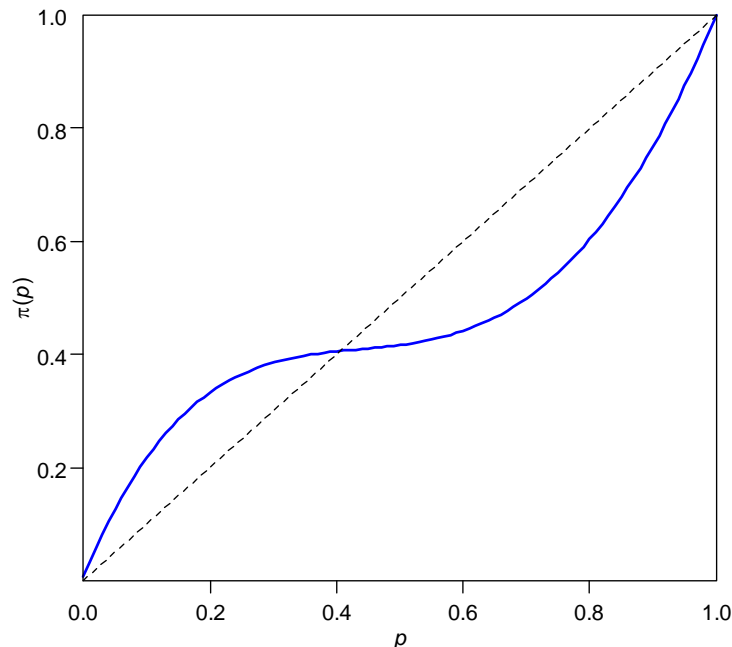


Abb. 4-2: *Subjektives Gewicht $\pi(p)$ als Funktion der aktuellen Wahrscheinlichkeit p (nach Tversky & Kahneman, 1992).*

Die Überschätzung von kleinen Wahrscheinlichkeiten mag mit verantwortlich dafür sein, dass Leute an Lotterien teilnehmen mit astronomisch geringen Gewinnwahrscheinlichkeiten (von mehr als 10^{-8} für den Hauptgewinn).

Eine weitere Folge besteht in der Tendenz, ungünstige Versicherungen zu akzeptieren. So haben z.B. die meisten Finanzprodukte mit Versicherung eine schlechte Performance (abgesehen davon, dass – wie im Zuge der Lehman-Pleite offensichtlich wurde – Versicherungen auch keinen perfekten Schutz bieten, was jedoch die meisten Anleger nicht wussten bzw. ignorierten).

Die Tendenz der Überschätzung kleiner Risiken ist vor allem für Personen in leitenden Positionen (Politiker, CEOs) bedeutsam: *Die Reduktion eines Risikos um – sagen wir 1% – hat grundsätzlich den gleichen Nutzen, unabhängig davon, ob es sich um eine Reduktion von 50% auf 49% handelt oder um eine Reduktion von 1% auf 0%.* Oftmals ist jedoch erstere Reduktion mit geringeren Kosten verbunden als letztere. Daher kann man oft mit viel geringerem Aufwand eine Verbesserung erzielen, indem ein mässig hohes Risiko weiter verringert wird anstatt ein kleines Risiko auf Null zu reduzieren.

4.2.2 Probability Matching



Kognitiver Mechanismus 4-2: Probability Matching

Probability Matching beschreibt die Tendenz, vorgegebene Wahrscheinlichkeiten von Ereignissen in den Antworten widerzuspiegeln.

Die Strategie des Probability Matching ist im Allgemeinen eine nicht optimale Strategie.

Das Phänomen wurde in einfachen Lernexperimenten in den 60er-Jahren ausgiebig untersucht. Diese Experimente haben folgende prototypische Struktur: Es gibt zwei Ereignisse (z.B. es leuchtet entweder eine rote oder eine grüne Lampe auf), die mit unterschiedlicher Wahrscheinlichkeit auftreten (z.B. 70% vs. 30%). Die Aufgabe der Person besteht darin, möglichst fehlerlos vorherzusagen, welche Lampe als nächstes aufleuchten wird. Sodann erhält sie ein Feedback, d.h. eine der Lampen leuchtet auf.

Die meisten Vpn spiegeln in ihren Antworten die Auftretenswahrscheinlichkeiten wider, d.h. sie sagen in 70% der Fälle vorher, dass die rote Lampe aufleuchten wird und in 30% der Fälle, dass die grüne Lampe aufleuchtet.

Die Strategie des Probability Matching ist nicht die optimale Strategie, welche den erwarteten Fehler minimiert. Die optimale Strategie besteht in der permanenten Wahl des am häufigsten auftretenden Ereignisses, im aktuellen Fall ist dies das Ereignis, dass die rote Lampe aufleuchtet. Die Tatsache, dass Probability Matching eine suboptimale Strategie darstellt, lässt sich leicht am aktuellen Beispiel demonstrieren.

Tab. 4-2 zeigt die Verbundwahrscheinlichkeiten der beiden Ereignisse: (a) *Experimenteller Outcome* und (b) *Antwort der Vp*, für die aktuellen Wahrscheinlichkeiten (70% [rot] vs. 30% [grün]) und bei Anwendung der Strategie des Probability Matchings durch die Versuchsperson.

Outcome	Antwort der Vp		Σ
	»Rote Lampe«	»Grüne Lampe«	
Rote Lampe	0.49	0.21	0.70
Grüne Lampe	0.21	0.09	0.30
Σ	0.70	0.30	

Tab. 4-2: Verbundverteilung über die beiden Variablen: *Experimenteller Outcome* und *Antwort der Versuchsperson (Vp)*.

Die Einträge in der Tabelle ergeben sich aufgrund folgender Überlegung: Das Ereignis, welche Lampe aufleuchtet ist – da rein zufällig – stochastisch unabhängig von der Antwort der Versuchsperson. Damit ergibt sich die Verbundverteilung der beiden Ereignisse aus dem Pro-

dukt der beiden Randverteilungen [Vgl. Anhang, Abschnitt 2.1.3.1, Gleichung (2-3)]:

$$\begin{aligned} &P(\text{Aufleuchten der Lampe, Antwort der Vp}) \\ &= P(\text{Aufleuchten der Lampe}) \times P(\text{Antwort der Vp}) \end{aligned}$$

So gilt z.B.:

$$P(\text{Lampe} = \text{rot}, \text{Antwort} = \text{"rot"}) = 0.7 \times 0.7 = 0.49$$

Eine korrekt Antwort liegt vor, falls die Antwort mit der Farbe der Lampe übereinstimmt, d.h. die Vp sagt »rot« und die rote Lampe leuchtet oder die Vp sagt »grün« und die grüne Lampe leuchtet. Da es sich hier um disjunkte Ereignisse handelt, können die zugehörigen Wahrscheinlichkeiten addiert werden und es ergibt sich:

$$\begin{aligned} P(\text{Antwort korrekt}) &= P(\text{Lampe} = \text{rot}, \text{Antwort} = \text{"rot"}) \\ &\quad + P(\text{Lampe} = \text{grün}, \text{Antwort} = \text{"grün"}) \\ &= 0.49 + 0.09 = 0.58 \end{aligned}$$

Die Wahrscheinlichkeit einer korrekten Antwort liegt daher im Falle von Probability Matching deutlich unterhalb jener, welche sich bei Anwendung der optimalen Strategie, nämlich der konstanten Wahl des am häufigsten auftretenden Ereignisses, ergibt: 0.70.



Bsp.4-10: Probability Matching bei Ratten und Studenten (Gallistel, 1990):

Gallistel (1990, Seite 351-352) berichtet von einer Demonstration mit Studenten, der er 1960 in Yale beiwohnte.

Eine Ratte wurde trainiert ein T-Labyrinth zu durchlaufen. Das Futter befand sich in 75% der Fälle in dem einen Zweig des Labyrinths und mit 25% im anderen Zweig.

Die Studenten des Anfängerkurses beobachteten das Tier und machten selbst in jedem Durchgang ihre Vorhersagen, wobei ein Licht (unsichtbar für die Ratte), anzeigte, in welchem Zweig sich das Futter befand.

Am Ende lernte die Ratte, immer den Zweig mit 75% Futter aufzusuchen. Die Studenten hingegen zeigten fast perfektes Probability Matching: Der durchschnittliche Prozentsatz für den 75%-Zweig lag nur ein oder zwei Prozentpunkte von 75 entfernt.

Die Studenten mussten daher zu ihrer Überraschung feststellen, dass das Verhalten der Ratte intelligenter war als ihr eigenes.

Bemerkung:

Unter der gleichen Bedingung wie die Studenten – nämlich mit Feedback – hätte die Ratte ebenso wie die Studenten Probability Matching gezeigt.

Beachte:

Das an- oder abwesende Futter war kein explizites Feedback, da die Ratte ja nicht wissen konnte, was sich im anderen Zweig des Labyrinths befand.

Die Strategie des Probability Matchings ist kein experimentelles Artefakt. Es wird sowohl von Menschen wie von Tieren durchgeführt. Hierzu zwei Beispiele aus dem Tierreich (Gallistel, 1990).

*Bsp.4-11: Probability Matching bei Tieren (Gallistel, 1990):*

Smith und Dawkins (1971) studierten das Jagdverhalten von grossen Meisen und erlaubten einzelnen Tieren auf verschiedenen Flecken mit unterschiedlicher (experimentell genau festgelegter) Ausbeute zu jagen.

Die Jagdperioden waren jedoch so kurz, sodass keine substantielle Abnahme der Beute in den verschiedenen Gebieten entstand.

Man könnte nun annehmen, die Tiere würden immer das Gebiet mit der grössten Ausbeute aufsuchen. Dies war nicht der Fall, *sondern die relative Häufigkeit, mit der ein Gebiet aufgesucht wurde, entsprach ungefähr der relativen Ausbeute.*

Harper (1982) führte ein Experiment mit einer Schar von 31 wild lebenden Enten durch.

Während mehrerer aufeinander folgender Tage gingen 2 Experimentatoren, jeder ausgestattet mit einem Sack von geschnittenen 2-Gramm Brotstückchen, zum See. Sie positionierten sich ca. 20 Meter getrennt am Rand des Sees und begannen die Brotstückchen in den See zu werfen.

Die relative Wurf-Rate der beiden Experimentatoren wurde zufällig bestimmt, so dass sie nicht vorhersagbar war.

Ergebnis:

- Am Beginn des Experiments versammelten sich die Enten vor den beiden Werfern, wobei die Aufenthaltsdauer der letzten Wurfrate entsprach.
- Innerhalb von 1 Minute (in welcher Zeit nur 12 bis 18 Stückchen ins Wasser geworfen worden waren und die meisten Enten noch kein Stückchen erhalten hatten) positionierten sich die Enten neu, so dass deren relative Verweildauer vor den beiden Werfern der aktuellen Wurfrate entsprach.

- In manchen Durchgängen am Ende wurden die Brotstückchen mit gleicher Rate geworfen, jedoch waren die Stückchen des einen Werfers doppelt so gross wie jene des anderen. In diesem Fall verteilten sich die Enten zuerst gleich über die beiden Werfer, aber nach 5 - 6 Minuten passten sie ihre Verteilung dem Produkt aus Häufigkeit und Grösse der Stückchen an.

4.2.2.1 ERKLÄRUNG VON PROBABILITY MATCHING

Die Tendenz zum Probability Matching im Tierreich scheint eine sinnvolle Strategie zu sein, wenn man die Konkurrenz um Futter einbezieht. Würde ein Tier nur den Platz mit dem grössten Ertrag aufsuchen, so würde dies Konkurrenten anlocken, was zu einer Verknappung des Futters führen würde. In diesem Fall wären dann jene Tiere bevorzugt, welche Plätze mit geringerem Ertrag aufsuchten. Die Maximierungsstrategie des Aufsuchens von Plätzen mit dem grössten Ertrag ist daher *keine evolutionär stabile Strategie*.

Die Strategie der Abstimmung der Verweildauer mit dem Anteil an Ertrag oder der Rate des Aufsuchens eines Platzes mit der Ertragsrate scheint unter den Bedingungen der Konkurrenz eine weitaus sinnvollere Strategie, weil sie es den Tieren ermöglicht, auch bei grosser Konkurrenz, sowie bei sich ändernden Verhältnissen, zu ihrem Anteil an Futter zu gelangen.

Diese Erklärung scheint auch für den Menschen zutreffend. Betrachtet man nämlich die verschiedenen möglichen Ereignisse als Quellen um im Experiment Pluspunkte zu sammeln, so scheinen die Vpn versucht, auch die weniger ertragreiche Quelle »anzuzapfen«. Ihnen entgeht dabei, dass sie hierfür die ertragreichere Quelle vernachlässigen müssen. Man beachte, dass, was für Tiere unter Konkurrenzbedingungen als durchaus sinnvolle Strategie erscheint, für Menschen mit fehlender Konkurrenz keine optimale Strategie darstellt.



Bemerkung

Wie bei fast allen Fällen von Urteilsfehlern, gibt es zahlreiche Versuche, das Verhalten als sinnvoll und völlig rational zu interpretieren. Ich möchte hier nicht auf die einzelnen – mir bekannten – Argumente eingehen, da ich sie alle für falsch und gekünstelt halte.

Ein anderer interessanter Aspekt betrifft das Auftreten von individuellen Unterschieden beim Probability Matching.

4.2.2.2 INDIVIDUELLE UNTERSCHIEDE BEIM PROBABILITY MATCHING

Stanovich und West (2000) führten zahlreiche Untersuchungen zur Beantwortung der Frage nach individuellen Unterschieden beim Urteilen durch. Im Spezifischen interessierte Sie die Frage, ob intelligentere Personen weniger Urteilsfehler machen.

Bezüglich der Anwendung der Strategie des Probability Matchings fanden West und Stanovich (2003) klare individuelle Unterschiede: Personen, welche die optimale Strategie verwenden, sind im Durchschnitt kognitiv leistungsfähiger (gemessen mittels SAT [Scholastic Aptitude Test]) gegenüber solchen, welche die Strategie des Probability Matchings verwenden.

Interessanterweise gibt es auch einen deutlichen Unterschied zwischen den Geschlechtern: Frauen verwenden häufiger die Strategie des Probability Matchings als Männer. So verwendeten z.B. im 3. Experiment von West und Stanovich (2003) 41% der Männer ($N = 165$) die optimale Strategie und 59% Probability Matching. Bei den Frauen ($N = 232$) betragen die entsprechenden Zahlen: 28% und 72%. Insgesamt gab es in allen drei Experimenten der Studie deutliche Geschlechtsunterschiede (alle in die gleiche Richtung). Die Ursachen für diese Unterschiede sind nicht geklärt.

4.2.3 Umgang mit bedingten Wahrscheinlichkeiten

Die bedingte Wahrscheinlichkeit $P(A|B = b)$ gibt die Wahrscheinlichkeit bzw. die Verteilung der Zufallsvariable A unter der Voraussetzung, dass die Variable B den Wert b aufweist. Die bedingte Wahrscheinlichkeitsverteilung repräsentiert daher die Verteilung der Werte der Zufallsvariable A nicht mehr in Bezug auf die Gesamtpopulation sondern in Bezug auf die eingeschränkte Teilpopulation, deren Einheiten alle die Bedingung $B = b$ erfüllen.



Bemerkung

Die gegebene Charakterisierung gilt im Rahmen der frequentistischen Konzeption von Wahrscheinlichkeit, in welcher der Begriff »Population« Sinn macht (vgl. Anhang).

Die bedingte Wahrscheinlichkeit wird zur Charakterisierung des Konzepts der (bedingten) stochastischen Unabhängigkeit verwendet (vgl. Anhang, Konzept 2-5 und Konzept 2-6), wobei gilt: Zwei Zufallsvariablen A und B sind stochastisch unabhängig, falls gilt:

$$P(A|B = b) = P(A), \text{ für alle Werte } b \text{ von } B. \quad (4-1)$$

Die zentrale Aussage von Gleichung (4-1) besagt, dass sich die Verteilung der Werte der Zufallsvariablen A nicht ändert, gleichgültig, welche Werte Variable B annimmt. In diesem Fall ist B irrelevant für die Kenntnis der Verteilung von A . Das Konzept der stochastischen Unabhängigkeit repräsentiert genau diese Irrelevanz.



Bsp.4-12: Stochastische Unabhängigkeit

Gegeben:

Variable T repräsentiere ein Trainingsprogramm für Legastheniker mit zwei Ausprägungen:

$t \Leftrightarrow$ Training angewendet

$\bar{t} \Leftrightarrow$ kein Training

Variable V repräsentiere die Tatsache, ob eine Verbesserung eintrat oder nicht:

$v \Leftrightarrow$ Verbesserung vorhanden

$\bar{v} \Leftrightarrow$ Verbesserung abwesend

Falls gilt: $P(V|T = t) = P(V|T = \bar{t}) = P(V)$, d.h. die beiden Variablen sind stochastisch unabhängig, so hat das Training offensichtlich keine Wirkung.

4.2.3.1 ASYMMETRIE BEDINGTER WAHRSCHEINLICHKEITEN

Bedingte Wahrscheinlichkeiten sind asymmetrisch, d.h. es gilt im Allgemeinen, dass:

$$P(A|B) \neq P(B|A)$$

Dies ist einfach mit Hilfe von Venn-Diagrammen einzusehen, wenn man sich dabei die Definition der bedingten Wahrscheinlichkeit vor Augen hält:

$$P(A|B) = \frac{P(A, B)}{P(B)} \quad \text{und} \quad P(B|A) = \frac{P(A, B)}{P(A)}$$

Man beachte, dass für beide bedingte Wahrscheinlichkeiten der Zähler des Quotienten identisch ist, jedoch sind die Nenner verschieden. Die Asymmetrie bedingter Wahrscheinlichkeiten lässt sich nun sehr einfach demonstrieren



Bsp.4-13: Asymmetrie bedingter Wahrscheinlichkeiten:

Gegeben:

K repräsentiere die Menge der Personen, die an Lungenkrebs erkrankt sind.

