

## **Induktives Schließen: Umgang mit Wahrscheinlichkeiten**

Klaus Fiedler & Henning Plessner

### BEGRIFFLICHE UND METHODISCHE GRUNDLAGEN

#### **Was versteht man unter "induktiv-statistischem Schließen"?**

Der Gegenstandsbereich dieses Kapitels lässt sich am besten extensional definieren, durch Aufzählen der Forschungs-Paradigmata, die man gewöhnlich dem Thema induktiv-statistisches Schließen zurechnet. Diese befassen sich mit Kategorisieren und Vorhersagen von Ereignissen unter Unsicherheit, subjektiver Wahrscheinlichkeit, Erkennen und Nutzen von Kontingenzen, und komplexeren Formen des Regellernens (vgl. Holland, Holyoak, Nisbett & Thagard, 1986; Holyoak & Nisbett, 1988; Simon & Kotovsky, 1963). All diesen experimentellen Paradigmata gemeinsam sind die beiden Konstituenten des Begriffs „induktiv-statistisches Schließen“. Einerseits besagt die Bezeichnung „induktives Schließen“, dass es um datengetriebene kognitive Prozesse geht, also um Schlüsse vom Konkreten zum Abstrakten, von beobachteten Stimulusdaten auf latente dahinter stehende Eigenschaften. Andererseits wird durch die Einschränkung „induktiv-statistisch“ deutlich gemacht, dass es sich um Schlüsse aufgrund nicht-deterministischer, fehlerbehafteter Regeln handelt. Es geht mithin in diesem Kapitel häufig – ganz analog zur Methodik der Inferenzstatistik – um die Erforschung der kognitiven Prozesse beim Schließen von Stichprobendaten auf latente Eigenschaften einer Population.

Wann werden diese kognitiven Funktionen benötigt? Bei welchen realen Problemsituationen, bei welchen realen Urteils- und Entscheidungsaufgaben sind induktiv-statistische Leistungen maßgeblich beteiligt? – Zentrale Anwendungsbereiche sind insbesondere: Probleme der diagnostischen Klassifikation (z. B. Schlüsse von beobachteten Symptommustern auf eine ursächliche Krankheit), der Prognose (z. B. Vorhersage von

Gewinn oder Verlust aufgrund aktueller Daten am Aktienmarkt), Risikoabschätzung (z. B. Bewertung des Risikos der Invalidität in einer bestimmten beruflichen Situation), Erkennen von Umwelt-Kontingenzen zwischen Ursachen und Wirkungen (z. B. Ernährung und Wohlbefinden) oder zwischen Indikatoren und verdeckten Realitäten (z. B. Körpersprache und Täuschung).

Auch an diesen Anwendungsfeldern sieht man wieder: Verlangt werden Schlüsse von einer gegebenen Datenstichprobe (z.B. medizinische Symptome, Aktienmarkt-Daten, berufliche Situation) auf latente, (noch) nicht direkt beobachtbare Gegebenheiten (Krankheit, Gewinne oder Verluste, Invalidität). Und diese Schlüsse müssen unter Unsicherheit gezogen werden, weil eine deterministische, fehlerfreie Regel nicht bekannt ist. Kategorisierungen (von Krankheiten), Vorhersagen (von Aktien) oder Entdeckungen (von Kontingenzen, Täuschungen) sind immer nur mit einer bestimmten Wahrscheinlichkeit möglich.

Typisch für den genannten Gegenstands- und Anwendungsbereich ist auch, dass es für solche Probleme meist Experten gibt (z. B. Entscheidungsforscher, Finanzmanager, Mediziner), die Expertenwissen und formale oder technische Methoden für die beste Lösung derartiger Probleme einsetzen. Aber dieselben Probleme stellen sich eben nicht nur Experten und Wissenschaftlern. Auch Laien und Novizen müssen im Alltag ständig entscheiden, wie sie auf körperliche Symptome reagieren, wie sie ihr Geld anlegen, wie sie sich ernähren, welche Bedeutung sie Signalen und Zeichen beimessen oder welche Ursache für einen Konflikt oder einen Unfall verantwortlich ist. Das heißt, alle Individuen sind zu kognitiven Schlüssen gezwungen, die denen ähneln, für die auch die wissenschaftliche Forschung und Methodik optimale Verfahren entwickelt hat. Es liegt daher in der Natur der Sache, dass man menschliche Leistungen bei induktiv-statistischen Problemen an der besten (bekanntesten) wissenschaftlichen Vorgehensweise misst. Oder anders ausgedrückt: es muss die Unterscheidung zwischen normativen Modellen (über optimale, rationale Problemlösung) und deskriptiven Modellen (über tatsächlich eingesetzte, oft nicht rationalen Lösungsstrategien)

beachtet werden. Bei genauerem, kritischem Hinsehen stellt sich indessen oftmals heraus, dass scheinbar rationale, normative Verfahren auf strittigen oder gar unhaltbaren Annahmen beruhen können und dass scheinbar irrationale kognitive Täuschungen durchaus einen Anpassungswert besitzen und zuweilen zu besseren Leistungen führen als aufwendige mathematische und logische Modelle (vgl. Fiedler, 1993).

Einige Paradigmata der kognitiv-psychologischen Forschung, die sich im Sinne dieser begrifflichen Systematik zumindest teilweise mit induktiv-statistischen Funktionen beschäftigen, werden hier indessen weitestgehend ausgeklammert, weil sich andere Kapitel ausführlich damit befassen. Hierzu zählt das gesamte weite Gebiet des kausalen Schließens (vgl. Waldmann & Hagmayer, Kapitel 5), die Erforschung von Analogie-Schlüssen bei der Kreativität (vgl. Hany, Kapitel 4) sowie die Rolle induktiver Leistungen für die Anpassung von Organismen an eine komplexe, schwer vorhersagbare Umwelt (vgl. Gigerenzer, Kapitel 7).

### **Eine kleine Taxonomie induktiver Funktionen**

Eine kleine Systematik der Begriffe, die im folgenden verwendet werden, sollte die Verständlichkeit erleichtern und die Beziehung der verschiedenen induktiven Leistungen zueinander verdeutlichen. Als elementare induktive Funktion kann man die *Kategorisierung* (Fried & Holyoak, 1984) eines beobachteten Exemplars B als zugehörig zu einer Kategorie oder Klasse C ansehen. Die Beobachtung besteht genauer betrachtet aus einer Reihe von Merkmalen  $m_1, m_2, \dots, m_K$  (oft auch als "Cues" bezeichnet) des beobachteten Exemplars (Ereignis, Person, Gruppe etc.). Das beobachtete Exemplar besitzt auf den Merkmalsdimensionen bestimmte Ausprägungen  $m_{B1}, m_{B2}, \dots, m_{BK}$ . Ein bekanntes oder angenommenes Gesetz  $G(C; m_1, m_2, \dots, m_K)$  dient als Grundlage, um die *Wahrscheinlichkeit*  $p(B \in C)$  zu schätzen, dass das Exemplar B in Anbetracht seiner Merkmale  $m_{B1}, m_{B2}, \dots, m_{BK}$  der Kategorie C zugehört. Beispielsweise wird ein Patient P mit einem Symptommuster  $m_{P1}$ ,

$m_{P_2}, \dots, m_{P_K}$  mit einer bestimmten Wahrscheinlichkeit in eine Diagnose-Klasse D

kategorisiert, auf der Grundlage einer medizinischen Gesetzmäßigkeit  $G(D; m_1, m_2, \dots, m_K)$ .

Die nächsthöhere Funktion in der Hierarchie induktiver Leistungen ist *Diskrimination*; sie setzt Kategorisierung als begrifflichen Baustein voraus. Diskriminieren verlangt einen Vergleich von zwei (oder mehr) Kategorisierungen. Entweder gegeben sind zwei Beobachtungen  $B_1$  und  $B_2$ , und es gilt zu entscheiden, ob beide derselben Äquivalenzklasse C angehören oder ob sie verschieden sind. Oder aber es liegt eine einzelne Beobachtung B vor, und es muss entschieden werden, ob B eher in eine fokussierte  $C_1$  als in alternative Kategorien  $C_0$  gehört. Das heißt, die *Wahrscheinlichkeit*  $p(B \in C_1)$  muss verglichen werden mit  $p(B \in C_0)$ . Beispielsweise könnte diskriminiert werden, welche von zwei Diagnosen eher auf einen Patienten mit einem bestimmten Merkmalsmuster zutrifft.

So wie Diskriminierung eine Relation oder ein Wettbewerb zwischen Kategorisierungen ist, kann man eine *Kontingenz* verstehen als Vergleich zwischen Diskriminierungen. Eine Kontingenz zwischen einem Kriterium C und einem Prädiktor P besteht dann, wenn sich eine Kategorisierung bzw. Diskriminierung von Beobachtungen hinsichtlich einer fokussierten Klasse mit den Prädiktorklassen ändert, das heißt wenn  $p(B \in C_1 / P_1)$  verschieden ist von  $p(B \in C_1 / P_2)$  oder anders ausgedrückt, wenn  $\Delta = p(B \in C_1 / P_1) - p(B \in C_1 / P_2) \neq 0$ .  $\Delta$  ist das am häufigsten verwendete statistische Kontingenzmaß. Eine Kontingenz könnte etwa bestehen zwischen einer Krankheit und dem Prädiktor Ernährung, so dass ein größerer Anteil von Menschen mit einer bestimmten Ernährung  $P_1$  unter die Krankheit fällt als mit einer anderen Ernährung  $P_2$ .

Im wesentlichen beschränkt sich das vorliegende Kapitel auf psychologische Forschung über mentale Operationen beim Kategorisieren, Diskriminieren und Erkennen von Kontingenzen, die in der Statistik den Operationen zur Berechnung von Mittelwerten, Varianzen und Korrelationen ähneln. Abschließend wird auf der komplexesten Stufe der begrifflichen Systematik die kognitive Verarbeitung solcher Probleme dargestellt, bei denen

zwei oder mehr Kontingenzen um die Erklärung gegebener Daten wetteifern. Ein solches Problem liegt etwa dann vor, wenn geklärt werden muss, welche von mehreren Kontingenzen das Auftreten einer Erkrankung eher erklärt (z.B. zwischen Krankheit und Ernährung, Krankheit und Auftreten eines Erregers, oder Krankheit und präventive Maßnahmen).

### **Experimentelle Methoden und Maße**

Um induktiv-statistisches Schließen und die dazu benötigten subjektiven Wahrscheinlichkeiten zu untersuchen, hat die experimentelle Psychologie eine Reihe von Methoden entwickelt. Die am wenigsten aufwendige und wohl auch am häufigsten verwendete Methode ist die *direkte Schätzung von Wahrscheinlichkeiten*, etwa auf einer numerischen Prozentskala oder aber auf einer grafischen Ratingskala. In vielen Untersuchungen wird den Teilnehmern in jedem Durchgang eine Kombination von Symptomen präsentiert, und ihre Aufgabe lautet, die Wahrscheinlichkeit bestimmter Diagnosen numerisch zu schätzen (Gigerenzer, Hoffrage & Kleinböling, 1991; Juslin, 1994; Koehler, White & Grondin, 2003; Tversky & Koehler, 1994).

Direkte Schätzungen dieser Art stellen hohe Anforderungen an die Versuchsteilnehmer, ihre subjektiven Wahrscheinlichkeiten introspektiv zu erkennen und auf eine numerische, verbale oder grafische Antwortskala abzubilden. Eine vorsichtiger Methode besteht daher in qualitativen Urteilen darüber, ob eine bestimmte Diagnose oder Kategorisierung auf eine Liste von Symptomen oder Merkmalen passt oder welche von zwei im Paarvergleich dargebotenen Kategorien eher zutrifft (z.B. Liegt bei solchen Symptomen eine Neurose oder eine Psychose vor? Vgl. Gigerenzer et al., 1991; Goldberg, 1968). Wahrscheinlichkeiten lassen sich dann schätzen über die relative Häufigkeit von Antworten, die eine bestimmte Kategorisierung unterstützen.

Beide Verfahren, qualitative Entscheidung und quantitative Schätzung lassen sich auch kombinieren, indem Urteiler zunächst angeben, ob eine bestimmte Kategorisierung zutrifft, und sodann in einem zweiten Schritt ihre *subjektive Sicherheit* bei dieser Entscheidung

einschätzen. Es ist hierbei wichtig, im Auge zu behalten, dass eine solche Angabe der Wahrscheinlichkeit als subjektive Sicherheit oder Glaubensgrad bei einer einzelnen Entscheidung nicht gleichzusetzen ist mit einer frequentistische Wahrscheinlichkeit (Gigerenzer & Murray, 1987), verstanden als relative Häufigkeit (Wieviel Prozent aller Patienten mit einer solchen Symptomatik haben die Krankheit?).

Anstelle von direkten Urteilen kann man Wahrscheinlichkeiten auch indirekt skalieren, etwa über den *Preis*, der für eine Lotterie mit einer bestimmten Gewinnwahrscheinlichkeit bezahlt wird (Becker, Degroot & Maschak, 1963; Stewart, Chater, Stott & Reimers, 2003; Slovic, 1995) oder über die Messung der *Reaktionszeit oder Latenz*, die jemand benötigt, um die Überlegenheit der einen über die andere Kategorisierung zu erkennen. Eine andere indirekte Methode besteht darin, die *Lerngeschwindigkeit* zu erfassen, mit der jemand eine Kategorisierung, Diskriminierung oder Kontingenz trainieren kann (Hirsch, Hammond & Hirsch, 1964; Slovic & Lichtenstein, 1971). Je schneller man lernt, dass ein bestimmtes Muster von Symptomen eine bestimmte Krankheit anzeigt, um so höher sollte die subjektive Wahrscheinlichkeit sein, bzw. die Bereitschaft, eine Kategorisierung zu akzeptieren. An dieser Stelle überschneidet sich die Forschung zum induktiven Schließen mit dem großen Gebiet des Lernens und Konditionierens (Alloy & Tabachnik, 1984; Cheng, 1997; Gluck & Bower, 1988).

Wesentlich ist auch die Unterscheidung verschiedener Modi der Darbietung von Stimulusinformation. In vielen Untersuchungen werden statistische Kennwerte einfach in Form von Tabellen oder Statistiken vorgegeben. Die Teilnehmer bekommen etwa mitgeteilt, dass die bedingte Wahrscheinlichkeit des Auftretens verschiedener Symptome, gegeben eine bestimmte Krankheit, bestimmten Prozentwerten entspricht, und müssen daraufhin schätzen, wie wahrscheinlich eine fragliche Krankheit angesichts mancher dieser Symptome ist. In anderen Untersuchungen bekommen die Teilnehmer die Wahrscheinlichkeiten oder Gesetze, die sie zum induktiven Schließen benötigen, nicht vorgegeben, sondern müssen sie über viele

Beobachtungen hinweg selbst herausfinden. Beispielsweise könnten sie über eine größere Stichprobe von Patienten hinweg beobachten, welche Symptome mit welcher Krankheit einhergehen, und die Erkenntnisse aus diesem langsamen Prozess des assoziativen Lernens bilden dann die Grundlage für induktive Schlüsse. Beide Verfahren – passive Rezeption von Statistiken versus aktive Beobachtung – führen zu durchaus unterschiedlichen Befunden (Hertwig, Barron, Weber & Erev, 2004; Weber, Shafir & Blais, 2004).

Damit sind die wichtigsten, aber beileibe nicht alle Methoden aufgezählt. Eine zunehmende Rolle spielen mathematische Modelle, die man durch geschicktes Experimentieren einsetzen kann, um kognitive Prozesse beim induktiven Schließen zu analysieren und zu quantifizieren. Zu diesen Modelle, deren ausführliche Behandlung den Rahmen dieses Kapitels sprengen würde, gehören vor allem die Signalentdeckungsanalyse (Swets, Dawes, & Monahan, 2000), multinomiale Prozessmodelle (Batchelder & Riefer, 1990; Meiser & Bröder, 2002) oder der Einsatz von Modellen der Computersimulation kognitiver Schlussprozesse (Dougerty, Gettys & Ogden, 1999; Fiedler, 1996; Hintzman, 1984; Roe, Busemeier & Townsend, 2001). Sofern im weiteren bei der Darstellung der Forschungsergebnisse keine besonderen Anmerkungen zur Methode erfolgen, beruhen die Befunde auf direkten Schätzungen oder Entscheidungen, der am häufigsten verwendeten Messmethode.

## PARADIGMATA DER EMPIRISCHEN FORSCHUNG

### **Subjektive Wahrscheinlichkeiten bei einfachen Kategorisierungen**

In diesem Abschnitt werden empirische Befunde und Theorien zum elementaren Paradigma dargestellt, Kategorisierungen unter Unsicherheit. Welches ist die Wahrscheinlichkeit, dass Menschen an einer bestimmten Todesursache sterben, dass in diesem Jahr eine Flutkatastrophe eintritt, dass jemand im Lotto gewinnt oder dass ein Verdächtiger in einem Strafprozess der wirkliche Täter ist? Viele alltägliche und existentielle

Entscheidungen (Gesundheit, Geldanlage, Kaufentscheidungen etc.) verlangen induktive Schlüsse dieser Art. Ein zentraler und vielfach bestätigter Befund besagt, dass subjektive Wahrscheinlichkeiten meist nicht genau kalibriert sind. Das bedeutet, subjektive und objektive Wahrscheinlichkeiten weichen oft beträchtlich voneinander ab. Zur Erklärung dieser Fehler und Verzerrungen subjektiver Urteile gibt es verschiedenartige Gründe. Es gibt aber auch einen plausiblen gemeinsamen theoretischen Nenner, der darin besteht, dass die Erfassung der Wahrscheinlichkeiten, mit denen vielfältige Ereignisse in einer unüberschaubar komplexen Umwelt auftreten, nur unter großer Unsicherheit möglich ist. Zum einen entziehen sich viele Ereignisse und Sachverhalte der direkten Beobachtung, oder es sind nur begrenzte Stichproben verfügbar, um latente Wahrscheinlichkeiten zu schätzen. Zum anderen ist die Kapazität der menschlichen Wahrnehmung und des Gedächtnisses beschränkt, so dass nicht einmal alle verfügbaren Daten vollständig und fehlerfrei genutzt werden können (Kahneman, Slovic & Tversky, 1982; Simon, 1990). Daher sind subjektiv geschätzte Wahrscheinlichkeiten auf vereinfachende Heuristiken oder "Daumenregeln" angewiesen (Gigerenzer, Todd & the ABC Group, 1999; Kahneman, Slovic & Tversky, 1982; Nisbett, Krantz, Jepson & Kunda, 1983), die nur eine suboptimale, näherungsweise Schätzung der objektiven Wahrscheinlichkeiten erlauben und unter bestimmten Bedingungen zu erheblichen Fehlschätzungen führen können.

### *Theoretische Fragen*

*Subjektive versus objektive Wahrscheinlichkeiten.* Eine grundlegende Frage, was Inferenzschlüsse über einzelne Ereignisklassen angeht, lautet: Wie weichen subjektive von objektiven Wahrscheinlichkeiten ab? Sind die objektiven Wahrscheinlichkeiten für breite Klasse von sehr seltenen bis sehr häufigen Ereignissen bekannt, dann kann man subjektive Wahrscheinlichkeiten als Funktion dieser objektiven Größen darstellen. Abbildung 1 zeigt eine solche Funktion in ihrer idealtypischen Form, wie sie in der momentan einflussreichsten ökonomischen Entscheidungstheorie angenommen wird, der prospect theory (Kahneman &

Tversky, 1979; Tversky & Kahneman, 1981). In dieser Funktion werden mehrere Gründe für die Abweichung subjektiver von objektiven Wahrscheinlichkeiten kombiniert. Zunächst sieht man im unteren Bereich der Kurve, dass objektiv seltene Ereignisse subjektiv oft überschätzt werden. Diese Tendenz manifestiert sich etwa darin, dass viele Menschen Geld und große Hoffnung in Glücksspiele setzen und dass viele seltene Restrisiken oft unverhältnismäßig gefürchtet werden (z.B., BSE, Terrorismus). Ereignisse von mittlerer Auftretenshäufigkeit werden dagegen im allgemeinen leicht unterschätzt. Dementsprechend verläuft die Kurve im mittleren Bereich unterhalb der Diagonalen, die das Maß für perfekte Kalibrierung anzeigt (d.h. wo subjektive und objektive Wahrscheinlichkeiten übereinstimmen). Der Grund hierfür könnte darin liegen, dass die Wahrscheinlichkeiten vieler alltäglicher Ereignisse enumerativ geschätzt werden, also durch mentales Aufzählen von Beispielen (Brown, 1995), und dieser Prozess ist meist unvollständig. Es fallen einem in der Regel weniger Beispiele ein als sie objektiv existieren. Erst im obersten Bereich gegen 100% begegnet die Kurve wieder der Diagonalen; objektive Sicherheit (100%) ist also auch subjektiv perfekte Sicherheit.

Wie Hertwig et al. (2004) jüngst gezeigt haben, hängt die Überschätzung seltener Ereignisse vom Darbietungsmodus der Information ab. Nur wenn Wahrscheinlichkeiten verbal benannt ("in der Lotterie zu gewinnen, ist extrem selten") oder numerisch quantifiziert werden ("die Gewinnwahrscheinlichkeit beträgt  $10^{-4}$ "), wird die wahre Seltenheit nicht deutlich genug gesehen. Verhaltensmaße (z.B. der Kauf einer Lotterie) messen solchen seltenen Kategorien zu viel Bedeutung bei. Müssen Urteiler indessen selbst durch eigene Beobachtung die Seltenheit dieser Ereignisse herausfinden und erleben – und somit sehr lange auf das Auftreten seltener Ereignisse in einer Stimulusserie warten – dann werden seltene Ereignisse unterschätzt; sie werden leicht übersehen und für das Verhalten nicht mehr berücksichtigt.

Unabhängig von der *absoluten* Über- oder Untergewichtung seltener Ereignisse gilt für die *relative* Beurteilung seltener und häufiger Ereignisse grundsätzlich, dass subjektive

Wahrscheinlichkeiten regressiv sind. Dieses zentrale Merkmal der Regressivität besagt, dass die Steigung der Funktion kleiner als 1 ist. Wahrscheinlichkeiten werden um so mehr unterschätzt, je größer sie sind, und sie werden um so mehr überschätzt, je geringer sie sind (Fiedler, 2002; Fiedler & Armbruster, 1994; Sedlmeier, 1999, 2002). Regressivität ist eine notwendige Konsequenz der Tatsache, dass subjektive und objektive Werte nicht perfekt korrelieren (Furby, 1973) oder anders ausgedrückt, dass der induktive Lernprozess des Erfassens von Wahrscheinlichkeiten fehlerbehaftet und unvollständig ist. Zu Beginn des Lernprozesses, bevor irgendetwas über die Auftretensrate verschiedener Ereignisklassen bekannt ist, muss die subjektive Funktion völlig flach und waagrecht verlaufen; alle Klassen erscheinen subjektiv gleich wahrscheinlich, da ja noch keinerlei Wissen existiert. Dann, wenn der Lernprozess fortschreitet und die unterschiedliche Auftretensrate bestimmter Ereignisklassen mehr und mehr sichtbar wird, wird die Kurve allmählich steiler. Aber solange Lernen in einer unsicheren Welt immer unvollständig und fehlerbehaftet ist, bleibt die Kurve auch nach einem längeren Lernprozess noch regressiv. Dies ist eine der häufigsten Fehlerquellen bei induktiven Schlüssen (Erev, Wallsten & Budescu, 1994; Fiedler, 1991); kleine Wahrscheinlichkeiten werden überschätzt und große werden relativ unterschätzt. Und diese Fehlertendenzen nehmen mit wachsender Extremität zu.