R repräsentiere die Population der Raucher.

Das Venn-Diagramm in Abb. 4-3 veranschaulicht die beiden Populationen der Raucher und der an Lungenkrebs Erkrankten. Zentral ist hierbei die relative Grösse der beiden Populationen: Das Diagramm drückt aus, dass die Population der Raucher um einiges grösser ist als jene der an Lungenkrebs Erkrankten.

Die bedingte Wahrscheinlichkeit:

$$P(K|R) = \frac{P(K \wedge R)}{P(R)}$$

entspricht dem Verhältnis der gemeinsamen Fläche der beiden Kreise geteilt durch die Fläche des Kreises, welcher die Raucher repräsentiert.

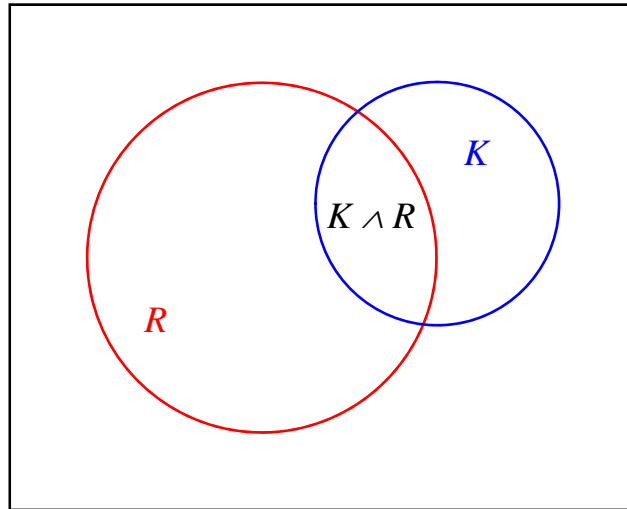


Abb. 4-3: Venn-Diagramm zur Illustration der Asymmetrie bedingter Wahrscheinlichkeiten: $R = \text{Raucher}$, $K = \text{Krebskranke}$.

Analog ergibt sich die bedingte Wahrscheinlichkeit:

$$P(R|K) = \frac{P(K \wedge R)}{P(K)}$$

aus dem Verhältnis der Schnittfläche der beiden Kreise geteilt durch die Fläche des Kreises, welcher die an Lungenkrebs Erkrankten repräsentiert.

Da die Population der an Lungenkrebs Erkrankten viel geringer ist als die Population der Raucher (bzw. die Fläche des die Krebskranken repräsentierenden Kreises kleiner ist als die Fläche des die Raucher darstellenden Kreises) ergibt sich die Ungleichung:

$$P(R|K) > P(K|R).$$

Die Wahrscheinlichkeit einen Raucher innerhalb der Population der Krebskranken zu finden ist also weitaus grösser als die Wahrscheinlichkeit einen Krebskranken innerhalb der Population der Raucher zu finden.



Bemerkung zur Notation:

Die Symbole $P(K,R)$ und $P(K \wedge R)$ repräsentierten beide die gleiche Wahrscheinlichkeit, nämlich die Verbundwahrscheinlichkeit von Rauchen und Lungenkrebs.

Eddy (1982) berichtet, dass die meisten Ärzte – wie die meisten Laien – den Unterschied zwischen den beiden bedingten Wahrscheinlichkeiten nicht berücksichtigen. So verwechseln die meisten die beiden bedingten Wahrscheinlichkeiten:

$$P(\text{Krankheit} = \text{vorhanden} | \text{Diagnose} = \text{positiv})$$

und

$$P(\text{Diagnose} = \text{positiv} | \text{Krankheit} = \text{vorhanden}).$$

Dies kann zu fatalen Fehlern führen, da letztere Wahrscheinlichkeit in der Regel weitaus höher ist als die erste. Werden die Arten von bedingten Wahrscheinlichkeiten verwechselt, so kann das zur Einleitung einer möglicherweise invasiven Behandlung führen, die aufgrund der korrekten Wahrscheinlichkeit nicht gerechtfertigt ist.

4.2.3.2 DER EINFLUSS VON KAUSALEN UND DIAGNOSTISCHEN SCHLÜSSEN AUF DIE BEURTEILUNG BEDINGTER WAHRSCHEINLICHKEITEN

Für die Beurteilung von bedingten Wahrscheinlichkeiten spielen kausale Überlegungen eine Rolle: Personen fokussieren vorwiegend auf den kausalen Einfluss eines Ereignisses auf ein zukünftiges und vernachlässigen die Rolle der diagnostischen Bedeutung eines späteren Faktums für die Beurteilung eines früheren Ereignisses.

Betrachten wir hierzu ein Beispiel von Tversky & Kahneman (1982a).



Bsp.4-14: Kausalität und die Beurteilung bedingter Wahrscheinlichkeiten (Tversky & Kahneman, 1982a):

Welche der folgenden beiden Wahrscheinlichkeiten ist grösser:

- (i) Die Wahrscheinlichkeit, dass innerhalb der nächsten 5 Jahre ein Gesetz zur Quecksilberverunreinigung durch Verbrennung von Kohle erlassen wird, wenn die Zahl der nachgewiesenen Toten aufgrund von Quecksilber innerhalb der nächsten 5 Jahre die Zahl 500 überschreitet.
- (ii) Die Wahrscheinlichkeit, dass innerhalb der nächsten 5 Jahre ein Gesetz zur Quecksilberverunreinigung durch Verbrennung von Kohle erlassen wird, wenn die Zahl der nachgewiesenen Toten aufgrund von Quecksilber innerhalb der nächsten 5 Jahre die Zahl 500 *nicht* überschreitet.

Welche der folgenden beiden Wahrscheinlichkeiten ist grösser:

- (iii) Die Wahrscheinlichkeit, dass die Zahl der nachgewiesenen Toten aufgrund von Quecksilber innerhalb der nächsten 5 Jahre die Zahl 500 überschreitet, wenn innerhalb der nächsten 5 Jahre ein Gesetz zur Quecksilberverunreinigung durch Verbrennung von Kohle erlassen wird.
- (iv) Die Wahrscheinlichkeit, dass die Zahl der nachgewiesenen Toten aufgrund von Quecksilber innerhalb der nächsten 5 Jahre die Zahl 500 überschreitet, wenn innerhalb der nächsten 5 Jahre *kein* Gesetz zur Quecksilberverunreinigung durch Verbrennung von Kohle erlassen wird.

Ergebnis:

Die meisten Personen (140 von 166) schätzen die Wahrscheinlichkeit (i) höher ein als (ii) und gleichzeitig (iv) höher als (iii).

Analyse:

Es sei:

G = Gesetz wird innerhalb der nächsten 5 Jahre beschlossen,

\bar{G} = Gesetz wird innerhalb der nächsten 5 Jahre nicht beschlossen.

T = mehr als 500 Tote aufgrund von Quecksilber innerhalb der nächsten 5 Jahre.

\bar{T} = weniger als 500 Tote aufgrund von Quecksilber innerhalb der nächsten 5 Jahre.

Das modale Antwortmuster [(i) > (ii) und (iv) > (iii)] sieht dann wie folgt aus:

$$P(G|T) > P(G|\bar{T}) \text{ und } P(T|G) < P(T|\bar{G}).$$

Dieses Muster widerspricht der Wahrscheinlichkeitstheorie, denn es muss gelten (Übung 4-3):

$$P(G|T) > P(G|\bar{T}) \Leftrightarrow P(T|G) > P(T|\bar{G}).$$

Interpretation:

Das modale Antwortmuster entspricht den Schlussfolgerungen:

- (i) Viele Tote aufgrund von Quecksilber veranlassen den Gesetzgeber dazu, ein Gesetz zur Regulierung des Quecksilberausstosses zu beschliessen.

- (ii) Ein Gesetz zur Regulierung des Quecksilberausstosses verhindert den Tod von Personen aufgrund von Quecksilber.

Diese Art von Schlussfolgerung übersieht jedoch die diagnostische Implikation, *dass das Vorliegen eines Gesetzes auf wenige Tote bzw. die Abwesenheit eines Gesetzes auf viele Tote durch Quecksilber hinweist* (und umgekehrt). Im Speziellen gilt:

1. Sieht man T als Ursache für G , so sollte $P(G|T)$ hoch sein, indem viele Tote zum Erlass eines Gesetzes führen. Interpretiert man andererseits T als diagnostischen Hinweis für das Vorliegen eines Gesetzes, so sollte $P(G|T)$ gering sein, da ein Gesetz das Auftreten vieler nachfolgender Toter verhindert haben sollte.

Somit: $P(G|T)$ ist hoch, wenn T als Ursache von G gesehen wird und gering, wenn T als diagnostischer Hinweis für G betrachtet wird.

2. Analog gilt für $P(G|\bar{T})$: Falls \bar{T} als Hinweis für das Vorliegen von G betrachtet wird, so ist $P(G|\bar{T})$ hoch. Wird hingegen \bar{T} (bzw. T) als mögliche Ursache für G gesehen, so sollte $P(G|\bar{T})$ gering sein.

Somit: $P(G|\bar{T})$ ist gering, wenn T als Ursache von G betrachtet wird und hoch, falls \bar{T} als diagnostischer Hinweis auf das Vorliegen von G betrachtet wird.

Die Tatsache, dass $P(G|T) > P(G|\bar{T})$ spricht dafür, dass die Personen die kausale Interpretation bevorzugen.

3. $P(T|\bar{G})$ sollte – bei kausaler Interpretation von G – hoch sein, indem die Abwesenheit eines Gesetzes zu vielen nachfolgenden Toten führt. Andererseits sollte bei diagnostischer Interpretation von \bar{G} $P(T|\bar{G})$ gering sein, da die Abwesenheit eines Gesetzes eher auf wenige Tote in der Vergangenheit hinweist.

Die kausale Interpretation von G führt also zu einer hohen Wahrscheinlichkeit $P(T|\bar{G})$ und die diagnostische zu einer geringen.

4. $P(T|G)$ ist bei kausaler Interpretation von G gering, da das Vorliegen eines Gesetzes viele Tote verhindert. Bei diagnostischer Interpretation von G als Hinweis auf viele Tote in der Vergangenheit ist $P(T|G)$ hoch.

Die kausale Interpretation von G führt also zu einer geringen Wahrscheinlichkeit $P(T|G)$ und die diagnostische zu einer hohen.

Die Tatsache, dass $P(T|G) < P(T|\bar{G})$ spricht wiederum dafür, dass die Personen die kausale Interpretation bevorzugen.

Das Beispiel macht eine zentrale Differenz zwischen dem menschlichen Urteiler und dem Wahrscheinlichkeitskalkül deutlich: Für den Menschen ist Kausalität eine entscheidende Determinante in der Beurteilung von Wahrscheinlichkeiten.

Innerhalb des Wahrscheinlichkeitskalküls ist die Kausalrichtung völlig irrelevant. Hier spielt nur der Grad der Assoziation zwischen den Variablen eine Rolle. Die diagnostische Implikation eines späteren Ereignisses auf das Vorliegen einer früheren Ursache hat eine gleich-grosse Bedeutung wie das Auftreten einer Ursache für ein später eintretendes Ereignis.

4.2.3.3 DAS PROBLEM NONMONOTONER WAHRSCHEINLICHKEITEN

Im Zusammenhang mit bedingten Wahrscheinlichkeiten soll hier noch kurz ein fundamentaler Aspekt, welche das Schlussfolgern mit Wahrscheinlichkeitsinformation betrifft, angesprochen werden.



Prinzip 4-1: *Fehlende Monotonie von Wahrscheinlichkeiten:*

Konditionierung auf ein neues Ereignis kann zur Umkehrung der gesamten Wahrscheinlichkeitsverhältnisse führen.

Es ist möglich, dass die folgenden Ungleichungen beide (zum gleichen Zeitpunkt) gültig sind:

$$P(A|B) > P(\bar{A}|B) \text{ und } P(A|B,C) < P(\bar{A}|B,C) \quad (4-2)$$

Hierbei bezeichnen die Buchstaben die An- oder Abwesenheit von verschiedenen Ereignissen.



Bsp.4-15: Non-monotone Wahrscheinlichkeiten

Gegeben:

A = Es hat letzte Nacht geregnet.

B = Der Rasen ist nass.

C = Die Besprengungsanlage war an.

Offensichtlich gilt die Beziehung:

$$P(A|B) > P(\bar{A}|B) \text{ und } P(A|B,C) < P(\bar{A}|B,C) \quad (4-2)$$

Eine zentrale Folgerung aus der fehlenden Monotonie besteht darin, dass bei Schlussfolgerungen mit Wahrscheinlichkeiten immer auf die gesamte relevante Information in der Wissensbasis zurückgegriffen werden muss. Die Gültigkeit eines Schlusses lässt sich somit nicht lokal auf der Basis von lokal beschränktem Wissen beurteilen.

Dieses Merkmal unterscheidet Wahrscheinlichkeitsschlüsse von deduktiven Schlussfolgerungen. Deren Gültigkeit kann rein lokal – ohne Betrachtung anderer Informationen als die unmittelbar in der Schlussfolgerung enthaltenen – beurteilt werden.

Die fehlende Monotonie hat wichtige praktische Auswirkungen im Zusammenhang mit induktiven Schlussfolgerungen, wie z.B. der Beurteilung von Evidenz für die Gültigkeit einer Theorie oder für induktive probabilistische Erklärungen.

4.2.4 Unterbewertung und Ignorierung von Basisraten

Es wurde bereits im Zusammenhang mit der Besprechung der Unterbewertung von Konsensusinformation in Abschnitt 2.5.2.3 darauf hingewiesen, dass es sich hier um ein generelleres Phänomen handelt, nämlich der Unterbewertung von Basisraten.



Konzept 4-1: Basisraten:

Basisraten repräsentieren die Auftretenswahrscheinlichkeit eines Ereignisses (oder einer Einheit) innerhalb der zugehörigen Referenzklasse (Population).

Basisraten repräsentieren somit die Wahrscheinlichkeit, dass ein zufällig aus der Population gezogenes Ereignis eines bestimmten Typs auftritt.

Beispiele:

- Anteil der Frauen und Männer in der Schweizer Bevölkerung [Referenzklasse = Schweizer Bevölkerung];
- Anteil der Personen mit endogener Depression innerhalb der Schweizer Studentinnen [Referenzklasse = Schweizer Studentinnen].

Basisraten enthalten klarerweise relevante Information über das mögliche Auftreten eines Ereignisses. Diese Information wird jedoch ignoriert bzw. zu wenig beachtet, falls zusätzliche Information vorhanden ist, welche als diagnostisch für die jeweilige Referenzklasse betrachtet wird, z.B. dass »die Studentin X oft traurig ist«.



Bemerkung

Man erinnere sich an den T. Barnum-Effekt (Konzept 2-1) und die Pseudodiagnostizität bestimmter Beschreibungen: Die meisten Menschen sind von Zeit zu Zeit traurig.

In einer Serie von Experimenten demonstrierten Kahneman und Tversky (1973), dass Personen Basisraten eine zu geringe Bedeutung beimessen.



Bsp.4-16: Basisraten-Neglect (Kahneman & Tversky, 1973):

Gegeben:

Die folgende Personenbeschreibung:

Jack ist 45 Jahre alt, verheiratet und hat 4 Kinder. Er ist im Allgemeinen konservativ, sorgfältig, und ehrgeizig. Er hat keinerlei Interesse für politische und soziale Fragen und verbringt den grössten Teil seiner freien Zeit mit seinen zahlreichen Hobbys. Zu diesen gehören Heimwerken, Segeln und mathematische Rätsel.

Aufgrund dieser Personenbeschreibung sollten die Vpn folgendes Wahrscheinlichkeitsurteil abgeben:

Die Wahrscheinlichkeit dass Jack einer von 30 (70) Ingenieuren aus der Gruppe von 100 ist, beträgt: _____

Für die Hälfte der Vpn ($N = 85$) waren 30 der 100 Personen Ingenieure und die anderen 70 Rechtsanwälte.

Für die andere Hälfte ($N = 86$) war das Verhältnis umgekehrt: 70 Ingenieure vs. 30 Rechtsanwälte.

Nach der Beschreibung erhielten beide Gruppen noch eine so genannte *Nullbeschreibung*:

Angenommen, Du hast keinerlei Information über ein zufällig aus der Stichprobe gezogenen Person.

Die Wahrscheinlichkeit dass dieser Mann einer von 30 (70) Ingenieuren aus der Gruppe von 100 ist, beträgt: _____

Die Manipulation der Basisraten (30/70 vs. 70/30) hatte bei den vollen Personenbeschreibungen einen sehr geringen (aber statistisch signifikanten) Effekt: Die Gruppe (30/70) gab als mittlere Schätzung den Wert 50% und die Gruppe (70/30) den Wert 55% (dafür, dass es sich um einen Ingenieur handelt).

Nur im Falle der Nullbeschreibung wurden die Basisraten adäquat berücksichtigt.