Dass Wahrscheinlichkeitsurteile und –inferenzen regressiv sind, sagt auch das Range-Frequency Modell (Parducci, 1965; Parducci & Wedell, 1986) korrekt vorher. Danach versuchen Urteiler, wenn sie Stimuli nach  $k$  aufsteigenden Kategorien klassifizieren und quantifizieren sollen, einen Kompromiss zwischen zwei Tendenzen zu finden: Nach der Range-Tendenz versuchen Urteiler, bei der Kategorisierung und Repräsentation von Stimuli im Gedächtnis die Kategoriengrenzen so zu legen, dass alle  $k$  Kategorien gleich groß sind. Nach dem Range-Prinzip versuchen sie, die Stimuli so zu repräsentieren, dass die Breite bzw. der Abstand aller Kategorien möglichst gleich ist. Nach dem Frequency-Prinzip versuchen sie zu erreichen, dass auf jede Kategorie gleich viele Stimuli oder Exemplare entfallen. Aus

beiden Tendenzen muss der beste Kompromiss gefunden werden. Hieraus folgt, dass wenn manche Kategorien tatsächlich wahrscheinlicher bzw. häufiger sind als andere, die Frequency-Tendenz dazu beiträgt, diese Unterschiede zu verringern. Dies bedeutet Regression: Unterschätzung von häufigen und Überschätzung von seltenen Kategorien.

Regressive Urteilstendenzen entstehen allein durch Rauschen oder unsystematische Fehler. Daneben gibt es jedoch auch systematische Einflüsse auf subjektive Wahrscheinlichkeiten, vor allem solche, die mit der selektiven Zugänglichkeit (Accessibility) und Erinnerbarkeit von Informationen zusammen hängen. Eine häufig zitierte Untersuchung von Combs und Slovic (1979) betrifft subjektive Schlüsse auf die relative Häufigkeit von insgesamt 41 Todesursachen. Die Urteile spiegelten erstens die Reihenfolge der Prävalenz recht genau wider; sie sind also durchaus sensibel für die objektive Wirklichkeit. Zweitens weisen sie die üblichen Regressionseffekte auf; seltene Risiken (z.B. Blitzschlag) werden zu hoch eingeschätzt, während die häufigsten (Herz- und Kreislauferkrankungen) unterschätzt werden. Aber drittens spiegeln die Daten die selektive Verfügbarkeit der Information wider. Im allgemeinen wurden diejenigen Risiken relativ überschätzt, über die die Medien am meisten und lebhaftesten berichten (z.B. Katastrophen, Gewaltverbrechen). Jene Todesarten, die selten öffentlich werden und daher weniger verfügbar sind, wurden indessen relativ unterschätzt (vgl. Combs & Slovic, 1979). Subjektive Urteile sind also systematisch in Richtung auf diejenigen Ereignisklassen verzerrt, die im Sinne der Verfügbarkeits-Heuristik (Tversky & Kahneman, 1973) eine höhere Chance haben, in die urteilsleitende Gedächtnisstichprobe aufgenommen zu werden.

Eine theoretische Frage, die in den letzten Jahren viel Aufmerksamkeit gefunden hat, betrifft den genauen kognitiven Prozess, der der Verfügbarkeits-Heuristik zugrunde liegt. Werden Urteile in Richtung auf diejenigen Stimulus-Klassen verzerrt, deren Erinnerung (retrieval) am leichtesten fällt, oder deren Erinnerung die meisten Beispiele hervorbringt (Schwarz et al., 1991). Wird etwa das Risiko einer bestimmten Krankheit dann am größten

wahrgenommen, wenn die Erinnerung eines Beispiels für die Krankheit besonders leicht fällt oder wenn die Anzahl der erinnerbaren Beispiele besonders groß ist? Die häufig zitierten Befunde von Schwarz et al. (1991) sprachen zunächst eindeutig für das Prinzip der Leichtigkeit, aber in der Folgezeit ergaben viele Experimente, dass je nach Randbedingung sowohl die Anzahl als auch die Leichtigkeit erinnerbarer Beispiele für subjektive Wahrscheinlichkeiten ausschlaggebend sein können (Rothman & Schwarz, 1998).

Die Gründe für selektive Verfügbarkeit und Zugänglichkeit sind vielfältig. Ob urteilsrelevante Beobachtungen beachtet oder ignoriert, stark oder schwach gewichtet werden und ob sie somit in die Stichprobe von Informationen mit eingehen, auf denen induktive Schlüsse beruhen, hängt von motivationalen Faktoren ab (z.B. Verdrängen unangenehmer Erinnerungen; Christianson, 1992; Singer 1990), vom Selbstbezug (Ross & Sicoly, 1979) oder von sogenannten Entpackungs-Effekten (Tversky & Koehler, 1994) oder Kategorien-Spaltungs-Effekten (Betsch et al., 1999; Fiedler & Armbruster, 1994). Demnach kann man eine Kategorie dadurch stärker kognitiv zugänglich machen und sie damit insgesamt wahrscheinlicher erscheinen lassen, dass man sie in verschiedene Teilaspekte zerlegt. So erscheint die Wahrscheinlichkeit einer Krankheit angesichts bestimmter Symptome höher, wenn die Kategorie in ihre Komponenten "entpackt" wird, indem man für jedes Symptom separat einen Schluss auf die Krankheit erfragt, statt ein einziges Urteil für das gesamte Symptommuster auf einmal zu erfragen (Koehler et al., 2003). Durch das Aufspalten in viele einzelne Symptome erscheint die gesamte Kategorie psychologisch mächtiger. Um das zu erklären, benutzen Tversky und Koehler (1994) in ihrer Support Theory den Begriff der Subadditivität. Damit soll ausgedrückt werden, dass der Support für eine gesamte Kategorie geringer ist als die Summe des Supports, der sich aus den einzelnen Aspekten (z.B. aufgrund jedes einzelnen Symptoms) ergibt.

Davon abgesehen kann man solche subjektiven Vergrößerungen von Kategorien ("Zooming Effekte", Fiedler 2002) auch einfach durch Regression erklären. Durch Aufspalten

einer Kategorie entstehen mehrere kleine Kategorien, die im Sinne der normalen Regression überschätzt werden, so dass die Summe der Teile größer wird als das Ganze, welches durch Regression eher reduziert wird.

Eine letzte theoretische Frage, die hier erwähnt werden soll, und ein weiterer Grund, warum subjektiv gelernte Wahrscheinlichkeiten häufig nicht den objektiven Verhältnissen entsprechen, ergibt sich aus der sequentiellen Natur des Lernprozesses. Über die Zeit hinweg müssen immer wieder neue Beobachtungen berücksichtigt werden, so dass sich die Wahrscheinlichkeiten ständig im Lichte neu hinzukommender Information verändern. Dies erfordert einen kognitiven Prozess der Anpassung (Updating). Beispielsweise muss die Wahrscheinlichkeit des Vorliegens einer Krankheit korrigiert werden, wenn zu den bisher beobachteten Symptomen ein oder mehrere neue Symptome hinzukommen. Die normative Regel für diese Korrektur liefert das Bayes-Theorem:

$$\Omega_{\text{posterior}} = L \cdot \Omega_{\text{prior}},$$

wobei  $\Omega_{\text{posterior}}$  die a-posteriori "Odds" sind, das heißt, die Wahrscheinlichkeit, dass die Krankheit vorliegt, geteilt durch die Wahrscheinlichkeit, dass die Krankheit nicht vorliegt, nachdem das neue Symptom bekannt geworden sind.  $\Omega_{\text{prior}}$  sind die entsprechenden a-priori "Odds" vor Bekanntwerden der neuen Information, und L ist der sogenannte Likelihood-Quotient für das neue Symptom,  $p(\text{Symptom} / \text{Krankheit gegeben}) / p(\text{Symptom} / \text{Krankheit nicht gegeben})$ .

Gemessen an dieser normativen Vorschrift verhalten sich subjektive Anpassungen von Wahrscheinlichkeiten häufig *konservativ*. Das heißt, die notwendigen Anpassungen sind nicht deutlich genug. Oder anders ausgedrückt, der Likelihood-Quotient L, der den Einfluss der Korrektur im Lichte der neuen Information beschreibt, bekommt nicht genügend Gewicht; die neue Wahrscheinlichkeit  $\Omega_{\text{posterior}}$  wird zu stark von der alten  $\Omega_{\text{prior}}$  bestimmt. Mit anderen Worten, der Lerneffekt durch die zuletzt erhaltene Information bleibt zu gering.

Im Gegensatz zu diesem Phänomen der Konservativität (d.h., zu starke Gewichtung von  $\Omega_{\text{prior}}$  und zu schwache Gewichtung von  $L$ ), gibt es auch zahlreiche Belege für die Vernachlässigung von Basisraten ("baserate neglect"; d.h. Ignorieren oder zu schwache Gewichtung von  $\Omega_{\text{prior}}$  und somit Übergewichtung von  $L$ ). Wird beispielsweise ein positiver HIV-Testbefund beobachtet, ist die subjektive Wahrscheinlichkeit, dass AIDS vorliegt, sehr hoch, denn der Likelihood-Ratio  $L$  ist sehr hoch (d.h.  $p(\text{HIV+} / \text{AIDS liegt vor})$  ist viel höher als  $p(\text{HIV+} / \text{AIDS liegt nicht vor})$ ). Dass  $\Omega_{\text{prior}}$  hingegen ein sehr kleiner Quotient ist, und zwar noch extremer als  $L$  groß ist, weil  $p(\text{AIDS liegt vor})$  um mehrere Zehnerpotenzen größer ist als  $p(\text{AIDS liegt nicht vor})$ , wird leicht ignoriert (vgl. Gigerenzer & Hoffrage, 1995; Swets et al., 2000). Die Folge ist, dass die Wahrscheinlichkeit von AIDS gegeben ein positiver HIV Test (HIV+) wegen des hohen  $L$  und der gleichzeitigen Vernachlässigung des der überaus geringen Basisraten meist deutlich überschätzt wird (Eddy, 1982; Fiedler, Brinkmann, Betsch & Wild, 2000; Gigerenzer & Hoffrage, 1995).

Theoretisch stellt sich zwangsläufig die Frage, unter welchen Bedingungen die beiden gegensätzlichen Befunde – Konservativismus oder Baserate Neglect – auftreten. Die bisherige Forschung gibt zwar keine erschöpfende Antwort, aber eine Reihe von Moderator-Variablen sind wohl bekannt. Basisraten werden dann eher beachtet und stark gewichtet, wenn sie sequentiell beobachtet oder erlebt werden können und nicht durch abstrakte Statistiken oder numerische Quantoren mitgeteilt werden (Hertwig et al., 2004). Eine wichtige Rolle kann auch Kontextwissen spielen; beispielsweise werden Basisraten, die durch technisch statistische Zahlen ausgedrückt werden (Schwarz, Strack, Hilton & Naderer, 1991), im Kontext eines computerbasierten Problems stärker berücksichtigt als im Kontext eines individuellen Urteilsproblems.