Wird anstelle der Nullbeschreibung folgende völlig uninformative Beschreibung präsentiert:

Dick ist 30 Jahre alt, verheiratet und ohne Kinder. Er ist ein Mann von grossen Fähigkeiten, der verspricht auf seinem Gebiet sehr erfolgreich zu werden. Er wird von seinen Kollegen gemocht.

In diesem Falle werden die Basisraten – anders als im Falle der Nullbeschreibung – vollständig ignoriert. Der mediane Schätzwert betrug in beiden Basisraten-Gruppen (30/70 vs. 70/30) 50%.

Das Beispiel demonstriert sehr schön die Vernachlässigung von Basisraten bei Vorliegen diagnostischer Information auch dann, wenn letztere von zweifelhaftem Wert ist.

Es gibt einen wichtigen Umstand, unter dem Basisraten beachtet werden. Dies ist dann der Fall, wenn die Basisraten kausal interpretiert werden können.



Bsp.4-17: Kausale Basisraten (Tversky & Kahneman 1982b):

Gegeben:

Zwei Beschreibungen einer Situation:

- (a) *Vor zwei Jahren gab es eine Abschlussprüfung für einen Kurs an der Yale Universität. Hierbei fielen (kamen) 75% der Studentinnen durch.*
- (b) *Vor zwei Jahren gab es eine Abschlussprüfung für einen Kurs an der Yale Universität. Ein Entwicklungspsychologe, der an schulischen Leistungen interessiert ist, interviewte eine grosse Anzahl von Studenten, die den Kurs besucht hatten. Da er vor allem an Reaktionen auf das Scheitern (den Erfolg) interessiert war, wählte er 75% der Studenten, die beim Examen durchgefallen (durchgekommen) waren.*

Die in den beiden Beschreibungen präsentierten Basisraten unterscheiden sich in einem zentralen Punkt: Die Basisraten in (a) enthalten kausale Information über den Schwierigkeitsgrad der Prüfung.

Dementsprechend hatten die Basisraten unterschiedliche Wirkung:

- Für (a) ergab sich eine Differenz von 34% zwischen der (75/25) und der (25/75) Bedingung.
- Für (b) betrug die Differenz 12%.

Zusammenfassend lässt sich daher sagen, dass Basisraten bei Vorliegen von (pseudo-) diagnostischer Information unterbewertet oder sogar völlig ignoriert werden, es sei denn die Basisraten besitzen kausale Implikationen.

4.3 Probabilistische Schlussfolgerungen

Probabilistisches Schlussfolgern wurde vor allem im Kontext des so genannten Bayes'schen Schliessens untersucht. Hierbei wurde das Urteil von Personen mit dem Ergebnis des *Bayes-Theorems* verglichen, wobei letzteres als normativer Standard betrachtet wurde.

**Bemerkung**

Einen zum Bayes'schen Schlussfolgern komplementären Forschungsansatz welcher – im weitesten Sinne – der Kategorie des probabilistischen Schlussfolgerns zugerechnet werden kann, betrifft Studien im Zusammenhang mit dem linearen Regressionsmodell. Einen frühen Vergleich beider Arten von Forschungsrichtungen präsentieren Slovic und Lichtenstein (1971).

Bevor wir die einschlägigen Experimente und Ergebnisse näher ins Auge fassen, sei hier das Bayes-Theorem in seiner einfachsten Form wiedergegeben.

**Konzept 4-2: Bayes-Theorem:****Gegeben:**

- Eine Menge von erschöpfenden und einander ausschließenden Hypothesen H_1, H_2, \dots, H_n . Dies bedeutet, dass genau eine der Hypothesen wahr ist. Die so genannte *a priori* Wahrscheinlichkeiten der einzelnen Hypothesen sind ebenfalls gegeben: $P(H_1), P(H_2), \dots, P(H_n)$ [Diese Wahrscheinlichkeiten müssen sich – da die Hypothesen exhaustiv und exklusiv sind auf 1.0 summieren].
- Ein Ereignis (bzw. Evidenz) E mit den bedingten Wahrscheinlichkeiten $P(E|H_1), P(E|H_2), \dots, P(E|H_n)$, welche auch *Likelihoods* genannt werden.

Bemerkung: Bei den Likelihoods handelt es sich nicht notwendigerweise um Wahrscheinlichkeiten. In den hier zu besprechenden Beispielen handelt es sich jedoch auch bei den Likelihoods immer um Wahrscheinlichkeiten.

Gesucht: Die so genannten *Posterior*-Wahrscheinlichkeiten:

$$P(H_1|E), P(H_2|E), \dots, P(H_n|E).$$

Das *Bayes-Theorem* besagt, dass sich die Posterior-Wahrscheinlichkeit $P(H_i|E)$ von Hypothese H_i wie folgt ergibt:

$$\begin{aligned} P(H_i|E) &= \frac{P(H_i, E)}{P(E)} \\ &= \frac{P(E|H_i) \cdot P(H_i)}{\sum_{j=1}^n P(E|H_j) \cdot P(H_j)} \end{aligned} \quad (4-3)$$

Bemerkungen:

1. Das Bayes-Theorem wird häufig als Regel zur Aktualisierung der a priori Wahrscheinlichkeiten betrachtet, indem diese durch Einbeziehung von neuer Evidenz E in Posterior-Wahrscheinlichkeiten überführt werden.
2. Das Bayes-Theorem ist eine normative Regel, welche direkt aus den Wahrscheinlichkeitsaxiomen und der Definition der bedingten Wahrscheinlichkeit folgt (siehe hierzu die Ausführungen im Anhang).
3. Das Bayes-Theorem vereinigt die drei elementaren Operationen des probabilistischen Schlussfolgerns: Kombination, Marginalisierung und Konditionierung:

- (i) Mittels Kombination von a priori Wahrscheinlichkeit $P(H_i)$ und Likelihood $P(E|H_i)$ wird die Verbundverteilung gebildet:

$$P(H_i, E) = P(E|H_i) \cdot P(H_i), \text{ für alle } H_i.$$

- (ii) Aufgrund der Verbundverteilung ergibt sich durch Marginalisieren die die Marginalverteilung $P(E)$:

$$\begin{aligned} P(E) &= P(H_1, E) + P(H_2, E) + \dots + P(H_n, E) \\ &= P(E|H_1) \cdot P(H_1) + P(E|H_2) \cdot P(H_2) + \dots + P(E|H_n) \cdot P(H_n) \\ &= \sum_{j=1}^n P(E|H_j) \cdot P(H_j) \end{aligned}$$

Bemerkung:

Die Verbundwahrscheinlichkeiten $P(H_i, E)$ sind disjunkt (Abb. 4-4). Daher folgt das Ergebnis der Marginalisierung (= Summierung der einzelnen Verbundwahrscheinlichkeiten) direkt aus dem dritten Wahrscheinlichkeitsaxiom.

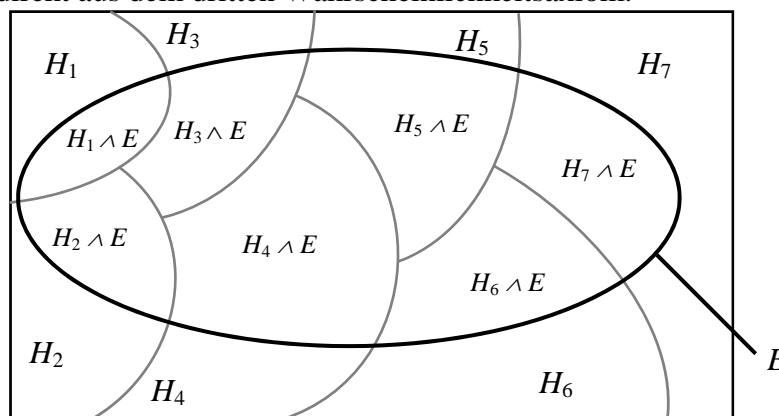


Abb. 4-4: Verhältnis von Evidenz E und den exklusiven und exhaustiven Hypothesen: Die einzelnen Verbundwahrscheinlichkeiten $P(H_i \wedge E)$ sind disjunkt.

- (iii) Die gesuchte bedingte Wahrscheinlichkeit ergibt sich nun durch Konditionierung auf die neue Evidenz E :

$$P(H_i|E) = \frac{P(H_i, E)}{P(E)}$$

Das klassische Problem zur Prüfung der Frage, ob Menschen gemäss Bayes-Theorem urteilen, bildet das so genannte Taxiproblem (Cab-Problem) von Tversky und Kahneman (1982b).



Bsp.4-18: Das Taxiproblem [Cab-Problem] (Tversky & Kahneman 1982b):

Ein Taxi wurde in der Nacht in einen Verkehrsunfall mit Fahrerflucht verwickelt. Es gibt zwei Taxi-Gesellschaften in der Stadt: Die Grünen und die Blauen. Du hast die folgenden Daten:

- (a) 85% der Taxis in der Stadt sind Grün (G) und 15% sind Blau (B).
- (b) Ein Zeuge identifizierte das Taxi als Blau ($\gg B \ll$). Das Gericht testet die Reliabilität des Zeugen unter den gleichen Bedingungen, die in der Nacht des Unfalls existierten und folgerte, dass der Zeuge in 80% der Fälle die Farbe korrekt identifiziert und in den anderen 20% falsch (unabhängig von der Farbe des Taxis).

Wie gross ist die Wahrscheinlichkeit, dass das Taxi tatsächlich Blau war?

Ergebnis:

Die modale (häufigste) Antwort der Vpn war 0.80.

Interpretation:

Die Antworten der Vpn zeigen klar eine Verwechslung der bedingten Wahrscheinlichkeiten:

$$P(\text{Taxi ist blau} | \text{Zeuge sagt blau}) = 0.41 \text{ und}$$

$$P(\text{Zeuge sagt blau} | \text{Taxi ist blau}) = 0.80.$$

Die Frage betrifft eindeutig die erste der beiden bedingten Wahrscheinlichkeiten.

Eine bequeme Art zur Analyse des Problems besteht in dessen Repräsentation als Ereignisbaum:



Bsp.4-19: Das Taxiproblem als Ereignisbaum repräsentiert:

Gegeben:

B = Das Taxi ist Blau.

G = Das Taxi ist Grün.

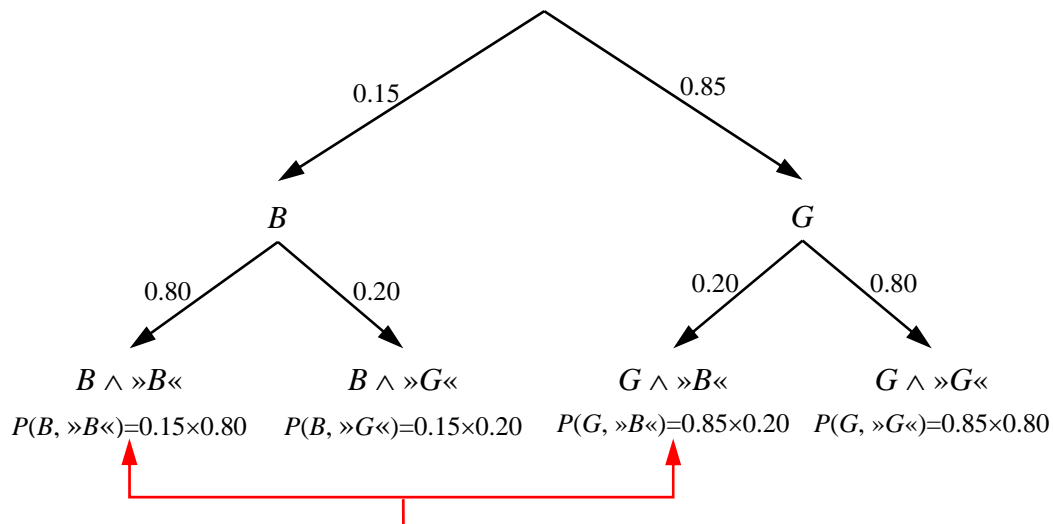
$\gg B \ll$ = Der Zeuge identifiziert das Taxi als Blau.

$\gg G \ll$ = Der Zeuge identifiziert das Taxi als Grün.

Abb. 4-5 zeigt die Repräsentation des Taxiproblems als Ereignisbaum.

Die Pfeile sind mit den relevanten Wahrscheinlichkeiten bezeichnet. Im oberen Bereich handelt es sich hierbei um die a priori Wahrscheinlichkeiten (Basisraten), $P(B)$ und $P(G)$, und im unteren Teil um die Likelihoods, d.h. um die bedingten Wahrscheinlichkeiten der Antwort des Zeugen, gegeben die Farbe): $P(\gg B \ll | B)$, $P(\gg G \ll | B)$, $P(\gg B \ll | G)$ und $P(\gg G \ll | G)$.

Im unteren Bereich befinden sich die Verbundwahrscheinlichkeiten, die durch Multiplikation der Wahrscheinlichkeiten entlang des Pfades ermittelt werden. Dies repräsentiert die *Kombination* der Information.



Hier handelt es sich um die beiden relevanten Verbundwahrscheinlichkeiten:

$$\begin{aligned}
 P(B | \gg B \ll) &= \frac{P(B, \gg B \ll)}{P(B, \gg B \ll) + P(G, \gg B \ll)} \\
 &= \frac{0.12}{0.12 + 0.17}
 \end{aligned}$$

Abb. 4-5: Repräsentation der Taxi-Problems mit Hilfe eines Ereignisbaumes.

Sobald die Verbundwahrscheinlichkeiten konstruiert sind, werden jene Verbundereignisse herausgegriffen, die mit der vorliegenden Evidenz verträglich sind. Im aktuellen Fall hat der Zeuge »B« gesagt, damit sind alle Verbundereignisse mit »G« irrelevant, da unverträglich mit der vorliegenden Evidenz.

Die relevanten Verbundwahrscheinlichkeiten werden sodann zur Durchführung der Konditionierung auf die aktuell vorliegende Evidenz herangezogen.

Die Wahrscheinlichkeit von $P(\gg B \ll)$ im Nenner ergibt sich als Summe: $P(\gg B \ll) = P(B, \gg B \ll) + P(G, \gg B \ll)$ [Marginalisierung der Verbundverteilung] und die relevante Verbundverteilung $P(B, \gg B \ll)$ bildet den Zähler.

Die Ereignisbaumdarstellung macht deutlich, warum trotz der hohen Reliabilität des Zeugen (80%) die Wahrscheinlichkeit, dass es sich aktuell um ein grünes Taxi handelt, nur ungefähr 41% beträgt: Aufgrund der Tatsache, dass es viel mehr grüne Taxis gibt, kommt es zu vielen *falschen Alarmen*, d.h. grüne Taxis werden irrtümlich als blau identifiziert. Dies ist im Baum durch den Pfad, der auf das Ereignis $G \wedge \gg B \ll$ hinführt, repräsentiert.

Die falschen Alarme machen mehr als die Hälfte der »Blau« - Urteile des Zeugen aus. Greift man also zufällig ein »Blau« - Urteil aus dieser Menge heraus, so ist es wahrscheinlicher, dass es sich um einen falschen Alarm und damit um ein grünes Taxi handelt als um einen Treffer (und damit um ein blaues Taxi).

4.3.1 Das Bayes-Theorem in Wettquotienten-Format

Die Darstellung von Bayes'schen Problem mit Hilfe von Ereignisbäumen führt zu einem guten Verständnis des Problems. Für die praktische Berechnung ist es jedoch einfacher, das Theorem im Format von *Wettquotienten* (=Odds-Format) zu verwenden.



Konzept 4-3: Das Bayes-Theorem im Wettquotienten-Format

Gegeben:

- Hypothesenmenge H_1, H_2, \dots, H_n mit a priori Wahrscheinlichkeiten: $P(H_1), P(H_2), \dots, P(H_n)$
- Likelihoods $P(E|H_1), P(E|H_2), \dots, P(E|H_n)$ der Evidenz E , gegeben die Hypothesen.

Bei Verwendung des Theorems in Form von Wettquotienten werden anstelle der Posterior-Wahrscheinlichkeit $P(H_i|E)$ die Wettquotienten der Posterior-Wahrscheinlichkeiten:

$$\omega_{i,n} = \frac{P(H_i|E)}{P(H_n|E)}, \quad i = (1, 2, \dots, n-1)$$

berechnet.

Hierbei wird eine der Posterior-Wahrscheinlichkeiten als Referenzwahrscheinlichkeiten verwendet. Im aktuellen Fall wurde die Wahrscheinlichkeit $P(H_n|E)$ verwendet. Es ist jedoch gleichgültig, welche der n Posterior-Wahrscheinlichkeiten verwendet wird.

Der Vorteil der Verwendung von Quotienten liegt darin, dass die aufwendige Berechnung des Nenners in Gleichung (4-3) entfällt, denn es gilt:

$$\frac{P(H_i|E)}{P(H_n|E)} = \frac{P(E|H_i)}{P(E|H_n)} \cdot \frac{P(H_i)}{P(H_n)} \quad (4-4)$$

Oder einfacher:

$$\omega_{i,n} = \lambda_{i,n} \cdot \pi_{i,n}, \quad (4-5)$$

wobei gilt:

$$\omega_{i,n} = \frac{P(H_i|E)}{P(H_n|E)} \quad \text{ist der Posterior Wettquotient (Posterior-Odds)}$$

$$\lambda_{i,n} = \frac{P(E|H_i)}{P(E|H_n)} \quad \text{ist die Likelihood-Ratio}$$

$$\pi_{i,n} = \frac{P(H_i)}{P(H_n)} \quad \text{ist der a priori Wettquotient (Prior-Odds)}$$

Sobald die Posterior-Wettquotienten berechnet sind, können diese zu Wahrscheinlichkeiten normiert werden, indem man jeden Quotienten durch die Summe aller Quotienten dividiert.

Man beachte, dass immer einer der n Quotienten den Wert 1 hat, im aktuellen Falle:

$$\omega_{i,n} = \frac{P(H_n|E)}{P(H_n|E)} = 1$$

Somit ergibt sich die Posterior-Wahrscheinlichkeit durch:

$$P(H_i|E) = \frac{\omega_{i,n}}{\sum_{j=1}^n \omega_{j,n}} \quad (4-6)$$

Für den Laien erscheint das Bayes-Theorem in Form von Wettquotienten meist als (unnötig) kompliziert. Dem ist aber nicht so, wie die folgenden Beispiele zeigen.



Bsp.4-20: Das Taxiproblem im Wettquotienten-Format

Gegeben:

B = Das Taxi ist Blau.

G = Das Taxi ist Grün.

»B« = Der Zeuge identifiziert das Taxi als Blau.

»G« = Der Zeuge identifiziert das Taxi als Grün.

Aufgrund der Wahrscheinlichkeiten (siehe Bsp.4-18) ergibt sich der folgende Wettquotient:

$$\begin{aligned}\frac{P(B|B\ll)}{P(G|B\ll)} &= \frac{P(B\ll|B) \cdot P(B)}{P(B\ll|G) \cdot P(G)} \\ &= \frac{0.8}{0.2} \cdot \frac{0.15}{0.85} \\ &= \frac{0.12}{0.17}\end{aligned}$$

Der zweite Quotient muss nicht berechnet werden, da gilt:

$$\frac{P(G|B\ll)}{P(G|B\ll)} = 1$$

Wir erhalten somit nach Normierung:

$$\begin{aligned}P(B|B\ll) &= \frac{P(B|B\ll)/P(G|B\ll)}{1 + P(B|B\ll)/P(G|B\ll)} \\ &= \frac{0.12/0.17}{1 + 0.12/0.17} = \frac{0.12}{0.17 + 0.12} = \frac{12}{29} = 0.41\end{aligned}$$

Das nächste Beispiel demonstriert noch eindrücklicher die Vorteile des Wettquotientenformats des Bayes-Theorems.



Bsp.4-21: Das Bayes-Theorem in Wettquotienten-Format

Gegeben:

Ein Dozent hat zwei Studentinnen S_1 und S_2 , die seine Übungen regelmässig machen.

Studentin S_1 ist in der Lage durchschnittlich $2/3$ der Probleme zu lösen, Studentin S_2 löst im Durchschnitt $1/3$ der Probleme.

Der Dozent wählt nun eine der beiden Studentinnen aus, wobei er folgendes Zufallsverfahren wählt: Er verwendet einen fairen Würfel mit folgender Auswahlregel: Falls der Würfel eine Augenzahl kleiner 3 zeigt, so wird S_1 gewählt, andernfalls S_2 .

Die gewählte Studentin erhält 5 Aufgaben, die zufällig aus dem gesamten Pool gleich schwieriger Aufgaben ausgewählt wurden. Sie löst 3 der 5 Aufgaben.

Wie wahrscheinlich ist es, dass es sich um S_1 handelt?

Bemerkung:

Die gegebene Formulierung legt das Problem nicht eindeutig fest. Wir müssen noch annehmen, dass die Wahrscheinlichkeit der Lösung einzig von der Fähigkeit der Studentin, wie sie in den oben präsentierten Wahrscheinlichkeiten repräsentiert ist, abhängt.

Dies schliesst Transfer- und Positionseffekte aus. Technisch ausgedrückt: Die Wahrscheinlichkeit einer Lösung (L) ist stochastisch unabhängig von Situations- (S) und Persönlichkeitseinflüssen (R), wie z.B. Ermüdung oder Training, gegeben die Fähigkeit (F) der Studentin: $P(L|S, R, F) = P(L|F)$.

Aufgrund dieser Annahme ergibt sich die Wahrscheinlichkeit, dass 3 von 5 Aufgaben von Studentin S_i gelöst werden mit Hilfe der Binomialverteilung:

$$P(N = 3|S_i) = \binom{5}{3} \cdot \pi_i^3 \cdot (1 - \pi_i)^2 \quad i = (1, 2).$$

Hierbei gilt:

$\pi_i = P(L|F_i)$ repräsentiert die Fähigkeit von S_i , eine Aufgabe zu lösen. Für das aktuellen Beispiel gilt:

$$\pi_1 = 2/3 \text{ und } \pi_2 = 1/3.$$

Zur Lösung des Problems benötigen wir noch die a priori Wahrscheinlichkeiten der beiden Studentinnen. Diese ergeben sich aufgrund des Auswahlprozesses wie folgt:

$$P(S_1) = 1/3 \text{ und } P(S_2) = 2/3.$$

Das Bayes-Theorem in Quotientenformat ergibt nun:

$$\frac{P(S_1|N = 3)}{P(S_2|N = 3)} = \frac{(2/3)^3 \cdot (1/3)^2 \cdot (1/3)}{(1/3)^3 \cdot (2/3)^2 \cdot (2/3)} = 1$$

$$\text{Somit: } P(S_1|N = 3) = P(S_2|N = 3) = 1/2.$$

Man beachte, dass der Binomial-Quotient:

$$\binom{5}{3} = \frac{5!}{3!2!}$$

im Nenner und Zähler identisch ist und daher weggelassen werden kann, was die Berechnung bedeutend vereinfacht.

Nehmen wir nun an, der Dozent hätte anstelle des Würfels eine faire Münze verwendet und je nach Kopf oder Zahl S_1 bzw. S_2 gewählt. In diesem Fall ergäbe sich:

$$\frac{P(S_1|N = 3)}{P(S_2|N = 3)} = \frac{(2/3)^3 \cdot (1/3)^2 \cdot (1/2)}{(1/3)^3 \cdot (2/3)^2 \cdot (1/2)} = \frac{2}{1}.$$

Die Wahrscheinlichkeit, dass S_1 ausgewählt wurde, beträgt in diesem Falle: $P(S_1|N=3) = 2/3$.

Bsp.4-21 demonstriert nun sehr deutlich die Vorteile des Wettquotientenformats.

Wettquotienten sind auch noch in anderer Hinsicht von Interesse. Sie ermöglichen es, die Unsicherheit von Ereignissen zu quantifizierten ohne die Normierung, welche Wahrscheinlichkeiten unterworfen sind (da sich die Wahrscheinlichkeiten von disjunktiven und exhaustiven Ereignissen auf 1 summieren müssen), zu berücksichtigen. Es hat sich gezeigt, dass die Verwendung von Wettquotienten zu konsistenteren Schlüssen führt (Slovic & Lichtenstein, 1971).



Bemerkung

Bezüglich des Umgangs mit Wettquotienten scheint es deutliche Unterschiede zwischen dem deutschen und dem angelsächsischen Raum zu geben: Wettquotienten sind im englischen Sprachraum weit verbreitet (speziell in England, das eine sehr spielfreudige Nation zu sein scheint). Im deutschen Raum können viele Leute mit Wettquotienten wenig anfangen.

Bei der Behandlung von Wahrscheinlichkeitsschlüssen darf – schon des Vergnügens wegen – ein Problem nicht fehlen, nämlich das Ziegenproblem (Monty Hall Dilemma).

4.3.2 Exkurs: Das Ziegenproblem



Bsp.4-22: Das Ziegenproblem (Monty Hall Dilemma):

Sie nehmen an einer Spielshow im Fernsehen teil, bei der Sie eine von drei Türen (A, B, oder C) wählen sollen. Hinter einer der drei Türen befindet sich der Preis, ein Auto, hinter den beiden anderen stehen Ziegen.

Sie entscheiden sich für eine Tür, sagen wir Tür A. Sie bleibt vorerst geschlossen. Der Moderator, der weiß hinter welcher Tür sich das Auto befindet, öffnet eine andere Tür, sagen wir Tür B und eine meckernde Ziege schaut ins Publikum.

Der Moderator fragt: Bleiben Sie bei Tür A oder wollen Sie zu Tür C wechseln.

Bemerkungen:

1. Dieses Problem hat Anfang der 90er-Jahre ein ungeheures Aufsehen erregt. Der Grund hierfür war, dass viele Leute – darunter der berühmte Mathematiker Paul Erdős, der die meisten mathematischen Publikationen des 20. Jahrhunderts aufweist – intuitiv glaubten, dass es gleichgültig sei, ob man die Tür wechselt oder nicht.

2. Es handelt sich hierbei um ein Problem das zu einem viel älteren Problem strukturell äquivalent ist: dem *Gefangenen-dilemma* (Mosteller, 1965).
3. Das Problem ist nicht vollständig formuliert. Es fehlt die Spezifikation der Präferenz des Moderators falls sich der Preis hinter der vom Kandidaten gewählten Tür A befindet. Dies macht einen Unterschied.
Nehmen wir an, der Moderator wählt in diesem Falle immer Tür B. Hat er nun tatsächlich B gewählt, so ist es gleichgültig, ob der Kandidat wechselt oder nicht, d.h. es besteht die gleiche Chance, dass sich der Preis hinter A oder C befindet (siehe Übung 4-6).

Intuitive Lösung des Problem:

Nehmen wir an, die Kandidatin wählt Tür A (Es ist gleichgültig, welche Tür sie wählt), dann beträgt die Wahrscheinlichkeit, dass sie gewinnt $1/3$. Die Wahrscheinlichkeit, dass sich der Preis hinter einer der beiden anderen Türen befindet, beträgt $2/3$.

Der Moderator beseitigt nun die Unsicherheit, hinter welcher der beiden Türen sich der Preis mit der Wahrscheinlichkeit $2/3$ befindet. Für die nicht geöffnete Tür besteht daher eine Wahrscheinlichkeit von $2/3$, dass sich der Preis dahinter befindet.

Das Problem lässt sich formal sehr einfach lösen. Wir verwenden nun – zur Abwechslung – eine Kontingenztabelle zur Repräsentation der Verbundverteilung über die beiden Ereignisse: *Tür, hinter der sich der Preis befindet* und *Die vom Moderator geöffnete Tür*. Hierbei nehmen wir an, dass die Kandidatin Tür A gewählt hat und der Moderator keinerlei Bevorzugung für Tür B oder C aufweist (d.h. falls sich das Auto hinter Tür A befindet, so wählt er mit gleicher Wahrscheinlichkeit entweder Tür B oder Tür C). Tab. 4-3 zeigt die Kontingenztabelle mit den Verbundwahrscheinlichkeiten.

Preis	Geöffnete Tür			Σ
	»A»	»B»	»C»	
A	0	1/6	1/6	1/3
B	0	0	1/3	1/3
C	0	1/3	0	1/3
Σ	0	1/2	1/2	

Tab. 4-3: *Tabelle der Verbundwahrscheinlichkeiten für das Ziegenproblem, unter der Annahme, dass der Kandidat Tür A gewählt hat und der Moderator keine der Türen (B oder C) bevorzugt.*

Hieraus ergeben sich durch Konditionierung auf die durch den Moderator geöffnete Tür sofort die gesuchten Wahrscheinlichkeiten:

$$P(A|»B») = P(A|»C») = \frac{1/6}{1/2} = \frac{1}{3} \text{ und}$$

$$P(B|»C») = P(C|»B») = \frac{1/3}{1/2} = \frac{2}{3}.$$

Man sieht also, dass die Kandidatin immer wechseln sollte, gleichgültig, ob der Moderator »B« oder »C« sagt.

4.4 Die Kritik am »Heuristics and Biases« Programm

In den 1980er und 1990er Jahren gab es vermehrt Kritik am »Heuristics and Biases« Ansatz. Die Kritikpunkte lassen sich in drei grobe Kategorien unterteilen:

- (1) Leugnung der Normativität von Axiomen und Regeln der Wahrscheinlichkeitsrechnung.
- (2) Kritik der Anwendbarkeit der Axiome der Wahrscheinlichkeitsrechnung.
- (3) Kritik der verwendeten Methode und Modelle.

Der gemeinsame Nenner der Kritik besteht darin, dass die gefundenen Fehler im Wahrscheinlichkeitsurteil durch Anwendung ungerechtfertigter Normen oder Methoden zustande kamen. Folglich sind auch die Implikationen für die Beurteilung der menschlichen Rationalität nicht gerechtfertigt.

Im Folgenden sollen nun diese Argumente näher untersucht werden.

4.4.1 Normativität von Axiomen und Regeln der Wahrscheinlichkeitsrechnung

Es gibt nur wenige Personen, welche die Normativität der Axiome der Wahrscheinlichkeitsrechnung in Zweifel ziehen. Einer davon war der britische Philosoph Jonathan Cohen (1923 - 2006). Dieser behauptete allen Ernstes, dass nichts als Fehler zählt, solange die Person, welche den Fehler macht, unter idealisierten Bedingungen nicht zugibt, dass sie einen Fehler gemacht hat. (Cohen, 1981, Seite 322, rechte Spalte, zweiter Absatz).

Ein zweites Argument von Cohen besagt, dass die Konstruktion von Axiomen und Regeln letztlich auf der Intuition gewöhnlicher Menschen beruhen, da es sich um nichts Anderes handelt als um Systematisierungen dieser Intuitionen. Daher können die gewöhnlichen Leute nicht als irrational betrachtet werden (Cohen, 1981, Seite 322, rechte Spalte, letzter Absatz).

Cohen schliesst daraus, dass der »Heuristics and Biases« Ansatz keine relevanten Implikationen bezüglich der Beurteilung der menschlichen Rationalität hätten.

Man kann natürlich der Auffassung sein, dass jeder rational ist, solange er nicht zugibt, dass er irrational handelt. Es fragt sich aber, ob diese Konzeption von Rationalität in irgendeiner Weise nützlich ist. Wenn nämlich jemand sich ständig aufgrund seines inkonsistenten Glaubenssystems selbst schadet, so scheint es sinnvoll diese Person als irrational zu bezeichnen (natürlich kann man auch von *sub-optimalem Verhalten* sprechen, wenn man das Wort irrational vermeiden will).

Eindeutig falsch ist jedoch die Folgerung von Cohen (1981), seine Ausführungen würden zeigen, dass der »Heuristics and Bases« Ansatz keine relevanten Implikationen bezüglich der Beurteilung der menschlichen Rationalität hat. Der Grund hierfür ist ganz einfach, dass die Leute – sobald sie über den wahren Sachverhalt aufgeklärt werden – ihre Fehler meist sofort einsehen und zugeben, dass sie einen Fehler gemacht haben, d.h. sie bemerken, dass sie einer fehlerhaften Intuition gefolgt sind.

Ein weiteres Problem für Cohens Argumentation bilden die Ergebnisse Stanovich und West (2000). Diese zeigen nämlich, dass Personen mit höheren kognitiven Fähigkeiten weniger Urteilsfehler produzieren.

Im Allgemeinen scheint es der Fall zu sein, dass die meisten Personen die Axiome und Regeln des Wahrscheinlichkeitskalküls als sinnvoll betrachten (So berichte Pinker (1997), dass ein Student, nachdem er auf vom Konjunktionsfehler im Zusammenhang mit dem Linda-Problem (siehe Bsp.4-8) gehört hatte, spontan ausrief: »Ich schäme mich für meine Spezies«). Das bedeutet aber noch lange nicht, dass sie diese Prinzipien auch korrekt für die Lösung komplexer Probleme anwenden können. Dies bringt uns zur zweiten Gruppe von Kritikpunkten, nämlich die Frage der Anwendbarkeit der normativen Kriterien.

4.4.2 Anwendbarkeit der Axiome und Regeln der Wahrscheinlichkeitsrechnung

Wir haben gesehen, dass der axiomatische Wahrscheinlichkeitsbegriff Wahrscheinlichkeiten als normierte Maße betrachtet, ähnlich wie Längen oder Flächen. Diese Konzeption stellt eine Idealisierung dar, die niemals exakt auf real existierende Ereignisse zutrifft.

Die beste Realisierung sind Casino-Spiele mit sehr langen Sequenzen, wobei zusätzlich darauf geachtet wird, dass die grundlegenden Annahmen – wie z.B. Unabhängigkeit der Ereignisse – weitgehend erfüllt sind. Daher bilden Wahrscheinlichkeitsmodelle, welche die Verteilung der möglichen Ergebnisse von Spielen beschreiben, die Wirklichkeit weitgehend korrekt ab. Auch in diesem Falle beschreibt das Modell die Realität niemals völlig exakt. Im Allgemeinen gilt das folgende Prinzip:



Prinzip 4-2: *Alle Modelle sind fehlerbehaftet*

Alle Modelle beschreiben den Anwendungsbereich nicht exakt, sondern nur approximativ. Daher sind alle Modelle – angewendet auf konkrete Bereiche – mit Fehlern behaftet.



Bemerkung

Die Maßaxiome treffen auch nicht exakt auf reale Längen von Dingen zu, denn es ist natürlich nicht der Fall dass die Länge zweier aneinander gelegter Stücke exakt der Summe der beiden Einzelstücke entspricht. Man kann aufgrund des Aufbaus der Stäbe diese vermutlich nicht nahtlos aneinander fügen.

Auch macht es einen Unterschied, wie die Stäbe aneinander gelegt werden, da die Enden der Stäbe nie vollständig eben sein können. Dies verletzt das Prinzip der Kommutativität der Addition, wonach gilt: $a + b = b + a$ (für beliebige Zahlen a und b).