Eine weitere moderierende Bedingung wurde von Ayton und Fischer (in press) aufgezeigt, mit Bezug auf zwei bekannte Alltagsbeispiele. Als "Hot-Hand" Phänomen bezeichnet man die Erwartung, dass etwa ein Basketball-Spieler, der gerade "heiß" ist oder

"einen Lauf hat", auch bei weiteren Korbwürfen Erfolg haben wird. Dies impliziert die Annahme, dass eine hohe Basisrate unverändert bleibt und entspricht somit dem Konservativismus. Es gibt aber in Sport und Spiel auch das umgekehrte Phänomen, "gambler's fallacy" genannt, also den Glauben, dass man nach mehreren Erfolgen im Glücksspiel nun geringere Chancen besitzt. Wie Ayton und Fischer (in press) gezeigt haben, entsteht die "Hot-Hand"-Illusion eher, wenn Ereignisse willentlich und durch Geschick kontrollierbar erscheinen, während Gamber's Fallacy in unkontrollierbaren, rein stochastisch erscheinenden Situationen entsteht.

### *Empirische Phänomene*

Um eine Übersicht der empirischen Forschung über induktive Schlüsse vom Typ der einfachen Kategorisierung zu gewinnen, sollen die verwendeten experimentellen Aufgaben nach folgenden Gesichtspunkten unterschieden werden: (a) Hinsichtlich der Validität und Rationalität der Inferenzschlüsse kann man unterscheiden zwischen Fällen, in denen zur Kategorisierung von Objekten tatsächlich die definierenden, ökologisch validen Merkmale der Kategorien herangezogen werden oder aber nicht-definierende, mit der Kategorie nur heuristisch verbundene Merkmale. (b) Hinsichtlich der Zahl der Merkmale oder – um das häufig verwendete englische Wort einzuführen – Cues, die zur Kategorisierung herangezogen werden, kann man Kategorisierungen aufgrund einzelner Cues von Kategorisierungen aufgrund multipler Cues abgrenzen, da letztere andere kognitive Prozesse erfordern. (c) Im Falle multipler Cues oder Merkmale macht es einen Unterschied, ob eine additive Kombinationsregel verlangt wird oder aber Wettbewerb und Wechselwirkungen zwischen den Stimulus-Merkmalen bestehen. (d) Ferner ist es psychologisch sinnvoll, zwischen selbst-generierten und vorgegebenen Stimulus-Merkmalen zu unterscheiden. (e) Und hinsichtlich des Skalenniveaus kann man dichotome und qualitative Merkmale von kontinuierlichen und quantitativen Merkmalen unterscheiden. Dies gilt auch für die abhängige Variable der Kategorisierungsaufgabe, die entweder eine einfache dichotome Entscheidung verlangt (liegt

eine Krankheit vor oder nicht) oder aber eine quantitative Ausprägung bzw. ein Urteil über die subjektive Sicherheit der Zugehörigkeit eines Stimulus zu einer Kategorie (Wie groß ist die subjektive Sicherheit, dass eine bestimmte Krankheit vorliegt?).

*Kategorisierungen aufgrund einzelner Merkmale.* Kategorisierungsschlüsse aufgrund einzelner, eindeutig definierender Merkmale werden gewöhnlich als Syllogismen bezeichnet und gehören in die Domäne des deduktiven Schließens. Sie werden daher in Kapitel xx behandelt. Induktiv-statistisches Schließen basiert indessen immer auf Merkmalen von begrenzter Validität, die keine eindeutigen Schlüsse zulassen, sondern nur Wahrscheinlichkeitsschlüsse, die von Natur aus fehlerbehaftet sind. So sind Krankheitsdiagnosen, juristische Schuldsprüche, Identifikationen von Gesichtern oder Interpretationen von Äußerungen als Lüge stets mit Unsicherheit behaftet. Selbst dann, wenn etwa für die psychologisch und biologisch ungemein wichtige Entscheidung, ob eine Kommunikation als Lüge oder Täuschung zu klassifizieren ist, die besten verfügbaren Cues herangezogen werden (z.B., Stimmhöhe, Pupillengröße oder Sprechverzögerungen; vgl. Zuckerman, DePaulo & Rosenthal, 1981), bleibt dennoch eine große Ungewissheit bestehen. Um damit umzugehen, ist es auf der meta-kognitiven Ebene dringend erforderlich zu erkennen, wie groß die Unsicherheit bei verschiedenen Schlussfolgerungen ist, um das Ausmaß der Konsequenzen von falschen Entscheidungen und Fehlhandlungen kontrollieren zu können. Angesprochen ist hiermit das Problem der Kalibrierung und der Überschätzung der subjektiven Sicherheit – genannt Overconfidence.

*Kalibrierung der subjektiven Sicherheit von Inferenzschlüssen.* Eine Vielzahl von empirischen Studien befasst sich mit der Frage, ob die subjektiv erlebte Sicherheit bei induktiven Schlüssen in einem Verhältnis zu der tatsächlichen Genauigkeit oder der Trefferrate steht, mit anderen Worten, ob subjektive Wahrscheinlichkeiten genau kalibriert (d.h., geeicht) sind. Eine typische Versuchsanordnung (vgl. Juslin, Winman & Olsson, 2000) sieht so aus, dass zunächst ein Gegenstandsbereich ausgesucht wird, für den die korrekten

Antworten für eine Vielzahl von Kategorisierungsaufgaben bekannt sind. Versuchsteilnehmer werden sodann über viele Durchgänge hinweg gebeten zu entscheiden, ob ein Stimulusobjekt in eine bestimmte Kategorie fällt oder nicht, und sie werden außerdem gefragt, ihre subjektive Sicherheit anzugeben, dass ihre Entscheidung korrekt war. Über alle Durchgänge hinweg kann man dann berechnen, wieviel Prozent aller Antworten korrekt waren, wenn die subjektive Sicherheit 60%, 70%, 80%, 90% oder 100% betrug. In dem Maße, wie der Anteil korrekter Antworten (ebenfalls in % gemessen) dem Ausmaß der subjektiven Sicherheit entspricht, gelten die Urteile als wohl kalibriert. Perfekte Kalibrierung ist jedoch in den meisten Studien nicht gegeben. Ein typischer Befund ist indessen Overconfidence: Die tatsächliche Rate richtiger Kategorisierungen ist niedriger als die subjektive Sicherheit, und diese Diskrepanz wird mit zunehmender Sicherheit immer größer (Lichtenstein, Fischhoff & Phillips, 1982). Es liegt auf der Hand, dass Fehlschlüsse im Sinne von Overconfidence großen ökonomischen, politischen und persönlichen Schaden verursachen können. Abgesehen von seiner enormen praktischen Bedeutung ist dieses Phänomen ebenfalls zentral für die theoretische Aufklärung der beim induktiven Schließen beteiligten kognitiven Prozesse. In jedem Falle ist daher die präzise Untersuchung des Ausmaßes von Overconfidence und seinen Grenzen von großer Bedeutung.

Eine typische Kalibrierungskurve ist in Abbildung 2 dargestellt, übernommen von Juslin et al. (2000). Wie man sieht, nimmt die Überschätzung der subjektiven Sicherheit von unsicheren zu sicheren Entscheidungen regelmäßig zu und erreicht im oberen Bereich ein beträchtliches Ausmaß. Ähnlich starke Befunde wurden auf den verschiedensten Gebieten der ökonomischen, medizinischen, sozialen und politischen Entscheidung erzielt (vgl. Meta-Analyse bei Juslin et al., 2000). In diesen Studien wurden starke und stabile Überschätzungen der subjektiven Wahrscheinlichkeit von Kategorisierungen beobachtet, was Lichtenstein, Fischhoff und Phillips (1982) zu der – ebenfalls induktiven – Schlussfolgerung veranlasste, dass „people were consistently and strongly biased toward overconfidence.“

Obwohl Overconfidence Jahrzehnte lang den Status eines gut bestätigten Phänomens inne hatte und zu dem festen Textbuch-Stoff gehörte, wurde Anfang der neunziger Jahre eine lebhaftige Kontroverse über die Gültigkeit und Grenzen dieses Phänomens entfacht. Kritik an der Interpretation vieler Overconfidence-Befunde entstand aus zweierlei Quellen. Zum einen wiesen Erev, Wallsten und Budescu (1994) darauf hin, dass die übliche Praxis, die Genauigkeitsrate als Funktion der subjektiven Sicherheit darzustellen, und nicht umgekehrt, die Gefahr eines Regressions-Artefaktes in sich birgt. Da die beiden Variablen, Genauigkeit und subjektive Sicherheit, nicht perfekt korrelieren, muss die mittlere Genauigkeit, wenn man sie konditional zu bestimmten Stufen der Sicherheit betrachtet, bei zunehmender Extremität der Sicherheit immer mehr abweichen. In der Tat konnten Erev et al. (1994) zeigen, dass Overconfidence in Underconfidence überführt werden kann, wenn man dieselben Daten umgekehrt analysiert, das heißt, die subjektive Sicherheit konditional zu verschiedenen Stufen der Genauigkeit darstellt. Wegen der Regressivität jeder unvollkommenen Korrelation zeigt sich dann zwangsläufig, dass die subjektive Sicherheit derjenigen Urteile, die mit extremer Genauigkeit (z.B. 90% oder 100%) abgegeben werden, im Mittel geringer ausfällt, also eine Unterschätzung der wirklichen Genauigkeit darstellt.

Diese Re-Analyse von Overconfidence im Sinne der statistischen Regression zur Mitte impliziert auch das inzwischen bekannte Phänomen des Hard-Easy Effektes (Juslin et al., 2000). Hiermit ist der empirisch leicht zu bestätigende Sachverhalt gemeint, dass schwierige Inferenzschlüsse mit einer geringen objektiven Trefferrate im Erwartungswert mehr Overconfidence erzeugen als leichte Inferenzschlüsse mit einer hohen objektiven Trefferrate. Dies ergibt sich aus der Tatsache, dass objektiv hohe Prozentraten schwerer zu überschätzen sind als objektiv niedrige Prozentraten.

Weder der Hard-Easy-Effekt noch das Regressionsartefakt können jedoch das gesamte Phänomen der Overconfidence in Frage stellen. Denn die Überschätzung der subjektiven Sicherheit zeigt sich ebenfalls – und manchmal ganz besonders deutlich – unter solchen

Bedingungen, die nicht ohne weiteres als Hard-Easy Effekt oder durch Regression alternativ erklärt werden können. Besonders markant ist die Illusion im Paradigma der Intervall-Produktion (Soll & Klayman, in press). In diesem experimentellen Paradigma werden Urteiler gebeten, quantitative Schätzungen abzugeben, beispielsweise die Einwohnerzahl von Städten zu schätzen. Die Instruktion lautet, die obere und untere Grenze einer Kategorie so zu wählen, dass die richtige Antwort (z.B. die tatsächliche Einwohnerzahl) mit einer bestimmten Wahrscheinlichkeit von 60%, 70%, 80% oder 90% innerhalb dieser Grenzen liegt. Wie sich herausstellt, ist die tatsächliche Wahrscheinlichkeit, dass die korrekte Antwort innerhalb der selbst bestimmten Kategorien liegt, systematisch geringer als die gegebene subjektive Sicherheit, das heißt, die selbst hergestellten Intervalle sind zu schmal, um die gewählte Wahrscheinlichkeit abzudecken.

Die andere Quelle von Kontroversen zum Overconfidence-Phänomen liegt in Brunswik's (1956) Idee repräsentativer Stichproben begründet. Diese Kritik setzt bei der Beobachtung an, dass praktisch alle traditionellen Kalibrierungsstudien nicht zufällig oder repräsentativ aus einem Inhaltsbereich ausgewählte Urteilsaufgaben verwenden, sondern selektiv ausgewählte Aufgaben, die – entsprechend der Intuition des Untersuchers – das vorgesagte Phänomen der Overconfidence tatsächlich wie gewünscht widerspiegeln. Verschiedene Forscher haben unabhängig voneinander aufgezeigt, dass Overconfidence immer dann deutlich auftritt, wenn die Aufgaben willkürlich oder selektiv ausgewählt werden, aber drastisch reduziert oder gar eliminiert wird, wenn die Urteilsaufgaben repräsentativ ausgewählt werden.

Auch diese Kritik kann jedoch nicht das gesamte Phänomen in Frage stellen, insbesondere nicht den zuvor erwähnten Befund mit selbst produzierten Intervallen, der auch im repräsentativen Design Bestand hat (Klayman, et al., 1999). Im übrigen haben auch nicht-repräsentativer Designs ihre Aussagekraft für die adaptive Intelligenz. Vertreter des repräsentativen Designs argumentieren, dass die subjektive Sicherheit dann wohl kalibriert ist,

wenn Urteile in demjenigen Bereich getestet werden, in dem sie gelernt wurden. Man könnte jedoch auch argumentieren, dass ein wichtiger Aspekt des angepassten Verhaltens darin besteht, Vorhersagen in neuen Umgebungen abzugeben, für die man nicht kalibriert ist und die nicht repräsentativ für die frühere Lernumgebung sind. Mit anderen Worten, wenn Overconfidence in nicht-repräsentativen Designs besonders ausgeprägt ist, dann bedeutet dies, dass Urteiler und Entscheider nicht bemerken, wenn sich die Problemumgebung verändert hat, das heißt, wenn Probleme auf einmal schwieriger und unsicherer sind als in der natürlichen Umgebung, in der die Kalibrierung gelernt wurde.

*Kategorisierungen aufgrund heuristischer Cues.* Viele Kategorisierungsentscheidungen, besonders wenn sie schnell und anhand geringer Information getroffen werden müssen, können sich nicht auf valide Cues verlassen, die eine logische oder intrinsische Beziehung zu der Zielkategorie haben, sondern können lediglich heuristische Cues nutzen. Heuristisch genannt seien solche Cues, die durch eine kognitive Heuristik (Gilovich, Griffin & Kahneman, 2002; Kahneman et al., 1982) als Hinweisreiz oder „Proxy“ mit der Zielkategorie verbunden sind. Diese Definition heißt mitnichten, dass heuristische Hinweisreize keine Validität besitzen; im Gegenteil, sie haben ihre heuristische Funktion in den meisten Fällen gerade dadurch entwickelt, dass sie wenigstens unter günstigen Bedingungen funktionell sind und entweder richtige oder (im Falle motivationaler Heuristiken) erwünschte und angenehme Inferenzschlüsse hervorbringen.

Die Verfügbarkeits-Heuristik wurde von Tversky und Kahneman (1973) als eine „heuristic for judging frequencies and probabilities“ eingeführt und ist als solche von primärer Bedeutung für induktiv-statistisches Schließen. Über weite Bereiche der alltäglichen Erfahrung hinweg ist der Verfügbarkeits-Cue – also die leichte Verfügbarkeit einer Kategorisierung im Gedächtnis – vermutlich sehr nützlich und verlässlich. Je höher die tatsächliche Auftretensrate einer Kategorie in der Welt, umso größer ist vermutlich ihre Verfügbarkeit im Gedächtnis, denn das Gedächtnis nimmt normalerweise (d.h. unbeachtet

anderer Einflüsse) mit der Darbietungshäufigkeit zu. Sofern man diese Regel umkehren kann, müsste man von der Verfügbarkeit bzw. Gedächtnisstärke auch auf die objektive Häufigkeit von Reizklassen schließen können. In der Tat ist dies für zahlreiche Inhaltsbereiche zutreffend (Goldstein & Gigerenzer, 2002). Dennoch gibt es auch andere Faktoren, welche die Verfügbarkeit im Gedächtnis beeinflussen, die aber *nicht* mit der objektiven Häufigkeit zusammen hängen, wie zum Beispiel die affektive Bedeutsamkeit oder Salienz von Stimulus-Beobachtungen (Combs & Slovic, 1979), ihr Selbstbezug (Ross & Sicoly, 1979) ihre Erwünschtheit und hedonische Annehmlichkeit (Singer, 1990) oder die Kompatibilität von Beobachtungen mit dem Vorwissen. Aufgrund dieser Faktoren können Schlüsse von der Verfügbarkeit auf die Häufigkeit oder Wahrscheinlichkeit irreführend sein.

Außerdem gibt es Faktoren, die zuweilen verhindern, dass die objektiven Häufigkeiten von Ereignisklassen tatsächlich beobachtet werden können, etwa weil bestimmte Ereignisse geheim bleiben oder der Wahrnehmung nicht zugänglich sind (z.B. latente Krankheiten wie AIDS, die Dunkelziffer von Verbrechen). Es sind diese beiden Klassen von Faktoren, die dafür sorgen, dass der Verfügbarkeits-Cue, wenn er unkritisch (hinsichtlich solcher Randbedingungen) verwendet wird, oft zu irreführenden Urteilen führt.

Dies ist immer dann der Fall, wenn die Stimuli oder Aufgaben in Experimenten gezielt so ausgewählt werden, dass die tatsächliche Häufigkeit von Reizklassen von ihrer Vertrautheit oder Verfügbarkeit abweicht. Werden etwa in einem Experiment 19 berühmte Namen und 20 nicht berühmte Namen präsentiert, dann wird das Gefühl der Verfügbarkeit im Gedächtnis nicht allein von der Häufigkeit bestimmt, sondern auch von der Berühmtheit. Infolge davon wird die Zahl der berühmten Namen trotz ihrer zahlenmäßigen Unterlegenheit geringer geschätzt als die Zahl der nicht berühmten Namen (Jacoby, Kelley, Brown & Jaeschko, 1989). In ähnlicher Weise können Fehlschlüsse durch den Verfügbarkeits-Cue dadurch experimentell erzwungen werden, dass selbstbezogene Stimuli mit fremdbezogenen verglichen werden (Ross & Sicoly, 1979), dass semantisch zusammen passende Reizpaare

(Löwe – Tiger) mit unzusammenhängenden verglichen werden (Tversky & Kahneman, 1973) oder dass manche Reizklassen durch eine Priming-Manipulation stärker aktiviert werden als andere (Rottenstreich & Tversky, 1997). Da selbstbezogene, semantisch plausible und durch Priming aktivierte Information leichter verfügbar ist als fremdbezogene, semantisch sinnlose und nicht aktivierte, bei ansonsten gleicher Häufigkeit der Darbietung, führt unter solchen Bedingungen die Heuristik zu Illusionen. Derartige Demonstrationen von experimentell forcierten Illusionen zeugen nicht nur von der Möglichkeit, dass Verfügbarkeit mit Fehlereinflüssen konfundiert sein kann und somit zu Fehlern führt, sondern ebenso sehr von der Sensibilität dieses Cue und von der Fähigkeit der Urteiler, Unterschiede im Grad der Verfügbarkeit in Wahrscheinlichkeits- oder Häufigkeitsurteile zu übersetzen.

Werden indessen Aufgaben-Situationen aufgesucht, in denen die Verfügbarkeit nicht mit anderen Einflüssen konkurrieren muss, dann spiegeln die Schätzungen von Häufigkeiten und relativen Häufigkeiten verschiedener Ereignisklassen oft die tatsächlich beobachtete Häufigkeit erstaunlich genau wider (Fiedler et al., 2000; Fiedler, Walther, Freytag & Plessner, 2002; Winman & Juslin, in press). Einen Beleg hierfür stellt auch der im Abschnitt über Kalibrierung berichtete Befund dar, wonach Urteiler den Anteil ihrer richtigen Antworten sehr genau schätzen können, wenn die Urteile repräsentativ sind für diejenige Aufgabenklasse, in der sie ihre eigene Erfolgsrate beobachten konnten. Man könnte dies so interpretieren, dass die gelernte Verfügbarkeit des Erlebnisses, richtig geantwortet zu haben, unter solchen Bedingungen zu genauen Urteilen führt. Wird jedoch die Ereignisklasse gewechselt – so dass die Verfügbarkeit des Erfolgs in einem Bereich gelernt und in einem anderen oder schwereren Bereich getestet wird – so gelangt die Verfügbarkeits-Heuristik zu verzerrten Urteilen (Juslin, 1994; Juslin & Persson, 2002). Im Grunde drückt die so weit referierte Forschung über das Auftreten und die Höhe von Illusionen, über die Genauigkeit und Ungenauigkeit der Schlüsse aufgrund der Verfügbarkeits-Heuristik eine völlig plausible Einsicht aus: Immer dann wenn Maße der Gedächtnisstärke als valides Maß für die

Auftretensrate von Ereignissen gelten können, funktioniert die Heuristik gut. Wenn die Gedächtnisstärke indessen wegen unkontrollierter oder absichtlich erzeugter Störeinflüsse nicht als unverzerrtes Maß der Auftretenshäufigkeit gelten kann, dann kann die Heuristik keine verlässlichen Ergebnisse liefern.

Während sich die frühe Forschung vornehmlich mit der Genauigkeit und der Rationalität heuristischer Urteile befasste, wendet sich die neuere Forschung zunehmend einer präziseren Bestimmung von Verfügbarkeit und des zugrunde liegenden kognitiven Algorithmus zu. Unterschieden werden können zumindest drei Interpretationen von Verfügbarkeit: (1) als Gedächtnisstärke schlechthin, am zuverlässigsten messbar durch die Zahl der Items, die aus einer Klasse erinnert werden können (Combs & Slovic, 1979); (2) als erlebte oder attribuierte Leichtigkeit der Erinnerbarkeit (Schwarz & Bless, 1992), unabhängig von der tatsächlich erinnerbaren Anzahl; (3) oder als bewusste oder unbewusste Gedächtnisspur, die von wenigstens einer einzigen früheren Exposition eines Reizes herrührt, im Sinne einer Rekognitions-Heuristik (Goldstein & Gigerenzer, 2002).

Obwohl die ursprüngliche Definition des Verfügbarkeits-Cues durch Tversky und Kahneman sich auf die Leichtigkeit bezog („ease with which relevant examples come to mind“), wurde in den meisten empirischen Untersuchungen zunächst die Anzahl erinnerter Reize aus einer Klasse als Maß der Verfügbarkeit verwendet. Ein solches Maß stellt nicht nur einen plausiblen Index genereller Gedächtnisstärke dar, sondern hat auch den Vorteil, dass durch die Berücksichtigung vieler erinnerter Items eine erhöhte Reliabilität erzielt wird. Tatsächlich korreliert die Anzahl erinnerter Items aus verschiedenen Kategorien in der Regel mit der geschätzten Häufigkeit oder Wahrscheinlichkeit dieser Kategorien.

Obwohl die Anzahl erinnerter Items vermutlich sehr hoch korreliert ist mit der erlebten Leichtigkeit der Erinnerbarkeit, so sind die beiden Interpretationen des zugrunde liegenden kognitiven Algorithmus doch verschieden. Schwarz et al. (1991) führten eine Reihe von viel beachteten Experimenten durch, um zu entscheiden, welche der beiden Interpretationen

zutrifft. In verschiedenen Bedingungen eines Experiments baten sie die Teilnehmer, entweder sechs Beispiele aufzuschreiben, in denen sie selbstsicheres Verhalten gezeigt hatten, oder aber zwölf Beispiele für selbstsicheres Verhalten. Hinterher wurden die Teilnehmer gebeten, auf einer Rating-Skala zu beurteilen, in welchem Maße auf sie das Attribut Selbstsicherheit zutrifft. Da es den Teilnehmern deutlich schwerer fiel, 12 Beispiele zu generieren als 6 Beispiele, wurden durch diese Versuchsanordnung die beiden Interpretationen von Verfügbarkeit gegeneinander getestet. Wie sich herausstellte, lag die selbst eingeschätzte Selbstsicherheit in der Bedingung mit wenigen (6) aber leicht erinnerbaren Beispielen systematisch höher als in der Bedingung mit vielen (12) aber schwer erinnerbaren Beispielen. Diese Befunde sprachen deutlich zugunsten eines kognitiven Prozesses, der die erlebte Leichtigkeit widerspiegelt.