Aufgrund der Tatsache, dass Modelle die Wirklichkeit immer nur approximativ beschreiben, läuft die Frage der Anwendbarkeit darauf hinaus, ob die durch ein Modell gegebene Annäherung hinreichend genau ist, sodass das Wahrscheinlichkeitsmodell als nützlich bezeichnet werden kann.



Bsp.4-23: Adäquatheit von Wahrscheinlichkeitsmodellen (Fortsetzung von Bsp.4-21)

In Bsp.4-21 wurde angenommen, dass die Wahrscheinlichkeit, mit der die Studentinnen die Aufgaben lösen, einzig von ihrer Problemlösefähigkeit abhängt (welche als konstant angenommen wurde).

Nur in diesem Falle ist das verwendete Binomialmodell zur Beschreibung der Wahrscheinlichkeit der Anzahl korrekter Antworten anwendbar. Natürlich sind die Bedingungen für die Anwendung des Modells niemals exakt erfüllt. Für die meisten praktischen Zwecke reicht es jedoch, wenn diese Bedingungen weitgehend erfüllt sind.

Die Nützlichkeit von Wahrscheinlichkeitsmodellen gilt in vielen Bereichen als erwiesen. Beispiele hierfür sind: Empirische Sozialforschung, medizinische Forschung, Versicherungsstatistik, Statistische Mechanik usw.

Im Zusammenhang mit der »Heuristics and Bases« Forschung stellen sich nun zwei Fragen:

1. Lassen sich Wahrscheinlichkeitsmodelle auch auf die Tversky und Kahnemann beschriebenen Probleme und Szenarien anwenden?
2. Sind die von Tversky und Kahnemann angewandten Modelle korrekt?

Wir werden in diesem Abschnitt nur die erste Frage behandeln und verschieben die Besprechung der zweiten Frage auf den nächsten Abschnitt.

Pinker (1997) behauptet, dass der im Linda-Problem (siehe Bsp.4-8) beobachtete Konjunktionsfehler nicht unbedingt einen Fehler darstellen muss, da es sich bei den involvierten Wahrscheinlichkeiten um subjektive Wahrscheinlichkeiten handle. Da die Träger subjektiver Wahrscheinlichkeiten geistige Inhalte sind und somit keinen extensionalen Charakter besitzen, sind Beziehungen zwischen Mengen (welche extensionale Entitäten darstellen) irrelevant.



Bemerkung:

Die Beziehung wonach $P(B) \geq P(A \cap B)$ für Mengen gilt, ergibt sich direkt aus dem dritten Axiom von Kolmogorov (Additivität von Wahrscheinlichkeiten disjunkter Ergebnisse), wie folgt:

$$B = (A \cap B) \cup (\bar{A} \cap B),$$

wobei die beiden Ereignisse $A \cap B$ und $\bar{A} \cap B$ disjunkt sind (da A und \bar{A} disjunkt sind). Somit:

$$\begin{aligned} P(B) &= P[(A \cap B) \cup (\bar{A} \cap B)] \\ &= P(A \cap B) + P(\bar{A} \cap B) \\ &\geq P(A \cap B) \end{aligned}$$

Gegen Pinkers Argument lassen sich zwei Einwände formulieren:

1. Eine Person, die einem Konzept, welches einen Spezialfall eines anderen darstellt, – wie im Falle des Linda-Problems die feministische Bankangestellte ein Spezialfall einer Bankangestellten ist – eine höhere Auftretenswahrscheinlichkeit zuordnet, ist nach allen bekannten Massstäben inkonsistent (zumindest, wenn der Begriff der Wahrscheinlichkeit eine Bedeutung hat, welche der im Alltag verwendeten annähernd entspricht). Diese Inkonsistenz kann – gemäss Dutchbook-Argument – dazu führen, dass die Person Verluste erleidet, d.h. ihre Ziele nicht optimal erreichen kann.
2. Für das Linda-Problem existieren Referenzmengen, nämlich die Menge der Bankangestellten bzw. die Menge der feministischen Bankangestellten. Daher lässt sich durchaus ein auf Menge beruhender Wahrscheinlichkeitsbegriff heranziehen, auch wenn es sich um einen Einzelfall handelt.

**Bemerkung:**

Auch wenn es sich um einen Einzelfall handelt, spielt die Wahrscheinlichkeit eine Rolle. Angenommen Herr Pinker wäre gezwungen, »Russisches Roulette« zu spielen, wobei er die Wahl zwischen einem Revolver mit einer Kugel oder mit 5 Kugeln (von 6 möglichen) hätte.

Ich glaube kaum, dass er in diesem Fall der Meinung wäre, dass es gleichgültig ist, welche Waffe er wählt.

Auch Gigerenzer sieht Probleme, was die Anwendbarkeit von Normen im Kontext der »Heuristics and Bases« Forschung betrifft. Er bleibt jedoch in seinen Aussagen vager und meint nur, dass inhaltsleere Normen nicht einfach auf konkrete Alltagsprobleme angewendet werden können, ohne genau zu prüfen, ob die Voraussetzungen erfüllt sind (siehe z.B. Gigerenzer, 1996, 2001; Vranas, 2000). Die Hauptkritik von Gigerenzer zielt im Spezifischen darauf, dass normative Modelle von Tversky und Kahneman falsch angewendet wurden. Dieser Kritik wenden wir uns im Folgenden zu.

4.4.3 Fehlerhaft angewendete normative Modelle

Der Kritikpunkt der fehlerhaften Anwendung des Bayes-Theorems wurde im Zusammenhang mit dem Taxi-Troblem (vgl. Bsp.4-18) von Birnbaum (1983) erhoben und von Gigerenzer und Murray (1987) wiederholt. Die Kritik von Birnbaum lautet wie folgt:

Die Bedingungen, unter denen der Zeuge die Farbe des Taxis ursprünglich identifizierte, sind verschieden von jenen, in denen die Reliabilität des Zeugen getestet wurde. Im Speziellen verändert der Zeuge bei der Prüfung der Reliabilität sein Entscheidungskriterium, um den unterschiedlichen Basisraten von blauen und grünen Taxis in der Natur (15/85) und während des Test (50/50) gerecht zu werden. Dadurch ändert sich die Trefferrate sowie die Rate des falschen Alarms und sie entsprechen nicht mehr den von Tversky und Kahneman angegebenen Werten.

Der zentrale Punkt dieser Kritik besteht in der Behauptung, dass *die Basisrate einen Einfluss auf die Likelihoods* (welche die Treffer und falschen Alarme repräsentieren) besitzt. Dies bedeutet, dass *die Basisrate einen Einfluss auf den Mechanismus ausübt, welcher der Kategorisierung der wahrgenommenen Taxis als »Grün« oder »Blau« zugrunde liegt.*

Aus der Wahrnehmungspsychologie ist bekannt, dass die Basisraten tatsächlich einen Einfluss auf die Kategorisierung von Stimuli haben. Das bekannteste Modell, welches dies erklärt und auf welches sich Birnbaum (1983) bezieht, ist das klassische Signalentdeckungsmodell (Abb. 4-6). Dieses Modell, das von bestechender Einfachheit ist, basiert auf die folgenden vier Annahmen:

1. Die Entscheidung des Zeugen basiert auf einer kontinuierlichen Grösse, welche als »Blaueit« bezeichnet werden kann. Diese Variable ist latent, d.h. von aussen nicht direkt beobachtbar. In Abb. 4-6 ist diese Grösse durch x -Achse repräsentiert
2. Die Präsentation ein und desselben Stimulus führt nicht immer exakt zum gleichen Wert auf der internen Skala »Blaueit«. Die Verteilung der »Blaueitswerte« für die beiden Stimuli ist in Abb. 4-6 durch die beiden Normalverteilungskurven repräsentiert: Die grüne Kurve repräsentiert die Verteilung der »Blaueitswerte«, wenn ein grünes Taxi präsentiert wird und die blaue Kurve die Verteilung bei Darbietung eines blauen Taxis. Man beachte, dass ein grünes Taxi manchmal einen höheren Wert auf der Variable »Blaueit« evoziert als ein grünes Taxi. Insgesamt aber führen blaue Taxis zu höheren »Blaueitswerten« als grüne Taxis.

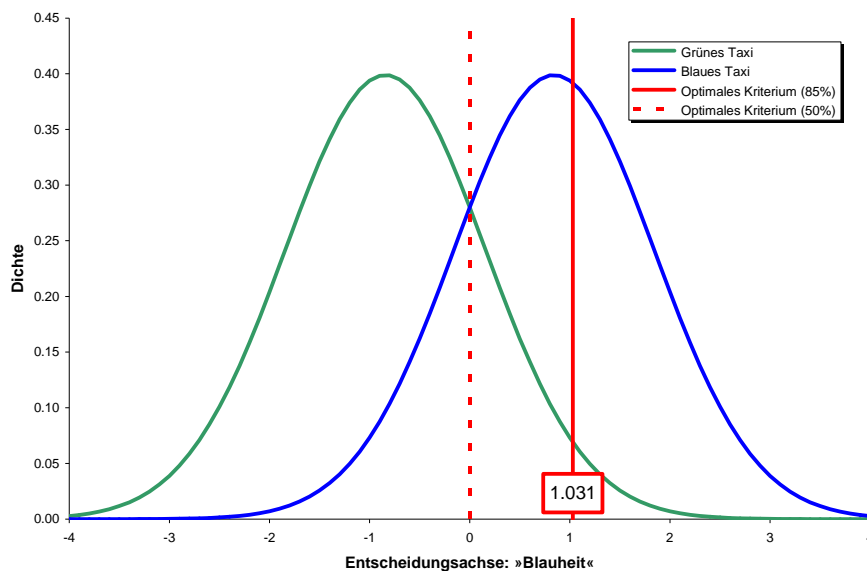


Abb. 4-6: Signalentdeckungsmodell als Modell des Zeugen beim Taxiproblem: Die grüne und blaue Kurve repräsentieren die Signalverteilungen von grünen und blauen Taxis. Die rote gestrichelte Linie zeigt die optimale Lage des Entscheidungskriteriums, falls 50% der Taxis grün sind, senkrechte durchgezogene Linie zeigt die optimale Lage des Kriteriums, wenn 85% der präsentierten Taxis grün sind.

3. Um zu einer Entscheidung zu gelangen, ob das präsentierte Taxi tatsächlich blau oder grün war, führt der Zeuge ein Entscheidungskriterium ein (die roten senkrechten Linien in Abb. 4-6), welche die »Blaueitsachse« in zwei Regionen teilt. Falls der aktuelle Stimulus zu einem »Blaueitsempfinden« führt, dass rechts vom Kriterium liegt, so antwortet der Zeuge mit »Blau«, andernfalls mit »Grün«.

4. Während der Zeuge keinerlei Einfluss auf sein »Blauheitsempfinden« hat, d.h. die Verteilung der Werte auf der »Blauheitsachse« unterliegen nicht seiner Kontrolle, kann er sein Entscheidungskriterium der aktuellen Situation anpassen.

Im Speziellen kann der Zeuge durch Anpassung des Entscheidungskriteriums seine Fehler minimieren. Falls 50% der präsentierten Taxis grün sind, so liegt die optimale Achse genau am Schnittpunkt der beiden Kurven. In Abb. 4-6 wird dieses Kriterium durch die gestrichelte Linie repräsentiert. Seine falsche Alarmrate beträgt in diesem Falle 20% (= Fläche unterhalb der grünen Kurve rechts von der gestrichelten Linie). Seine Trefferrate beträgt 80% (= Fläche unterhalb der blauen Kurve rechts von der gestrichelten Linie). Dies ergibt insgesamt einen Prozentsatz korrekter Antworten von 80%:

$$\begin{aligned} P(\text{korrekt}) &= P(\gg G \ll G) \cdot P(G) + P(\gg B \ll B) \cdot P(B) \\ &= 0.8 \cdot 0.5 + 0.8 \cdot 0.5 \end{aligned}$$

Falls der Zeuge weiss, dass die Anzahl der grünen Taxis 85% beträgt, so sollte er sein Kriterium nach rechts verschieben. Dies bedeutet, er sollte im Zweifel eher »grün« sagen, da es ja viel mehr grüne Taxis gibt. Die optimale Lage des Kriteriums für diesen Fall ist in Abb. 4-6 durch die rote durchgezogene Linie repräsentiert. In diesem Fall beträgt die Rate des falschen Alarms 0.031 und die Trefferrate beträgt 0.425. Durch das Verschieben des Kriteriums nach rechts verringern sich also beide Wahrscheinlichkeiten. Der Prozentsatz korrekter Antworten beträgt in diesem Falle 89%:

$$\begin{aligned} P(\text{korrekt}) &= P(\gg G \ll G) \cdot P(G) + P(\gg B \ll B) \cdot P(B) \\ &= 0.969 \cdot 0.85 + 0.425 \cdot 0.15 \end{aligned}$$

Die Maximierung der korrekten Antworten durch Veränderung der Antworttendenz aufgrund von geänderten Basisraten der verschiedenen Stimulustypen ist eine anerkannte Tatsache und das Signalentdeckungsmodell, welche dieses Phänomen erklärt, ist daher ein allgemein anerkanntes Modell zur Erklärung des Erkennens.

Setzt man nun aber die geänderten Likelihoods $P(\gg B \ll B) = 0.425$, sowie $P(\gg B \ll G) = 0.031$ in die Formel für das Bayes-Theorem ein, so ergibt sich die Posterior-Wahrscheinlichkeit $P(B | \gg B \ll) = 0.71$, was deutlich näher bei den modalen Schätzungen der Vpn von 0.80 liegt als die von Tversky und Kahneman berechneten »korrekten« Werte von 0.41.

Die hier gegebene ausführliche Analyse macht einen wichtigen Aspekt deutlich:

Geänderte Basisraten können zu geänderten Likelihoods führen, falls die Basisraten den Mechanismus, welcher der Generierung der Likelihoods unterliegt, beeinflusst.

Wahrscheinlichkeitstheoretisch betrachtet kann man das Ganze auch als eine Erweiterung des Problems mit den Variablen *Farbe des Taxis* und *Antwort des Zeugen* auf ein Problem mit drei Variablen betrachten, wobei die *Basisrate (BR)* als dritte Variable hinzukommt.

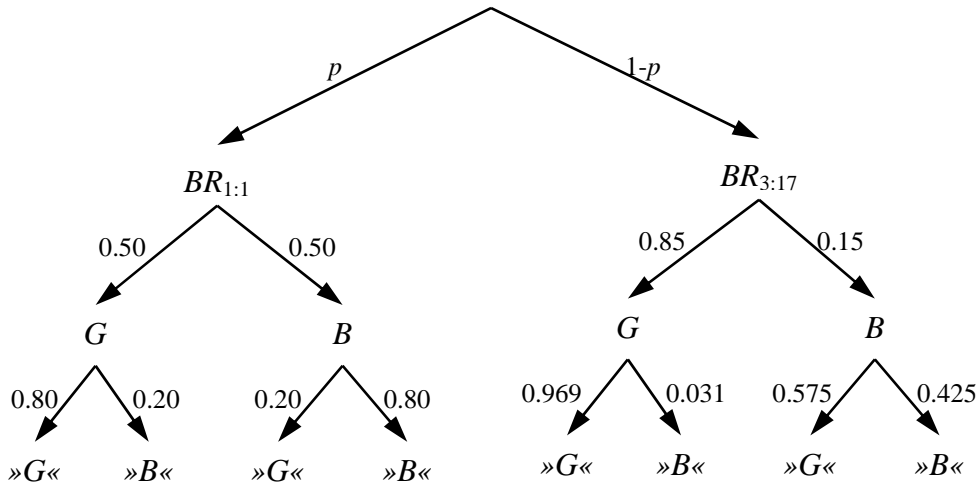


Abb. 4-7: Ereignisbaum des erweiterten Taxi-Problem mit der Variable »Basisrate« (BR) als zusätzliche Variable, mit den beiden Werten $\{BR_{1:1}$ und $BR_{3:17}\}$.

Abb. 4-7 zeigt den Ereignisbaum für das erweiterte Taxiproblem. Die Variable *Basisrate* kann die Werte $BR_{1:1}$ bzw. $BR_{3:17}$ annehmen, wobei der Index den Wettquotienten der Basisraten wiedergibt.

Gefragt sind nun nicht mehr die Wahrscheinlichkeiten $P(B|»B«)$ bzw. $P(G|»B«)$, sondern die Wahrscheinlichkeiten $P(B|»B«, BR = BR_{3:17})$ bzw. $P(G|»B«, BR = BR_{3:17})$.

Zentral ist, dass für die Likelihoods gilt:

$$P(»B«|B, BR = BR_{3:17}) \neq P(»B«|B, BR = BR_{1:1})$$

$$P(»B«|G, BR = BR_{3:17}) \neq P(»B«|G, BR = BR_{1:1})$$

Dies bedeutet, dass *keine* bedingte Unabhängigkeit zwischen der Antwort des Zeugen und der Basisrate besteht, gegeben die Farbe des Taxis, in Symbolen:

$$P(\text{Antwort}|Farbe, Basisrate) \neq P(\text{Antwort}|Farbe)$$

Es stellt sich nun die Frage, ob die vorgängige Analyse tatsächlich auf das Taxiproblem anwendbar ist. Hier zeigt sich nun, dass sich die Kritik von Birnbaum (1983) auf eine ältere Version des Taxiproblems von 1980 bezieht. Dort wurde das Problem so beschrieben, dass der Test unter den gleichen Lichtverhältnissen stattfindet wie in der Nacht des Unfalls. Ob bei der Testung des Zeugen die natürlichen Basisraten (85% grün, 15% blau) verwendet wurden, blieb offen und es wurde

von Birnbaum (1983) – ob zu Recht oder zu Unrecht – angenommen, dass im Test 50% der präsentierten Items grün und 50% blau waren. Die Kritik trifft jedoch keineswegs auf die Version von Tversky und Kahneman (1982) [Bsp.4-18] zu, denn in dieser Version wird ausdrücklich davon gesprochen, dass der Test des Zeugen *unter den gleichen Verhältnissen* durchgeführt wird, was die unterschiedlichen Basisraten mit einbezieht. Aus diesem Grund wirken sich die Basisraten während des Testens in genau der gleichen Weise auf das Urteil des Zeugen aus wie in der Originalsituation, die Likelihoods zwischen Original- und Testsituation sind daher völlig austauschbar.