Eine Reihe von anderen Experimenten bestätigten diese Befunde (z. B. Ruder & Bless, 2003; Wänke, Bless und Biller, 1996) und konnten darüber hinaus zeigen, dass es nicht einmal auf die Leichtigkeit eines tatsächlich vollzogenen Abruf-Prozesses aus dem Gedächtnis ankommt, sondern auf die bloße Attribution, dass Erinnerung leicht fällt, egal ob sie vollzogen wird (Wänke, Bohner & Jurkowitsch, 1997) und auch dann, wenn es sich um eine experimentell induzierte Fehlattriution handelt (Schwarz et al., 1991; Wänke, Schwarz & Bless, 1995). Bekommen Urteiler beispielsweise gesagt, dass die Leichtigkeit, mit der Wörter mit „t“ als ersten Buchstaben erinnert werden können, auf die blass gedruckten „t“s im Hintergrund des Fragebogens zurückzuführen ist, dann wird der so entstandene Cue (Leichtigkeit als Maß für Verfügbarkeit) nicht genutzt (Wänke et al., 1995). Die geschätzte Anzahl von Wörtern mit „t“ als ersten Buchstaben ist nicht erhöht, wenn das Generieren von Beispielen leicht fällt, dies aber externen Umständen zugeschrieben werden kann.

In diesen und vielen anderen Untersuchungen wird die Leichtigkeit des Abrufs wie bei Schwarz et al. (1991) manipuliert. Wenn wenige Beispiele erinnert werden müssen, wird das als leichter erlebt als, wenn viele erinnert werden müssen. Durch diese Manipulation wird ein

Konflikt zwischen der Anzahl und der Leichtigkeit des Erinnerns erzwungen. Dass möglicherweise beide Masse der Verfügbarkeit bzw. Gedächtnisstärke zugleich für induktive Schlüsse genutzt werden, kann mit dieser Methode nicht festgestellt werden. Es kann durchaus sein, dass auch die Anzahl erinnerter Exemplare ein nützlicher Cue ist, der jedoch von der erlebten Leichtigkeit überdeckt wird, wenn die beiden Masse im Konflikt sind.

Tatsächlich zeigten sich in neueren Untersuchungen mehrere Moderatoren, von denen es abhing, ob entweder die Leichtigkeit oder die Anzahl erinnerbarer Beispiele das Urteil bestimmen. Ruder und Bless (2003) fanden, dass sich Urteile in guter Stimmung eher auf die Leichtigkeit des Abrufs, in schlechter Stimmung hingegen auf die Anzahl verfügbarer Beispiele verlassen. Wänke et al. (1996) konnten zeigen, dass der Aspekt der Leichtigkeit überwiegt, wenn Urteiler selbst Beispiele für eine Kategorie generieren, dass jedoch die Anzahl als Maß der Verfügbarkeit genutzt wird, wenn Urteiler beobachten, wenn Beispiele durch andere generiert werden. Ein weiterer Befund besagt, dass hoch involvierte Urteiler (z.B. von Krankheitsfällen selbst Betroffene) bei der Bewertung von Gesundheitsrisiken eher die Anzahl erinnerbarer Fälle heranziehen, gering Involvierte hingegen eher die Leichtigkeit. Um den relativen Einfluss beider Aspekte des Verfügbarkeits-Cues sauber trennen und schätzen zu können, werden dringend Experimente benötigt, in denen Leichtigkeit und Anzahl erinnerter Beispiele orthogonal manipuliert werden, so dass man den Einfluss des einen Aspekts bei konstant gehaltenem Einfluss des anderen beobachten kann.

Seit ihrer ursprünglichen Formulierung wurde der Geltungsbereich der Verfügbarkeits-Heuristik in verschiedener Weise ausgeweitet. Nicht nur die Häufigkeiten und Wahrscheinlichkeiten werden aus der Verfügbarkeit von Gedächtnisinhalten erschlossen, sondern auch andere, davon abhängende Urteile und Entscheidungen, wie Diagnosen (McReynolds, 1989), Schuld- und Verantwortungsattributionen (Reyes, Thompson & Bower, 1980) oder Einstellungen (Wänke et al., 1996). Gleichzeitig wurden neben der Erinnerbarkeit ("ease of retrieval") auch andere Aspekte des kognitiven Erlebens gefunden, die sich nicht auf

den Gedächtnis-Abruf (retrieval) beziehen, sich aber ganz ähnlich auswirken, wie zum Beispiel die Flüssigkeit der Verstehens ("fluency"; Wittlessea & Leboe, 2000) oder die Leichtigkeit der mentalen Simulation eines Ereignisses (Galinski & Moskowitz, 2000; Kahneman & Tversky, 1982). Je besser man sich ein Geschehen oder Ergebnis vorstellen oder mental simulieren kann (z.B. Ansteckung mit einer Krankheit; Eintreffen einer Katastrophe), um so wahrscheinlicher werden entsprechende Schlüsse gezogen (Koehler, 1991; Wells & Gavanski, 1989).

Der Rekognitions-Cue (Goldstein & Gigerenzer, 2002) ermöglicht eine simple und klar definierte Heuristik, die auf den elementaren Unterschied reagiert, ob ein Stimulus-Objekt gänzlich neu ist oder aber wenigstens einmal zuvor erlebt wurde. Hierbei ist es nicht von Bedeutung, ob das Wiedererkennen des Stimulus bewusst erlebt wird. Es kann auch sein, dass das zentrale Nervensystem auf irgendeine unterbewusste Weise unterschiedlich auf neue und alte Reize reagiert. Goldstein und Gigerenzer (2002) zufolge ermöglicht der Recognition-Cue erstaunlich genaue Urteile und Entscheidungen in vielen Gegenstandsbereichen (Schätzung der Einwohnerzahl von Städten; Geldanlage in Aktien), denn der einfache dichotome Unterschied zwischen gänzlich neuen und schon einmal erlebten Stimuli enthält auch objektiv sehr wertvolle diagnostische Information. Ein wiedererkannter Stimulus reflektiert einen Gegenstand in der Umwelt, der eine zeitlich stabile Existenz besitzen muss, dem man vermutlich auch in Zukunft wieder begegnen wird, der bedeutsam und interessant genug war, um im Gedächtnis enkodiert zu werden, der zu einer vertrauten Umgebung gehört und der – nicht zu vergessen – nicht lethal ist (d.h., den man schon einmal überlebt und unbeschadet überstanden hat. Auch hier zeigt sich allerdings, dass die Rekognitions-Heuristik in Aufgaben-Situationen, in denen sie mit anderen erfahrungsbasierten Einflüssen konkurrieren muss, kaum verwendet wird (Betsch, Hoffmann, Hoffrage & Plessner, 2002). Ein wichtiges Ziel zukünftiger Forschung im Bereich der Verwendung heuristischer Cues besteht daher nicht nur in der Bestimmung der Randbedingungen von Situationen, in denen sie sinnvoll

eingesetzt werden können, sondern auch in der Bestimmung von Situationen, in denen sie tatsächlich von Menschen verwendet werden (Betsch & Pohl, 2002; Bröder, 2001).

Eine bedeutsame Rolle beim induktiven Schließen spielen motivationale Cues, wobei man drei Fälle unterscheiden kann: Die Bereitschaft, eine Kategorisierung vorzunehmen, ist dann erhöht, wenn (a) die resultierende Information für die persönlichen Ziele und die eigene Identität erwünscht und angenehm ist; (b) wenn damit eine Erwartung oder Meinung bestätigt wird; (c) und wenn die Kategorisierung konsistent ist mit früheren Entscheidungen oder Verpflichtungen. Diese drei Fälle von motiviertem Schließen (Pyszczynski & Greenberg, 1987; Kunda, 1990) lassen sich kurz mit den Schlagworten Wunschdenken, Konfirmations-Täuschung ("confirmation bias") und Selbstkonsistenz bezeichnen.

Eine prominente Form des Wunschdenkens äußert sich im unrealistischen Optimismus (Weinstein, 1980). Das ist die Tendenz, Risiken für die eigene Person geringer einzuschätzen als für die übrige Bevölkerung. Beispielsweise schätzen die meisten Menschen die Wahrscheinlichkeit, selbst Opfer eines Unfalls oder Gewaltverbrechens zu werden, geringer ein als dieselbe Risikowahrscheinlichkeit für die gesamte Bevölkerung. Technisch ausgedrückt, die eigene Person wird mit geringerer subjektiver Wahrscheinlichkeit in bedrohlich Kategorien (z.B. Unfall) eingeordnet als andere Personen. Der leitende Cue ist Selbstbezug; selbstbezogene Stimuli (eigener Besitz, eigene Freunde, eigene Bezugsgruppe) werden im Vergleich zu fremdbezogenen Stimuli eher positiv und weniger negativ kategorisiert (Greenwald et al., 2002).

Motivational bedingte Schlüsse hängen neben dem Selbstbezug auch davon ab, ob eine Kategorisierung eine Erwartung oder Hypothese bestätigt. Induktive Schlüsse werden eher gezogen, wenn sie eine Erwartung bestätigen als wenn sie die Erwartung widerlegen. In der sozialen Wahrnehmung werden negative Eigenschaften eher bei solchen Personen erschlossen, die unter negativen Stereotypen leiden (Bodenhausen, 1990). Und auch

Wissenschaftler gelangen bereitwilliger zu Schlüssen, die die eigene Hypothese bestätigen, als zu widerlegenden Schlüssen. Die Übereinstimmung mit einer vorherrschenden Erwartung kann ein starker inferenzleitender Cue sein.

Selbstkonsistenz ist – wie auch Erwartungskonsistenz – ebenfalls ein relationaler Cue. In dem Maße, wie eine Schlussfolgerung mit einer vorausgehenden Schlussfolgerung oder Handlung konsistent ist, nimmt die Wahrscheinlichkeit zu (Jonas, Schulz-Hardt, Frey & Thelen, 2001; Schulz-Hardt, Frey, Luethgens & Moscovici, 2000). Wie auch die anderen beiden zuvor behandelten motivationalen Cues sollte dieser Cue im Hinblick auf seinen funktionellen Wert nicht unterschätzt werden. Vor allem dann, wenn verlässlichere Cues nicht bekannt sind, führt die Konsistenz mit vorausgehenden Entscheidungen und Hypothesen in vielen Fällen zu richtigen Urteilen. So wie das Wetter des Vortages meist die beste Prognose für das Wetter des heutigen Tages darstellt, sollten politische Wahlen, Konsumentenentscheidungen oder Schülerleistungen meist gut aus vorausgehenden Handlungen derselben Personen vorhersagbar sein. Auch Erwartungskonsistenz ist keineswegs ein irrationaler Cue, sondern bedeutet nichts anderes als Vertrauen auf eine Theorie oder frühere Erfahrung. Wunschdenken oder Selbstreferenz als leitender Cue lässt sich zwar weniger leicht rationalisieren, aber auch diese Strategie des Schließens kann zumindest eine motivierende und selbstwertfestigende Funktion haben (vgl. Alloy & Abramson, 1979).