Aus dem genannten Grund ist daher die Anwendung des Bayes-Theorems mit den von Tversky und Kahneman vorgegebenen Werten durchaus gerechtfertigt und die Kritik von Gigerenzer (2001), wonach inhalts-blinde Normen angewendet wurden, ist nicht angebracht.

Die Tatsache, dass seine Analyse auf die neuere Version des Problems nicht zutrifft, wird übrigens von Birnbaum (1983) selbst am Ende seines Artikels zugegeben, was jedoch Gigerenzer und Murray (1987) keineswegs darin hindert, die Schlussfolgerungen von Birnbaum zu wiederholen.

In diesem Zusammenhang stellt sich noch eine andere Frage:

Frage:



Werden derartige Überlegungen bezüglich des Wahrnehmungsmechanismus des Zeugen und dessen Beeinflussbarkeit durch Basisraten von den Probanden bei der Lösung des Problems überhaupt angestellt?

Das Taxi-Problem kann als eine Lehrbuchaufgabe in Bayes'schem Schlussfolgern betrachtet werden. Warum sollte der Proband die gemachten Angaben anzweifeln und irgendwelche Annahmen über den Wahrnehmungsmechanismus des Zeugen machen, die dann dazu führen, dass die angegebenen Wahrscheinlichkeiten in Zweifel gezogen werden? Liegen die Schwierigkeiten der Vpn nicht vielmehr darin, dass sie die Funktionsweise des Bayes-Theorems und die Bedeutung der Basisraten (wie dies durch die Baumdarstellung (vgl. Bsp.4-19) expliziert wird) nicht verstehen?

Falls diese Fragen mit *Ja* beantwortet werden können, so zeigen obige Argumente höchstens, dass die Anwendung des Bayes-Theorems mit den vorgegebenen Zahlen für die ältere Version des Taxi-Problems problematisch ist. Dies bedeutet aber noch nicht, dass die Personen die Bedeutung von Basisraten oder das Bayes-Theorem verstehen. Ein Hinweis, dass letzteres zutrifft ergibt sich aus einer Modifikation des Taxiproblems durch Gigerenzer und Hoffrage (1995). Hier wird der Zeuge beim Test in der Nacht an die Kreuzung gestellt wird, wodurch garantiert werden soll, dass dieselben Bedingungen wie bei der Identifikation vorliegen (siehe Bsp.4-26). Diese Modifikation führte jedoch zu keiner nennenswerten Verbesserung der Leistung der Vpn.

Dies zeigt, dass für die Vpn derlei Details wenig Bedeutung haben. Vielmehr sieht es so aus, dass die Vpn die Struktur des Problems nicht verstehen und damit keine adäquate Repräsentation – wie z.B. einen Ereignisbaum (siehe Bsp.4-19) – bilden können, welche ihnen helfen würde, das Problem korrekt zu lösen (vergleiche hierzu die Diskussion in Kapitel 4.4.4.4).

Das Problem der Veränderung von Likelihoods durch Basisraten wurde auch im Kontext des Basisraten-Neglects (siehe Bsp.4-16) vorgebracht und es wurde ebenfalls eine Signalentdeckungsanalyse durchgeführt (Mueser, Cowan, & Mueser, 1999). Ähnlich wie Birnbaum (1983) behaupten die Autoren auch hier, dass die Likelihoods, welche aus der diagnostischen Information erschlossen werden, mit den Basisraten variieren. Ob der von ihnen vorgeschlagene Mechanismus etwas mit den Fehlern der Vpn zu tun hat, kümmert die Autoren wenig.

Die beschriebenen Signalentdeckungsanalysen haben noch ein anderes Problem. Obwohl diese Modelle sehr verbreitet sind, so sind sie keineswegs unumstritten. Speziell die Frage, ob Kriterien verändert werden, wurde von anerkannten Autoritäten auf dem Gebiet angezweifelt (vgl. Balakrishnan 1999; Balakrishnan, und MacDonald, 2002; Van Zandt, 2000). Diese Kritik zieht auch die Sinnhaftigkeit der vorgelegten Analysen in Zweifel.

Zusammenfassend lässt sich also sagen, dass die von Birnbaum, Gigerenzer und Murray vorgebrachten Argumente gegen die Anwendung des Bayes-Theorems auf das Taxi-Problem von Tversky und Kahneman (1982) nicht gerechtfertigt sind. Zudem wurde nicht gezeigt, dass die durch die Modelle postulierten Mechanismen die schlechten Leistungen der Vpn tatsächlich erklären können.

4.4.4 Verbesserung von Wahrscheinlichkeitsurteilen

Es gibt zahlreiche Versuche, die beobachteten Defizite im Wahrscheinlichkeitsurteil zu verringern. Diese Versuche wurden oft mit der Kritik am »Heuristics and Biases« Ansatz verknüpft, indem darauf hingewiesen wurde, dass Tversky und Kahneman ihre Probleme einfach nicht »artgerecht« präsentiert hätten und die beobachteten Defizite vorwiegend auf dieser fehlerhaften Problempäsentation beruhe.

Wir haben bereits einen Ansatz zur Verbesserung des Wahrscheinlichkeitsurteils im Zusammenhang mit der Verwendung von Basisraten besprochen, nämlich die Verwendung von *kausalen Basisraten* zur Verringerung der Unterbewertung von Basisraten (vgl. Bsp.4-16). Auch im Falle des Taxi-Problems führen kausale Basisraten zu einer stärkeren Berücksichtigung derselben. Gibt man anstelle der Information, dass die grünen Taxis 85% der Taxis in der Stadt stellen und die Blauen 15% die Information, dass *die grünen Taxis in 85% der Unfälle verwickelt sind*, so führt dies zu einer stärkeren Berücksichtigung der Basisraten.

Im Folgenden wollen wir kurz einige weitere Methoden zur Verbesserung des Wahrscheinlichkeitsurteils betrachten. Wir beschränken uns hierbei auf die folgenden Methoden:

1. Erhöhung der Salienz des Zufallscharakters,
2. Präsentation der Wahrscheinlichkeitsinformation als Häufigkeiten,
3. Verwendung von »natürlichen« Häufigkeiten,
4. Verwendung spezieller Hinweisreize, welche die Problemstruktur besser kenntlich machen.

4.4.4.1 VERBESSERUNG DES WAHRSCHEINLICHKEITSURTEILS DURCH ERHÖHUNG DER SALIENZ DES ZUFALLSCHARAKTERS

Eine Kritik am »Heuristics and Biases« Ansatz lautet, dass in den Darstellungen von Tversky und Kahneman der Zufallscharakter oft nur ungenügend dargestellt wurde. Macht man deutlicher, dass es sich um einen Zufallsprozess handelt, so verringert sich auch die Verzerrung im Wahrscheinlichkeitsurteil.

Gigerenzer, Hell, und Blank (1988) demonstrierten, dass die Ignorierung von Basisraten weitgehend verschwindet, wenn die Versuchspersonen den Prozess des Ziehens einer Stichprobe selbst durchführen können.

Die Ergebnisse dieser Studie unterliegen gewisser Einschränkungen, was deren Interpretation bzw. Verallgemeinerung betrifft:

1. In neueren Studien wird fast immer darauf verwiesen, dass es sich um einen Prozess des zufälligen Ziehens einer Stichprobe handelt. Dies führt aber nicht zum Verschwinden von Fehlerurteilen. Daher kann aus der Studie von Gigerenzer et al. (1988) nicht allgemein geschlossen werden, dass die Fehler im Wahrscheinlichkeitsurteil vorwiegend auf mangelnde Hervorhebung des Zufallsprozesses beruhen.
2. Im Alltag werden Wahrscheinlichkeitsinformationen präsentiert, die fast nie einen Hinweis auf den Prozess des Ziehens der Stichprobe geben. Die meisten Leute nehmen vermutlich von selbst an, dass es sich um eine repräsentative Stichprobe handelt.
3. Die Erfahrung des Prozesses des Ziehens der Stichprobe kann dazu beigetragen haben, dass für die Vpn die Problemstruktur verständlicher wird. Wir werden diesen Aspekt im Abschnitt 4.4.4.4 näher behandeln.

Eine zweite Möglichkeit zur Verbesserung des Urteils, der ausgiebig in der Literatur diskutiert wurde, betrifft das Format, in dem die Information über die Unsicherheit präsentiert wurde.

4.4.4.2 VERBESSERUNG DES WAHRSCHEINLICHKEITSURTEILS DURCH VERWENDUNG VON HÄUFIGKEITEN

Ein weitere Kritikpunkt am »Heuristics and Biases« Ansatz besagt, dass der Umgang mit Information über Unsicherheit dadurch erschwert wird, dass letztere in Form von Wahrscheinlichkeiten oder

Prozenten präsentiert wird und nicht als absolute oder relative Häufigkeiten.

Es ist offensichtlich, dass verschiedene Arten, Unsicherheit über ein Ereignis zu kommunizieren einen Unterschied macht. Eine Zeitungsmeldung wie:

33% der Schweizer werden vermutlich an der gefährlichen Grippe XYZ erkranken.

erzeugt vermutlich eine geringere Resonanz als die Meldung:

Jeder dritte Schweizer wird vermutlich an der gefährlichen Grippe XYZ erkranken.

Eine Studie von Slovic, Monahan, & MacGregor (2000) zeigt deutlich die unterschiedliche Wirkung von frequentistischen gegenüber probabilistischen Informationsformaten.



Bsp.4-24: Frequentistische vs. probabilistische Informationsformate (Slovic, Monahan & MacGregor, 2000).

Die Autoren gaben Mitgliedern der American Academy of Psychiatry and Law aktuelle Fälle gewaltbereiter Patienten, die auf einer Seite zusammenfassend beschrieben wurden.

- Die eine Gruppe sollte die *Wahrscheinlichkeit* beurteilen, dass Mr. Jones einen Gewaltakt verüben wird, falls er auf Bewährung freikommt.
- Die zweite Gruppe sollte beurteilen, wie viele von 100 Leuten wie Mr. Jones, die auf Bewährung freikommen, einen Gewaltakt begehen würden.

Die Wahrscheinlichkeitsurteile waren eindeutig höher als die Häufigkeitsurteile.

Gigerenzer (2001) bringt ein anderes Beispiel, welches die differentielle Wirkung der unterschiedlichen Formate demonstriert:



Bsp.4-25: Frequentistische vs. probabilistische Informationsformate (Gigerenzer, 2001).

Ein Psychiater, der seinen Patienten Prozac verschrieb, informierte diese darüber, dass sie eine 30-50% Wahrscheinlichkeit hätten, ein sexuelles Problem (Verlust der Libido, sexuelle Fehlfunktion) zu entwickeln.

Nachdem er von Gigerenzers Forschungen erfahren hatte, änderte er seine Methode der Kommunikation: Anstelle von Wahrscheinlichkeiten verwendete er nun Häufigkeiten:

Der teilte nun seinen Patienten mit, dass von 10 Patienten, denen er Procak verschreibe, 3-5 ein sexuelles Problem entwickelten.

Wenn die Information im Häufigkeitsformat präsentiert wurde, waren die Klienten weniger beunruhigt.

Die Verwendung von Häufigkeiten hat auch in verschiedenen Bereichen von Wahrscheinlichkeitsurteilen zu Verbesserungen geführt. So fanden verschiedene Autoren, dass die Verwendung von Häufigkeiten zur Verringerung des Konjunktionsfehlers führt (Tversky & Kahneman, 1983; Hertwig & Gigerenzer, 1999; Sloman, Over, Slovak, & Stibel, 2003).

Auch im Bereich des Bayes'schen Schlussfolgerns wurden von Cosmides & Tooby (1996) starke Verbesserungen bei Verwendung von relativen Häufigkeiten anstelle von Wahrscheinlichkeiten gefunden. Diese Ergebnisse sind jedoch insofern problematisch als die hohen Lösungswahrscheinlichkeiten (über 70%) nicht durch andere Studien repliziert werden konnten. So fanden Sloman et al. (2003) bei exakt den gleichen Problemen nur zwischen 30% und 50% korrekte Lösungen.

Aufgrund dieser Ergebnisse stellt sich die Frage, warum Frequenzen zu einer Verbesserung des Wahrscheinlichkeitsurteils führen. Wir verschieben die Diskussion dieser Frage auf Abschnitt 4.4.4.4.

4.4.4.3 NATÜRLICHE HÄUFIGKEITEN

Wir beginnen mit einer Definition:



Konzept 4-4: *Natürliche Häufigkeiten (Kleiter, 1994):*

Natürliche Häufigkeiten spiegeln die Basisraten wider.

Die Definition lässt sich anhand eines Beispiels verdeutlichen.



Bsp.4-26: Das Taxi-Problem mit natürlichen Häufigkeiten (Gigerenzer & Hoffrage, 1995)

In einer amerikanischen Stadt gibt es zwei Taxiunternehmen, das eine hat nur grüne Taxis, die Taxis des anderen Unternehmens sind alle blau.

Nachdem ein Taxi nachts einen Unfall verursachte und der Fahrer anschließend Fahrerflucht beging, kommt es zu einer Gerichtsverhandlung. Es gibt einen Zeugen, der das davonfahrende Taxi als blau identifizierte.

Das Gericht untersucht nun die Fähigkeit des Zeugen, die Farbe eines Taxis bei Nacht richtig zu identifizieren. Dazu geht ein Gerichtsdienstler mit dem Zeugen an den Ort, von dem aus der Zeuge den Unfall beobachtet hat und lässt ihn die Farbe der zufällig vorbeifahrenden Taxis identifizieren. Dabei ergibt sich folgendes:

- 15 von je 100 Taxis, die während des Beobachtungszeitraumes vorbeifahren, sind blau.
- Von je 15 blauen Taxis werden 12 vom Zeugen als blau identifiziert.

- Von je 85 grünen Taxis werden 17 vom Zeugen als blau identifiziert.

Wie viele von je 100 vorbeifahrenden Taxis, welche vom Zeugen als blau identifiziert wurden, sind blau?

_____ von 100

Man beachte, dass die natürlichen Häufigkeiten 12 von 15 und 17 von 85 die relativen Häufigkeiten die Basisraten der Taxis (15 von 100 vs. 85 von 100) direkt widerspiegeln. Abb. 4-8 zeigt den Ereignisbaum für das Taxi-Problem mit natürlichen Häufigkeiten

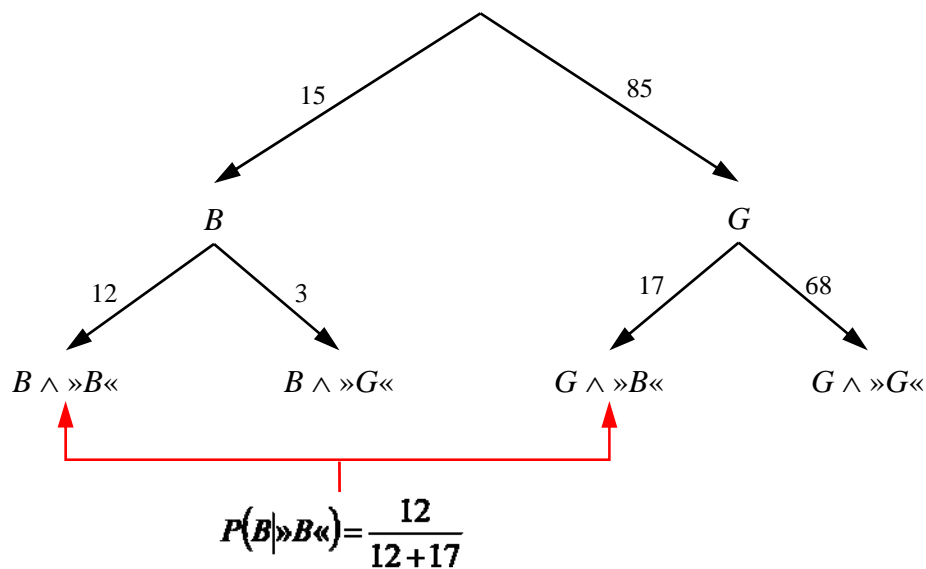


Abb. 4-8: Ereignisbaum des Taxi-Problems mit natürlichen Häufigkeiten.

Man erkennt sofort, dass das Taxi-Problem durch die Verwendung von natürlichen Häufigkeiten stark vereinfacht wird. Es stellt sich nun aber die Frage, wodurch diese Vereinfachung bewirkt wird.

Gigerenzer und Hoffrage (1995) argumentieren, dass natürliche Häufigkeiten jenes Format darstelle, mit dem unsere Vorfahren Information über Unsicherheit repräsentiert hätten. Es handelt sich also hierbei um das durch die Evolution herausgebildete Format zur Repräsentation von Unsicherheit.