*Kategorisierungen aufgrund multipler Merkmale.* In vielen Fällen müssen bei der Kategorisierung verschiedene Merkmale oder Cues zugleich beachtet, gegeneinander abgewogen und optimal gewichtet werden, beispielsweise, wenn für die Diagnose einer Krankheit verschiedene Symptome bekannt sind oder wenn mehr als ein Cue zur Lügendetektion vorliegt. Neben der Frage, wie diese verschiedenen Cues kombiniert und gewichtet werden, stellt sich im Kontext solcher Aufgaben mit multiplen Cues auch die Frage, wie (gut) die Validität der einzelnen Cues erkannt werden kann.

Experimente zum Kategorisieren und zum Wahrscheinlichkeitslernen mit mehreren Cues wurden bereits seit vielen Jahrzehnten durchgeführt (Goldberg, 1968, 1970; Hursch, Hammond & Hursch, 1964; Hammond & Summers, 1972). Neben der Frage, wie systematisch und effektiv solche komplexen Aufgaben gelöst werden (vgl. Slovic & Lichtenstein, 1971), interessierte von Anfang an die Frage nach den kognitiven Prozessen und Verarbeitungsstrategien. Während Urteiler – etwa Experten in Medizin und Wirtschaft (Goldberg, 1968, 1970) – überzeugt sind, die Cues auf komplexe und konfigurale Weise zu kombinieren, vermögen dennoch einfache lineare Modelle die Urteile meist sehr genau vorherzusagen (Dawes, 1979).

Auch in jüngeren Untersuchungen über Kategorisierung aufgrund multipler Cues steht die Analyse kognitiver Prozesse im Vordergrund. Während manche Autoren die These vertreten, dass die Komplexität vielfacher Cues reduziert wird, indem sich Urteiler im Sinne einer sogenannten lexikografischen Strategie (Martignon & Hoffrage, 1999) auf einzelne Cues konzentrieren, zum Beispiel auf den besten oder validesten Cue. Gigerenzer und Goldstein (1996) konnten zeigen, dass eine solche Take-the-Best Strategie in vielen Fällen korrekte Vorhersagen liefert, besonders, wenn die Validitäten der Cues bekannt sind und einer ausreichend starken Rangordnung unterliegen (Martignon & Laskey, 1999). Allerdings ist weniger darüber bekannt, wie häufig solche Einzel-Cue Strategien in der Realität von menschlichen Urteilern eingesetzt werden. Verschiedene Autoren haben indessen theoretische (Juslin & Persson, 2002) wie auch empirische Argumente (Bröder, 2003) für simultane Wirkungen mehrerer Cues und Wechselwirkungen bzw. Konkurrenz verschiedener Cues vorgelegt (Chapman & Robbins, 1990; Edgell, 1978, 1980). Beispielsweise wird der Einfluss einer bestimmten Ursache (z.B. ein Fahrfehler als Ursache eines Verkehrsunfalls) geringer gesehen, wenn eine andere Ursache bekannt ist (technischer Fehler im Auto), was gemeinhin als Discounting bezeichnet wird (McClure, 1993).

Sicher gibt es keine allgemeingültige Regel, nach der multiple Cues bei allen Kategorisierungsproblemen kognitiv verarbeitet werden. Verschiedene Probleminhalte und –anforderungen können vielmehr zu unterschiedlichen Prozessen führen. Juslin, Jones, Olsson und Winman (2003) konnten jüngst zeigen, dass sogar kleine Variationen in der Präsentation – bei völlig konstant gehaltenen Aufgabeninhalten – zu systematisch verschiedenen kognitiven Prozessen führen können. Die Teilnehmer dieser Untersuchung mussten aufgrund von vier grafisch dargestellten Cues die Giftigkeit von Insekten vorhersagen. Je nach experimenteller Bedingung war die Kategorisierung entweder als dichotomes Urteil abzugeben (giftig oder nicht) oder aber als kontinuierliches Urteil (Grad der Giftigkeit). Wie sich über mehrere Experimente hinweg systematisch zeigte, induziert ein dichotomes Antwortformat eine exemplarbasierte Repräsentation im Gedächtnis (d.h., Erinnern von diskreten Beispielen für die Kategorie "Giftig"), während ein kontinuierliches Antwortformat eine kognitive Repräsentation hervorruft, die auf abstrahierten Merkmalen beruht. Auch andere jüngere Forschungsansätze bekräftigen die Erkenntnis, dass die kognitiven Prozesse, die kategorialen Schlüssen zugrunde liegen, sehr flexibel und anpassungsfähig sind und nicht in eine enge Schablone gezwungen werden können (Huttenlocher, Hedges & Vevea, 2000).

*Notwendige und hinreichende Bedingungen für induktive Schlüsse.* Bei vielen alltäglichen aber auch wissenschaftlichen Problemen erfüllen Kategorisierungen die Funktion einer Erklärung oder Interpretation. So stellt beispielsweise die Erklärung „Vermutlich kam es zu dem katastrophalen Unfall, weil der Fahrer ermüdet war“ insofern auch eine Kategorisierung der Form wenn  $p(B \in C)$  dar, als die Beobachtung  $B$  (Unfall) mit einer bestimmten Wahrscheinlichkeit der erklärenden Kategorie  $C$  (Ermüdung) zugeschrieben wird. Dasselbe ist in der Wissenschaft der Fall, wenn beobachtete Daten  $B$  im Sinne einer theoretischen Kategorie  $C$  interpretiert werden. Alltägliche wie wissenschaftliche Erklärungen beziehen sich spontan häufiger auf hinreichende als auf notwendige Bedingungen für das zu erklärende Ereignis (Einhorn & Hogarth, 1986). Das heißt, eine als wahrscheinlich erachtete

erklärende Kategorisierung gibt eher eine Antwort auf die Frage: „Liefert die Kategorie einen hinreichenden Grund für die Beobachtung?“ als auf die Frage: „Bietet ausschließlich diese Kategorie eine notwendige Bedingung für die Beobachtung?“ Diese Tendenz, nach hinreichenden statt nach notwendigen Interpretationen zu suchen, schlägt sich darin nieder, dass multiple, konjunktive Begründungen oft für wahrscheinlicher gehalten werden als einfache Begründungen (Zuckerman, Eghrari & Lambrecht, 1986). Eine akzeptable Erklärung für einen schweren Verkehrsunfall lautet nicht nur, dass der Fahrer ermüdet war, sondern dass der Fahrer ermüdet war UND die Kreuzung gefährlich war UND der Unfallpartner auch nicht aufgepasst hat. Eine solche konjunktive Erklärung liefert zwar eher eine hinreichende Begründung für einen schweren Unfall, aber als notwendige Bedingung kann sie nicht wahrscheinlicher sein als die einfache Begründung. Denn die Konjunktion mehrerer Bedingungen (Ermüdung & Kreuzung & Unfallpartner) kann logisch nicht wahrscheinlicher sein als eine Bedingung allein (Ermüdung).

Tatsächlich gibt es jedoch eine Vielzahl von empirischen Belegen für die sogenannte Konjunktionstäuschung (Hertwig & Gigerenzer, 1999; Tversky & Kahneman, 1983). Es erscheint weniger wahrscheinlich, dass jemand aus heiterem Himmel plötzlich einen Herzanfall bekommt als dass jemand über 55 Jahre ist und plötzlich einen Herzanfall bekommt. Oder es erscheint weniger wahrscheinlich, dass in Kalifornien in den nächsten 10 Jahren eine Flutwelle über 100 Menschenleben fordert als dass in Kalifornien in den nächsten 10 Jahren ein Erdbeben eine Flutwelle auslöst, welche über 100 Menschenleben fordert.

Mengenlogisch ist eine konjunktive Kategorie, die durch viele Bedingungen gleichzeitig gekennzeichnet ist, enger und weniger inklusiver als eine einfache Kategorie. Die besagt Präferenz für konjunktive, mehrfache Erklärungen bedeutet also eine Präferenz für enge, spezifische Kategorien, die der vollen Natur des Ereignisses gerecht werden. Dabei nehmen sie in Kauf, dass diese Kategorisierungen womöglich zu eng und zu gewagt sind. Wichtig ist allerdings der Hinweis, dass die Konjunktionstäuschung nicht einen groben Denkfehler

widerspiegeln muss, sondern durch subtile Faktoren im Aufgabenverständnis zustande kommen kann (Hertwig & Gigerenzer, 1999).

### **Induktive Schlüsse bei Diskriminierungsaufgaben**

Unter Diskriminierung wird der Vergleich oder Wettbewerb zwischen zwei Kategorien verstanden. Die Grenze zwischen einfachen Kategorisierungen und Diskriminierungen ist fließend, da im Prinzip bei jeder Kategorisierung eine alternative Kategorie mitgedacht wird. Eine medizinische Diagnose beispielsweise (Kategorisierung) wird in dem Masse zu einer Differenzialdiagnose (Diskriminierung zwischen zwei Diagnosen), wie mögliche andere Diagnosen in Betracht gezogen werden. Dieser Abschnitt beschäftigt sich vor allem mit dem elementaren Fall der Auswahlentscheidung oder Diskriminierung zwischen zwei Kategorien.

#### *Theoretische Fragen*

Technisch formuliert verlangt diese Entscheidung den Vergleich der beiden Wahrscheinlichkeiten  $p(B \in C_1)$  und  $p(B \in C_2)$ , dass ein beobachteter Stimulus B der einen ( $C_1$ ) oder aber der anderen ( $C_2$ ) Kategorie angehört. Nach dem Bayes-Theorem sollte sich die Diskriminierung zwischen  $C_1$  und  $C_2$  nach dem Verhältnis der beiden a-posteriori Wahrscheinlichkeiten  $p(C_1/D)$  und  $p(C_2/D)$  richten, dass in Anbetracht der gegebenen Daten D die beiden Kategorisierungen zutreffen. Normativ hängt dieses Verhältnis nur ab vom Verhältnis der beiden a-priori Wahrscheinlichkeiten oder Basisraten  $p(C_1)$  und  $p(C_2)$  der beiden Kategorien (z.B. wie oft die beiden Krankheitsdiagnosen im allgemeinen auftreten) sowie (multipliziert mit) dem Verhältnis der beiden bedingten Wahrscheinlichkeiten  $p(D/C_1)$  und  $p(D/C_2)$ , dass die beiden Kategorien zu den vorliegenden Daten führen (z.B. dass die beiden Diagnosen die beobachteten Symptomen herbeiführen). Andere Faktoren sollten die Entscheidung nicht beeinflussen.

In der Realität trifft diese rationalistische Annahme jedoch meist nicht zu. Auswahlentscheidungen hängen von einer Vielzahl anderer Faktoren ab als nur von den Basisraten der Kategorien und der Nähe der Kategorien zu den Daten. Aus der

Umfrageforschung wohl bekannt ist die Tatsache, dass die Wahrscheinlichkeit der Kategorisierung von der Salienz und der Alternativen  $C_1, C_2, \dots, C_k$  beeinflusst wird. Eine solche Beeinflussung kann zum Beispiel dadurch geschehen, dass Interviewer oder Fragebogen-Konstrukteure bestimmte Antworten als plausibel vorgeben (Schwarz, 1999). Es kann auch sein, dass  $C_1$  vertrauter ist oder routinemäßig, ohne Grund häufiger gewählt wird als  $C_2$  (wenn man immer in dieselbe Bar geht oder denselben Heimweg nimmt; Betsch, Haberstroh, & Höhle, 2002).