Eine alternative Erklärung ergibt sich, wenn man das Problem mit natürlichen Wahrscheinlichkeiten vom Standpunkt der durchzuführenden Operationen betrachtet: In der Formulierung mit natürlichen Wahrscheinlichkeiten werden den Vpn die relevanten Verbundhäufigkeiten (nämlich die Werte 12 und 17) präsentiert. Damit muss bei Verwendung natürlicher Häufigkeiten die Operation der Kombination von Wahrscheinlichkeiten nicht durchgeführt werden. Zusätzlich vereinfacht sich die Operation der Marginalisierung, indem nur jene Ver-

bundhäufigkeiten, welche für die Durchführung der Marginalisierung benötigt werden, in der Problemstellung genannt werden.

Gemäss dieser Analyse konfundiert die Präsentation von Bayes'schen Problemen in natürlichen Häufigkeiten zwei Dinge:

1. Die Verwendung von unnormierten Häufigkeiten;
2. Die Vereinfachung des Problems durch Präsentation der relevanten Verbundhäufigkeiten, was die Operation der Kombination von Information überflüssig macht und die Operation der Marginalisierung vereinfacht.

Diese beiden Faktoren lassen sich separieren, indem man das Problem mittels Verbundwahrscheinlichkeiten präsentiert.

- 12% der der Taxis sind blau und wurden vom Zeugen als blau identifiziert.
- 17% der der Taxis sind grün und wurden vom Zeugen als blau identifiziert.

Falls der zentrale Faktor für die Vereinfachung des Problems in der Präsentation von Verbundwahrscheinlichkeiten und der damit einhergehenden Vereinfachung der durchzuführenden Operationen liegt, so sollte die Formulierung des Taxi-Problems (und anderer Bayes'scher Probleme) mittels Verbundwahrscheinlichkeiten eine ähnlich positive Wirkung haben wie die Verwendung von natürlichen Häufigkeiten. Genau dies wurde von Keren und Lewis (1999) gezeigt. Die von Gigerenzer und Hoffrage (1995) vorgebrachte Erklärung ist daher wenig plausibel.

4.4.4.4 ERHÖHUNG DER VERSTÄNDLICHKEIT DER PROBLEMSTRUKTUR

In Experimenten zu Wahrscheinlichkeitsurteilen werden den Probanden meist Textprobleme präsentiert (wie beim Taxi-Problem oder beim Linda-Problem). Es bietet sich daher an, die Erkenntnisse aus der klassischen Problemlöseliteratur auf diese Experimente anzuwenden.

Die zentrale Erkenntnis der Problemlöseforschung besteht darin, dass der wesentliche Schritt zur Lösung eines Problems darin besteht, eine geeignete mentale Repräsentation des Problems zu generieren. Wendet man diese Erkenntnis auf die klassischen Wahrscheinlichkeitsprobleme an, so ergeben sich einige interessante Einsichten.

Beginnen wir zuerst mit den Bayes'schen Problemen und nehmen wir wiederum als Beispiel das Taxi-Problem. Die Schwierigkeit des Problems besteht hier darin, dass die Vpn keine adäquate Repräsentation der verschiedenen Möglichkeiten, welche zum »Blau« Urteil des Zeugen führen, bilden. Im Speziellen repräsentieren sie nicht adäquat das Zustandekommen der vielen fehlerhaften Falschen Alarme aufgrund des Überwiegens grüner Taxis in der Grundgesamtheit. Sobald dieser Aspekt verstanden wird, sollte sich die Lösung ergeben.

Man beachte, dass bei Präsentation des Problems mittels natürlicher Wahrscheinlichkeiten bzw. mittels Verbundhäufigkeiten dieses Pro-

blem umgangen wird, da in die Verbundwahrscheinlichkeiten die Basisraten integriert sind. Daher kann in diesem Falle das Problem gelöst werden, ohne dass die Problemstruktur voll verstanden wird.

Ein ähnliches Argument ergibt sich für das Linda-Problem und den beobachteten Konjunktionsfehler. In diesem Falle geht das Verhältnis der beiden Mengen *Bankangestellte* und *feministische Bankangestellte* nicht in Repräsentation der Vpn ein, was den Konjunktionsfehler zur Folge hat.

Folgt man der hier dargestellten Argumentation, so ergeben sich zwei wichtige Schlussfolgerungen:

1. Jede Art von Hinweis, welche das Verständnis der Problemstruktur erhöht, sollte zu einer verbesserten Leistung führen.
2. Die Verbesserungen der Leistung bei Verwendung von Häufigkeiten anstelle von Wahrscheinlichkeiten und durch Hervorhebung des Zufallscharakters (vgl. Absatz 4.4.4.1 und 4.4.4.2) sollten darauf zurückzuführen sein, dass diese Manipulationen das Verständnis für die Problemstruktur erhöhen.

Um dies zu testen, haben Sloman et al. (2003) eine Reihe von Experimenten durchgeführt, wobei sie in einer Bedingung die beteiligten Mengen und deren Einbettung klar gemacht haben. Diese Klärung der Beziehung der relevanten Mengen zueinander führte zu einer klaren Verbesserung der Leistung der Vpn und zwar unabhängig davon, ob das Problem mittels Wahrscheinlichkeiten oder Häufigkeiten formuliert war.

Umgekehrt führte die Verunklarung der Einbettung der beteiligten Mengen (z.B. beim Linda-Problem, indem zwischen den beiden relevanten Fragen weitere Fragen eingeschoben wurden) zu einer Verschlechterung der Leistung und zwar unabhängig davon, ob Häufigkeiten oder Wahrscheinlichkeiten verwendet wurden. Der beobachtete Vorteil von Häufigkeiten wie er z.B. von Hertwig und Gigerenzer (1999) beobachtet wurde, löste sich dabei in »blauen Dunst« auf.

Aufgrund dieser Ergebnisse ergibt sich eine klare Schlussfolgerung:

Die zentrale Ursache für Fehler im Wahrscheinlichkeitsurteil liegt nicht – wie von Gigerenzer, Cosmides, Tooby et. al. behauptet – in der Tatsache, dass unser Modul zur Schlussfolgerung mit unsicherer Information mit Häufigkeiten anstelle von Wahrscheinlichkeiten arbeitet, sondern in der Tatsache, dass die Probleme nicht adäquat repräsentiert werden. Jede Methode, welche die Problemstruktur klärt, wird daher zu einer Verbesserung führen.

4.5 Übungen



Übung 4-1:

Gegeben:

Ein Experiment, mit drei möglichen Ereignissen: Es leuchte entweder eine rote (60% der Fälle) oder eine grüne (30%) oder eine blaue Lampe (10%) auf.

Die Vp, welche das Aufleuchten der Lampe für den jeweils nächsten Durchgang vorhersagen soll, macht ein Probability-Matching.

Geben Sie an:

- (i) Wie gross ist der durchschnittliche Prozentsatz korrekter Vorhersagen der Vp?
- (ii) Was ist der maximal zu erreichende durchschnittliche Prozentsatz?



Übung 4-2:

Geben Sie ein Beispiel zur Demonstration der Asymmetrie bedingter Wahrscheinlichkeiten. Zeichnen Sie ein Venn-Diagramm und erklären Sie anhand desselben, wie es zu dieser Asymmetrie kommt.



Übung 4-3:

Beweisen Sie die folgende Beziehung:

$$P(A|B) > P(A|\bar{B}) \Leftrightarrow P(B|A) > P(B|\bar{A})$$



Übung 4-4:

Gegeben: Die folgende Problemstellung:

Aufgrund der aktuellen Forschung weiss man, dass ein Lügendetektor in 90% der Fälle positiv reagiert, falls die Person schuldig ist.

Er reagiert aber auch in 50% der Fälle positiv, wenn die Person unschuldig ist.

Von den Personen, bei denen der Test angewendet wird, sind im Durchschnitt 10 % schuldig.

Wie wahrscheinlich ist es, dass die Person lügt, falls der Detektor positiv reagiert?

- (i) Repräsentieren Sie das Problem als Ereignisbaum und berechnen Sie hieraus die gesuchte Wahrscheinlichkeit.
- (ii) Berechnen Sie die gesuchte Wahrscheinlichkeit mit Hilfe des Bayes-Theorems in Wettquotienten-Format.



Übung 4-5:

Gegeben:

Ein Betrieb zur Herstellung von Druckformen hat drei Maschinen, eine neue Maschine M_1 und eine ältere M_2 und eine ganz alte, M_3 .

M_1 produziert täglich 5000 Druckformen, M_2 3000 und M_3 1000.

Die Fehlerraten der Maschinen unterscheiden sich ebenfalls:

M_1 produziert durchschnittlich 1 falsche Form auf 300 Stück.

M_2 produziert durchschnittlich 1 falsche Form auf 100 Stück.

M_3 produziert durchschnittlich 1 falsche Form auf 70 Stück.

Der Produktionsleiter zieht 1 Druckformen zufällig aus den an diesem Tag produzierten und bemerkt, dass sie fehlerhaft ist. Wie wahrscheinlich ist es, dass es sich um eine von M_1 , M_2 bzw. M_3 produzierte Druckform handelt?

- (i) Repräsentieren Sie das Problem als Ereignisbaum und berechnen Sie hieraus die gesuchte Wahrscheinlichkeit.
- (ii) Berechnen Sie die gesuchte Wahrscheinlichkeit mit Hilfe des Bayes-Theorems in Wettquotienten-Format.



Übung 4-6: Ziegenproblem

- (i) Repräsentieren Sie das Ziegenproblem (Bsp.4-22) mit Hilfe eines Ereignisbaumes, unter der Annahme, dass der Moderator keine Präferenz für eine der nicht gewählten Türen B und C aufweist, d.h. er öffnet Tür B und C mit gleicher Wahrscheinlichkeit von $1/2$, wenn sich der Preis hinter A befindet

Berechne die Wahrscheinlichkeit, dass die Kandidatin gewinnt, wenn der Moderator Tür B geöffnet hat und sie nicht wechselt.

- (ii) Wiederhole Aufgabe (i) jedoch unter der Annahme, dass der Moderator Tür B bevorzugt, d.h. er öffnet immer Tür B , wenn sich der Preis hinter A befindet.

5. Literatur

- Abelson, R. P. (1963). Computer simulation of "hot cognition". in S. S. Tomkins & S. Messick (Eds.), *Computer simulation of personality* (pp. 277-302). New York: Wiley.
- Ayton, P. & Fischer, I. (2004). The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness? *Memory & Cognition*, 32, 1369-1378.
- Balakrishnan, J. D. (1999). Decision processes in discrimination: Fundamental misrepresentations of signal detection theory. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 1189-1206.
- Balakrishnan, J. D., & MacDonald, J. A. (2002). Decision criteria do not shift: Reply to Treisman. *Psychonomic Bulletin & Review*, 9, 848-865.
- Bartlett, F. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.
- Birnbaum, M. H. (1983). Base rates in Bayesian inference: Signal detection analysis of the cab problem. *American Journal of Psychology*, 96, 85-94.
- Brase, G. L., Cosmides, L., & Tooby, J. (1998). Individuation, counting, and statistical inference: The role of frequency and whole-object representations in judgments under uncertainty. *Journal of Experimental Psychology: General*, 123, 3-21.
- Carroll, J. S. (1978). The effect of imaging an event on expectations for the event: An interpretation in terms of the availability heuristic. *Journal of Experimental Social Psychology*, 14, 88-96.
- Chapman, L. J., & Chapman, E. J. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, 74, 193-204.
- Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences*, 4, 317-370.
- Conway, M., & Ross, M. (1984). Getting what you want by revising what you had. *Journal of Personality and Social Psychology*, 47, 738-748.
- Cooper, w. H. (1981). Ubiquitous halo. *Psychological Bulletin*, 90, 218-244. .

- Cosmides, L. & Tooby, J. (1996). Are humans good intuitive statisticians after all ? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, 58, 1-73.
- Darley, J. M., & Batson, C. D. (1973). "From Jerusalem to Jericho": A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology*, 27, 100-108.
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, 58, 17-22.
- Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases* (pp. 249-267). Cambridge, England: Cambridge University Press.
- Ernst, E., & Singh, S. (2009). *Gesund ohne Pillen - was kann die Alternativmedizin?* München: Hanser.
- Fischhoff, B. & Beyth, R. (1975). "I knew it would happen": Remembered probabilities on once-future things. *Organizational Behavior and Human Performance*, 13, 1-16.
- Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology*, 65, 45-55.
- Gallistel, C. R. (1990). *The organization of learning*. Cambridge, Mass.: MIT Press.
- Gerry, M, Manning, C. G., Loftus, E. F., & Sherman, S. J. (1996). Imagination inflation: Imaging a childhood event inflates confidence that it occurred. *Psychonomic Bulletin & Review*, 3, 208-214.
- Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review*, 103, 592-596.
- Gigerenzer, G. (2001). Content-blind norms, no norms, or good norms? A reply to Vranas. *Cognition*, 81, 93-103.
- Gigerenzer, G. Hell, W., & Blank, H. (1988). Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 513-525.

- Gigerenzer, G. & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102, 684-704.
- Gigerenzer, G. & Murray, D. (1987). *Cognition as intuitive statistics*. Hillsdale, NJ: Erlbaum.
- Gilovich, T., & Savitsky (2002). Like goes with like: The role of representativeness in erroneous and pseudo-scientific beliefs. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp.617-624). Cambridge, UK: Cambridge University Press.
- Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, 17, 295-314.
- Hacking, I. (2001). *An introduction to probability and inductive logic*. Cambridge: Cambridge University Press.
- Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgment. *Journal of Experimental Social Psychology*, 12, 392-407.
- Harper, D. G. C. (1982). Competitive foraging in mallards: Ideal free ducks. *Animal Behaviour*, 30, 575-584.
- Janis, I. L., & Mann, L. (1979). *Decision making: A psychological analysis of conflict, choice, and commitment*. New York: Free Press.
- Johnson, M. K., Hashtroudi, S., & Lindsay, S. D. (1993). Source monitoring. *Psychological Bulletin*, 114, 3-28.
- Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, 88, 67-85.
- Juslin, P. (2001). Representative Design: Cognitive science from a Brunswikian perspective. In K. R. Hammond, & Steward, T. S. (Eds.), *The essential Brunswik: Beginnings, explications, applications* (pp. 404-408). Oxford: Oxford University Press.
- Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp.46-81). New York: Cambridge University Press.
- Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less. Adding a better end. *Psychological Science*, 4, 401-405.

- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237-251.
- Keren, G. & Lewis, C. (1999). On the difficulties underlying Bayesian reasoning: A comment on Gigerenzer and Hoffrage. *Psychological Review*, 106, 411-416.
- Kleiter, G. (1994). Natural sampling: Rationality without base rates. In G. H. Fischer & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics, and methodology* (pp. 375-388). New York: Springer.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108, 480-498.
- Larkey, P. D., Smith, R. A., & Kadane, J. B. (1989). It's okay to believe in the "hot hand". *Chance: New directions for Statistics and Computing*, 2, 22-30.
- Loftus, E. F. (1980). *Memory*. Reading, Mass.: Addison Wesley.
- Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, 13, 585-589.
- Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin*, 68, 304-305. .
- Marcus, G. B. (1968). Stability and change in political attitudes: Observed, recalled and explained. *Political Behavior*, 8, 19-31.
- Mather, M., Johnson, M. K., & DeLeonardis, D. M. (1999). Stereotype reliance in source monitoring: Age differences and neuropsychological test correlates. *Cognitive Neuropsychology*, 24, 437-458.
- McCloskey, M., & Zaragossa, M. (1985). Misleading postevent information and memory for events: Arguments against the memory impairment hypothesis. *Journal of Experimental Psychology: General*, 114, 1-16.
- Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, 67, 371-378.
- Miller, A. G., Gillan, B., Schenker, C., & Radlove, S. (1973). Perception of obedience to authority. *Proceedings of the 81st Annual Convention of the American Psychological Association*, 8, 213-225.
- Mitchell, K. J. & Johnson, M. K. (2000). Source monitoring: Attributing mental experiences. In E. Tulving, & F. I. M. Craig

- (Eds.), *The Oxford handbook of memory* (Chapter 12, pp. 179-195). Oxford: Oxford University Press.
- Mitchell, K. J. & Johnson, M. K. (2009). Source monitoring 15 years later: What have we learned from fMRI about the neural mechanisms of source memory? *Psychological Bulletin*, *135*, 638-677.
- Morris, M. W., & Peng, K. (1994) Culture and cause: American and Chinese attributions for social and physical events. *Journal of Personality and Social Psychology*, *67*, 949-971.
- Mosteller, F. (1965). *Fifty challenging problems in probability with solutions*. Reading, MA: Addison Wesley.
- Mueser, R. P., Cowan, N., & Mueser, K. T. (1999). A generalized signal detection model to predict rational variation in base rate use. *Cognition*, *69*, 267-312.
- Nisbett, R. E., & Borgida, (1981).
- Nisbett, R. E. & Ross, L. (1980). *Human inference: Strategies and shortcomings of human judgment*. Englewood Cliffs, NJ: Prentice-Hall.
- Paik, M. (1985). A graphical representation of a three-way contingency table: Simpson's paradox and correlation. *American Statistician*, *39*, 53-54.
- Pickrell, J. E., Bernstein, D. M., & Loftus, E. F. (2004). Misinformation effect. In R. Pohl (Ed.), *Cognitive Illusions: A handbook of fallacies and biases in thinking, judgment and memory* (pp. 345-361). Hove, UK: Psychology Press.
- Pietromonaco, P. R., & Nisbett, R. E. (1982). Swimming upstream against the fundamental attribution error: Subjects' weak generalizations from the Darley and Batson study. *Social Behavior and Personality*, *28*, 641-651.
- Pinker, S. (1997). *How the mind works*. New York: Norton.
- Popper, K. R. (1984). *Objektive Erkenntnis: Ein evolutionärer Entwurf* (4. Auflage). Hamburg: Hoffmann & Campe.
- Redelmeier, D., Katz, J., & Kahneman, D. (2003). Memories of colonoscopy: A randomized trial. *Pain*, *104*, 187-194.
- Rode, C., Cosmides, L., Hell, W., & Tooby, J. (1999). When and why do people avoid unknown probabilities in decisions under uncertainty? Testing some predictions from optimal foraging theory. *Cognition*, *72*, 269-304.

- Roediger, H. L., & Gallo, D. A. (2004). Associative memory illusions. In R. Pohl (Ed.), *Cognitive Illusions: A handbook of fallacies and biases in thinking, judgment and memory* (pp. 308-326). Hove, UK: Psychology Press.
- Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *21*, 803-814.
- Ross, L., Amabile, T. M., & Steinmetz, J. L. (1977). Social roles, social control and biases in social-perception processes. *Journal of Personality and Social Psychology*, *35*, 485-494.
- Ross, L., Lepper, M. R., Strack, F., & Steinmetz, J. L. (1977). Social explanation and social expectation: The effect of real and hypothetical explanation upon subjective likelihood. *Journal of Personality and Social Psychology*, *35*, 817-829.
- Ross, M. (1989). Relation of implicit theories to the construction of personal histories. *Psychological Review*, *96*, 341-357.
- Rotton, J. & Kelly, I. W. (1985). Much ado about the full moon: A meta-analysis of lunar-lunacy research. *Psychological Bulletin*, *97*, 286-306.
- Schacter, D. L. (1999). *Wir sind Erinnerung: Gedächtnis und Persönlichkeit*. Hamburg: Rowohlt.
- Schustack, M. W. & Sternberg, R. J. (1981). Evaluation of evidence in causal inference. *Journal of Experimental Psychology: General*, *110*, 101-120.
- Sherman, S. J., Cialdini, R. B., Schwartzman, D. F., & Reynolds, K. D. (1985). Imagining can heighten or lower the perceived likelihood of contracting a disease: the mediating effect of ease of imagery. *Personality and Social Psychology Bulletin*, *11*, 118-127.
- Slovic, P. & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgement. *Organizational Behavior and Human Performance*, *6*, 649-744.
- Slovic, P., Monahan, J., & MacGregor, D. G. (2000). Violence risk assessment, and risk communication: The effects of using actual cases, providing instructions, and employing probability versus frequency formats. *Law and Human Behavior*, *24*, 271-296.

- Smith, J. M. N., & Dawkins, R. (1971). The hunting behavior of individual Great Tits in relation to variations in their food density. *Animal Behaviour*, *19*, 695-706.
- Stainton, R. J. (2006). *Contemporary debates in cognitive science*. Malden, Mass: Blackwell.
- Stanovich, K. E. (2010). Decision making and rationality in the modern world. Oxford, UK: Oxford University Press.
- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate?. *Behavioral and Brain Sciences*, *23*, 645-726.
- Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, *4*, 469-477.
- Tu, Y-K., Gunnell, D., Gilthorpe, M. S. (2008). Simpson's paradox, Lord's paradox, and suppression effects are the same phenomenon – the reversal paradox. *Emerging Themes in Epidemiology*, *5*, 2.
Available from: <http://www.ete-online.com/content/5/1/2>.
- Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, *76*, 31-48.
- Tversky, A., & Gilovich, T. (1989a). The cold facts about the "hot hand" in basketball. *Chance: New directions for Statistics and Computing*, *2*, 16-21.
- Tversky, A., & Gilovich, T. (1989a). The "hot hand": Statistical reality or cognitive illusion? *Chance: New directions for Statistics and Computing*, *2*, 31-34.
- Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, *76*, 105-110.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207-232.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 453-458.
- Tversky, A., & Kahneman, D. (1982a). Causal schemas in judgments under uncertainty. In: D. Kahneman, P. Slovic, & A. Tversky, (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 117-128). Cambridge, England: Cambridge University Press.
- Tversky, A. & Kahneman, D. (1982b). Evidential impact of base rates. In: D. Kahneman, P. Slovic, & A. Tversky, (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 153-160). Cambridge, England: Cambridge University Press.

- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90, 293-315.
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representations of uncertainty. *Journal of Risk and Uncertainty*, 5, 297-323.
- Van Zandt, T. (2000). ROC Curves and confidence judgments in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 26, 582-600.
- von Randow, G. (1994). *Das Ziegenproblem: Denken in Wahrscheinlichkeiten*. Hamburg: Rowohlt.
- Vranas, P. B. M. (2000). Gigerenzer's normative critique of Kahneman and Tversky. *Cognition*, 76, 179-193. .
- Vranas, P. B. M. (2001). Single-case probabilities and content-neutral norms: A reply to Gigerenzer. *Cognition*, 81, 1059-111.
- Wagenaar, W. A. (1970). Appreciation of conditional probabilities in binary sequences. *Acta Psychologica*, 34, 348-356.
- Wainer, H., & Brown, M. L. (2007). Three statistical paradoxes in the interpretation of group differences: Illustrated with medical school admission and licensing data. In: C. R. Rao, & S. Sinharay (Eds.), *Handbook of statistics, Vol.26: Psychometrics* (pp. 893-918). Amsterdam: Elsevier. .
- West, R. F., & Stanovich, K. E. (2003). Is probability matching smart? Associations between probabilistic choices and cognitive ability. *Memory & Cognition*, 31, 243-251.
- Wilson, A., & Ross, M. (2004). Illusions of change and stability. In R. Pohl (Ed.), *Cognitive Illusions: A handbook of fallacies and biases in thinking, judgment and memory* (pp. 379-396). Hove, UK: Psychology Press.

6. Bücher zum Thema

- Baron, J. (2008). *Thinking and deciding*. Cambridge: Cambridge University Press. [FSES M121, TEB-26109, BP2 Ökonomie]
- Ein umfassendes Werk zum Thema Denken und Entscheiden. Leider etwas langatmig und nicht sehr spannende geschrieben.*
- Dawes, R. M. (1988/2001). *Rational choice in an uncertain world*. Philadelphia: Harcourt Brace Jovanovich. [HAED Q-1160, HPAED Q-1819].
- Ein interessantes Buch, welches ausgewählte Themen zu Entscheiden und Urteilen darstellt. Gut und kompetent geschrieben, informativ.*

Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth*. New York: Free Press. [HPAED Q-1510]

Ein sehr interessantes Buch zum Thema Expertentum in der Psychologie.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. New York: Cambridge University Press. [SPAED L-2-508]

Eine Sammlung von Aufsätzen zu verschiedensten Themen über Urteile und Urteilsfehler. Die meisten der Aufsätze sind bereits in Journalen publiziert worden. Dieses Buch ist nicht dazu gedacht als Ganzes gelesen zu werden. Es handelt sich um eine Fortsetzung des Werkes von Kahneman, Slovic und Tversky (1982) [siehe unten].

Kahneman, D., Slovic, P. & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press. [HPAED Q- 508]

Eine Sammlung von klassischen Aufsätzen zu verschiedensten Themen über Urteile und Urteilsfehler. Informativ, aber teilweise etwas langatmig.

Lilienfeld, S. O., Lynn, S. J., Lohr, J. M. (2003). *Science and pseudoscience in clinical psychology*. New York: Guilford Press. HPAED X-1615.

Ein Herausgeberbuch, in dem verschiedene Spezialisten Probleme, Kontroversen und auch zwielichtige Praktiken in der Psychotherapie behandeln.

Lilienfeld, S. O., Lynn, S. J., Rusco, J., & Beyerstein, B. L. (2010). *Fifty great myth of popular psychology: Shattering widespread misconceptions about human behavior*. Chichester: Wiley-Blackwell, [HAED Q-1927]

Eine informative und über weite Strecken sehr amüsante Darstellung verschiedener populärer fehlerhafte psychologischer Theorien.

Nisbett, R. E. & Ross, L. (1980). *Human inference: Strategies and shortcomings of human judgment*. Englewood Cliffs, NJ: Prentice-Hall. [HPAED Q-1060]

Ein Klassiker und eines meiner Tophits! Trotz des hohen Alters des Buches immer noch sehr informativ. Sehr gut geschrieben und unterhaltsam. Noch immer sehr empfehlenswert.

Plous, S. (1993). *The psychology of judgment and decision making*. New York: McGraw-Hill. [HPAED Q-1157]

Eine kompakte und sehr gut geschriebene Zusammenfassung der wichtigsten Aspekte von Entscheidungen und Urteilen. Informativ und unterhaltsam. Das Buch hat den William James Book Award erhalten.

Pohl, R. (2004). *Cognitive Illusions: A handbook of fallacies and biases in thinking, judgment and memory*. Hove, UK: Psychology Press. [HPAED Q-1713]

Gibt einen Überblick zur Psychologie der Urteilsfehler. Das Besondere and dem Buch besteht darin, dass zu den besprochen Themen konkrete Anweisung zur Durchführung einer Studie gegeben werden. Ein Nachteil besteht in der sehr schlecht geführten Rationalitätsdebatte: Das Buch will es jedem Recht machen. Leider enthält das Buch auch Fehler.

Rosenzweig, P. (2007). *The halo effect and eight other business delusions that deceive managers*. New York. Free Press. [HPAED W-2077]

Das Buch behandelt die negativen Auswirkungen des Halo-Effekts auf (halb-) wissenschaftliche Untersuchungen und journalistische Berichte über die Ursachen des Erfolgs von Unternehmen. Neben der Wirkung des Haloeffekts werden noch andere Probleme und Fehler in der Beurteilung von Manager und Unternehmen aufgezeigt. Das Buch ist unterhaltsam und leicht verständlich geschrieben und es ist äusserst interessant und lehrreich.

Stanovich, K. E. (2010). *Decision making and rationality in the modern world*. Oxford, UK: Oxford University Press. [HPAED Q-1920]

Kurz, informativ und gut geschrieben. Enthält auch eine gute Diskussion zur Rationalitätsdebatte. Empfehlenswert.

Sutherland, S. (2007). *Irrationality*. London: Pinter & Martin. [HPAED Q-1900]

Amüsant und gut geschrieben. Gibt einen guten Überblick über eine sehr breite Palette von Urteilsfehler. Einfach zu lesen.

7. Anhang: Paik-Diagramme zu den Beispielen zum Paradoxon von Simpson

Im diesem Anhang werden die Diagramme von Paik (1985) zur graphischen Veranschaulichung des Simpson-Paradoxons in Bsp.2-20, Bsp.2-21 und Bsp.2-22 präsentiert.

Die x -Achse repräsentiert jeweils die relevante unabhängige Variable. Die y -Achse repräsentiert die relevanten Anteile für die Abhängige Variable. Die Variable, über welche summiert wird, wird durch die Bezeichnungen innerhalb der Graphik kenntlich gemacht.

Die relative Anzahl von Beobachtungen innerhalb der verschiedenen Bedingungen (Gruppen) wird durch die relative Grösse der (blauen) Kreise repräsentiert.

Bsp.2-20: Simpson-Paradoxon: Todesurteile in Florida zwischen 1972-1979:

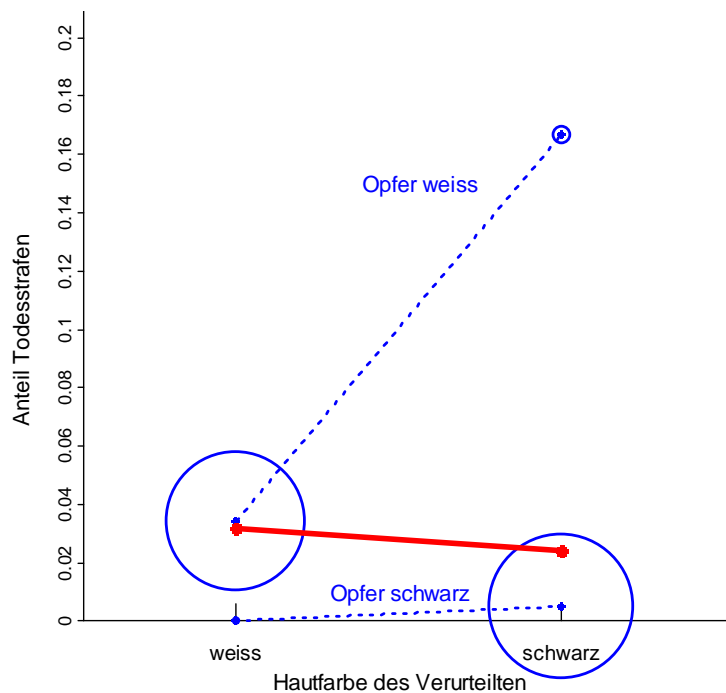


Abb. 7-1: Diagramm von Paik zur Veranschaulichung des Effekts der Summation über die Variable Hautfarbe des Opfers: Der positive Zusammenhang zwischen Todesurteilen und der Hautfarbe des Verurteilten (repräsentiert durch die gestrichelten blauen Linien) verwandelt sich bei Ignorieren der Hautfarbe des Opfers in einen negativen Zusammenhang (rote Linie).

Die blauen Symbole beziehen sich auf die Daten in der vollen Tabelle.

Die Endpunkte der dicken roten Linie repräsentieren die Anteile auf der abhängigen Variablen für die zusammengefassten Daten (marginale Tabelle).

Bsp.2-21: Simpson-Pardoxon: Wirkung einer neuen Therapie:

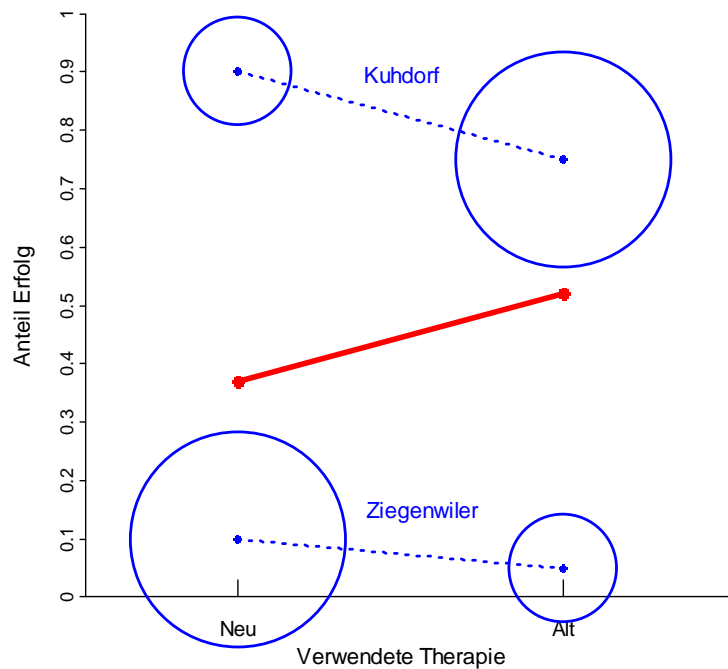


Abb. 7-2: Diagramm von Paik zur Veranschaulichung des Effekts der Summation über die Ortschaften: Der positive Zusammenhang zwischen neuer Therapie und Erfolg (repräsentiert durch die gestrichelten blauen Linien) verwandelt sich bei Ignorieren der Ortschaften in einen negativen Zusammenhang (rote Linie).

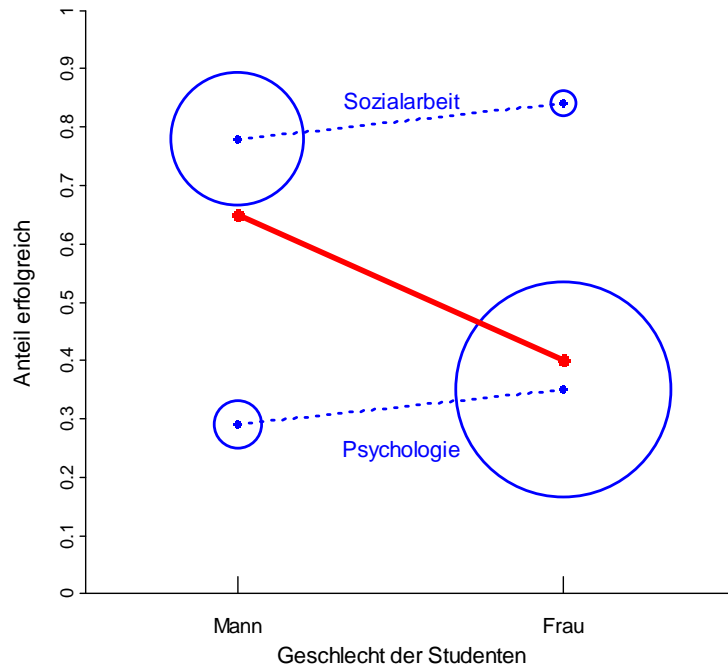
Bsp.2-22: Simpson-Pardoxon: Diskriminierung von Frauen:

Abb. 7-3: Diagramm von Paik zur Veranschaulichung des Effekts der Summation über die Studienfächer: Der positive Zusammenhang zwischen weiblichem Geschlecht und Studienerfolg (repräsentiert durch die gestrichelten blauen Linien) verwandelt sich bei Ignorieren der Studienfächer in einen negativen Zusammenhang (rote Linie).