Oft liegt der Entscheidung zwischen zwei Kategorien ein Kontinuum zugrunde, auf dem die Information  $D$  variiert. Liegt  $D$  auf diesem Kontinuum oberhalb eines bestimmten Cut-Off-Punktes oder Kriteriums  $K$ , dann erfolgt eine Kategorisierung in  $C_1$ ; ansonsten wird für  $C_2$  entschieden. Übersteigt der Grad der Depression eines Patienten ( $D$ ) etwa einen kritischen Wert, dann wird er als selbstmordgefährdet ( $C_1$ ) eingestuft und entsprechend behandelt; ansonsten gilt er als nicht selbstmordgefährdet ( $C_2$ ). Diese Entscheidung hängt nicht nur davon ab, wie gut die Daten über den Patienten wirklich zwischen den beiden Kategorien trennen, sondern auch von strategischen Faktoren. Da die Kosten für eine falsche Klassifikation in  $C_2$ , wenn in der Tat Selbstmordgefahr vorliegt, viel größer sind als die Kosten für den umgekehrten Fehler, wird vermutlich ein vorsichtiges Kriterium  $K$  gewählt. Dann gibt es einen Bias, im Zweifelsfall eher mehr Patienten als selbstmordgefährdet einzustufen. Die Signalentdeckungsanalyse (Swets, Dawes & Monahan, 2000) bietet ein methodisches Inventar, um solche strategischen Prozesse zu analysieren und Antwort-Bias von der eigentlichen Diskriminierbarkeit zu unterscheiden.

Schließlich richtet sich die Diskriminierung nicht nur nach der Wahrscheinlichkeit, mit der ein Stimulus  $B$  in die Kategorien  $C_1$  und  $C_2$  gehört, sondern wird auch von der semantischen oder symbolischen Ähnlichkeit beeinflusst (Tversky, 1977; Tversky & Gati, 1978). Solche Einflüsse sind fallen in das Gebiet von Stereotypen oder Aberglauben. Ein schwarzer Panther wird zum Beispiel eher als gefährlich eingestuft als ein gefleckter Panther,

nicht weil Information bekannt wäre, wonach  $p(\text{Gefahr/schwarz}) > p(\text{Gefahr/gefleckt})$ , sondern wegen der symbolischen Ähnlichkeit der Farbe schwarz zu Tod, Nacht, und Bedrohung (Shweder, 1977).

### *Empirische Phänomene*

*Wahrscheinlichkeitsurteile im Paarvergleich.* In zahlreichen Untersuchungen werden zwei Kategorien vorgegeben und Versuchsteilnehmer müssen diskriminieren, welche von beiden Kategorien wahrscheinlicher ist. Solche Diskriminationsaufgaben im Paarvergleich führen, analog zu den einfachen Kategorisierungsaufgaben im vorigen Abschnitt, häufig zu Overconfidence: Die Rate richtiger Entscheidungen liegt niedriger als die subjektive Sicherheit der Urteiler, die richtige Kategorie gewählt zu haben. Auch hier zeigt sich jedoch, dass Overconfidence verschwindet oder stark reduziert wird, wenn die Aufgaben repräsentativ aus einer vollständigen Referenzmenge und nicht nach der Intuition des Experimentators ausgewählt werden. Gigerenzer et al. (1991) gaben beispielsweise ihren Versuchsteilnehmern sämtliche Paare deutscher Städte über 100000 Einwohner vor und erhoben vergleichende Wahrscheinlichkeitsurteile darüber, welche von beiden Städten die größere ist. Unter diesen Bedingungen – wenn alle Paare einbezogen werden, die leichten wie die schwer zu schätzenden – erwiesen sich die subjektiven Wahrscheinlichkeiten der Urteiler als recht genau kalibriert.

An solchen Aufgaben entwickelten Gigerenzer et al. (1991) die im Moment vielleicht prominenteste theoretische Konzeption der subjektiven Wahrscheinlichkeit, genannt probabilistic mental models (PMM). In dieser Theorie wird angenommen, dass bei fehlendem Wissen über die wirkliche Einwohnerzahl ein mentales Modell auf der Grundlage verfügbarer Cues gebildet wird. Ein geeignetes mentales Modell für die Beurteilung der Einwohnerzahl könnte etwa folgende Cues umfassen: Ob die Stadt überhaupt bekannt ist (Recognition Cues), ob die Stadt einen Flughafen hat, ob es eine Landeshauptstadt ist, ob sie einen Bundesliga-Club stellt etc. Um eine Entscheidung zu treffen, welche der beiden Städte größer ist, werden

die Cues in einer bestimmten Reihenfolge abgearbeitet, beginnend mit dem Rekognitions-Cue (Bekanntheit der Stadt) und dann in der Reihenfolge der Cue-Validitäten (d.h. wie gut sie zwischen großen und kleinen Städten trennen). PMM nimmt also an, dass Vorzeichen und Stärke der Cue-Validitäten irgendwie bekannt sind. Die Entscheidungsregel sieht vor, dass der erste Cue, der zwischen den beiden Städten einen Unterschied macht, den Ausschlag gibt. Zeigt sich bereits ein Unterschied im ersten, dem Rekognitions-Cue, dann wird die bekannte Stadt als größer eingestuft als die unbekannt. Sind beide Städte bekannt (oder beide unbekannt), dann wird geprüft, ob der zweite Cue in der Hierarchie (z.B. Flughafen) einen Unterschied macht. Wenn ja, wird die Stadt mit Flughafen als die größere eingestuft; wenn nicht, geht der Prozess zum nächsten Cue über bis nach dem letzten verfügbaren Cue als letztes Mittel eine Zufallsentscheidung getroffen wird. So einfach und fast primitiv der PMM-Algorithmus aussehen mag, er liefert erstaunlich genaue Vorhersagen von induktiv-statistischen Schlüssen im Paarvergleich.

Im Bezugsrahmen des PMM wurden in den letzten Jahren Computersimulation eine Reihe von noch einfacheren Entscheidungsregeln untersucht, die ebenfalls recht genaue Entscheidungen unter Unsicherheit ermöglichen (Gigerenzer & Goldstein, 1996). Die am meisten diskutierte Einfachheuristik ist „Take the best“ (, wobei ausschließlich der validesten Cue benutzt wird); eine andere erstaunlich wirksame Regel ist Einheitsgewichtung aller Cues (, wobei nur das Vorzeichen, +1 bzw. -1, beachtet werden muss). Solche einfachen Regeln ermöglichen (unter bestimmten Annahmen über die Hierarchie der Cue-Validitäten) genaue Entscheidungen als aufwendige Regeln (wie multiple Regression), die alle verfügbaren Cues nach einem optimalen Verfahren zu gewichten versuchen.

*Anchoring and adjustment.* Vorausgehende vergleichende Urteile können sich im Sinne von Verankerungs-Effekten auf nachfolgende absolute und relative Urteile auswirken. In einem klassischen Experiment konnten Tversky und Kahnemann (1974) zeigen, dass Relativurteile, wie die bei Diskriminationsaufgaben typische Zuordnung eines Reizes in eine

von zwei Reizklassen, einen systematischen Einfluss auf anschließende Absoluturteile ausüben können. Die Versuchspersonen in diesem Experiment wurden gebeten, zu schätzen, wie groß die Anzahl der afrikanischen Staaten in der UNO in Prozent ist. Mit einer Art Glücksrad wurde vorgeblich eine Zahl zwischen 0 und 100 gelost. Die Versuchspersonen sollten zunächst angeben, ob der Prozentsatz der afrikanischen Staaten in der UNO größer oder kleiner als die geloste Zahl ist, und dann die genaue Zahl schätzen. Wurde mit dem Glücksrad die Zahl 65 gelost, schätzten die Versuchspersonen die Anzahl der afrikanischen Staaten in der UNO im Mittel auf 45%. War hingegen die geloste Zahl eine 10, so lagen die geschätzten Werte mit durchschnittlich 25% deutlich niedriger.

In einer Untersuchung von Plous (1987) wurden Versuchspersonen unter anderem dazu befragt, für wie wahrscheinlich sie den Ausbruch eines Atomkriegs zwischen der USA und der Sowjetunion hielten. In einer Bedingung wurden sie zunächst befragt, ob die Wahrscheinlichkeit größer oder kleiner als 90% ist, bevor sie einen genauen Wert schätzen sollten, in einer zweiten Bedingung wurden die 90% durch 1% ersetzt und in einer dritten Bedingung wurde kein Anker gegeben. Den Erwartungen entsprechend schätzten die Versuchspersonen die Wahrscheinlichkeit eines Atomkrieges bei einem Ausgangswert von 90% höher ein als unter der Bedingung ohne Anker und bei einem Ausgangswert von 1% niedriger als ohne Anker.

Northcraft & Neale (1987) untersuchten den Einfluss von Ankern auf Preisentscheidungen auf dem Wohnungsmarkt. Versuchspersonen waren neben Studenten auch Wohnungsmakler. Sie hatten 20 Minuten Zeit, ein Haus zu begutachten, dessen Wert sie einschätzen sollten. Für diese Aufgabe bekamen sie die üblichen Unterlagen mit allen möglichen Daten über das Haus und das Grundstück. Unter diesen Angaben befand sich ein fingierter Listenpreis, der in vier Versuchsbedingungen bei einem tatsächlichen Wert von \$ 135000 zwischen \$ 119900 und \$ 149900 variierte. Der Einfluss dieser Anker zeigte sich sowohl bei den Studenten als auch bei den Wohnungsmaklern. Je höher der fingierte

Listenpreis war, um so höher waren ihr geschätzter Wert des Hauses, ihr empfohlener Verkaufspreis, ihre Angabe eines vernünftigen Kaufpreises und ihr geringstes Angebot. Dabei gab nur einer von 10 Wohnungsmaklern zu, sich nach dem Listenpreis gerichtet zu haben.

Insgesamt konnten Ankereffekte in zahlreichen Domänen nachgewiesen werden und haben sich als ein sehr robustes Phänomen erwiesen (Chapman & Johnson, 1999; Englich & Mussweiler, 2001; Galinsky & Mussweiler, 2001; Mussweiler & Strack, 1999a, 1999b, 2000b; Strack & Mussweiler, 1997; Wilson, Houston, Etling & Brekke, 1996). Tversky und Kahnemann zählen die Tendenz zur Verankerung und Anpassung wie die Verfügbarkeitsheuristik zu den verkürzten Urteilsstrategien, mit denen Menschen Urteile treffen, wenn ihnen keine vollständige Information zur Verfügung steht. Eine alternative Erklärung besteht in der Annahme, dass Ankereffekt durch numerisches Priming zustande kommt (Jacowitz & Kahnemann, 1995; Wilson et al. 1996). Danach wird durch die Nennung einer Zahl bei der relativen Urteilsaufgabe genau dieser Wert aktiviert, und damit die Wahrscheinlichkeit erhöht, dass er bei der Generierung der absoluten Urteils verwendet wird. Den zur Zeit Überzeugendsten Ansatz zur Erklärung von Ankereffekten bietet jedoch das Selective Accessibility Model (Mussweiler & Strack, 1999a, 1999b, 2000b; Strack & Mussweiler, 1997). Danach generieren Personen mehr oder wenig automatisch vor allem solches Wissens, das mit dem Ankerwert konsistent ist. Die Zugänglichkeit von Anker-konsistentem Wissen nimmt dadurch zu und hat eine höhere Wahrscheinlichkeit, bei der absoluten Schätzaufgabe herangezogen zu werden, als Anker-inkonsistentes Wissen. Beispielsweise konnten Mussweiler und Strack (2000b) zeigen, dass bei Versuchspersonen, denen ein hoher Anker als Preis für ein Auto gegeben wurde, die Zugänglichkeit von mit einem hohen Preis assoziierten Konzepten (z.B. Luxus) anstieg. Arbeiten zum Ankereffekt fokussieren im Allgemeinen auf die Anpassung eines Urteils an einen vorgegebenen Wert und damit auf einen Assimilationseffekt. Generell zeigen sich bei sequentiellen Urteilen neben Assimilationseffekten aber auch häufig Kontrasteffekte (z.B. Petzold &

Georgieva, 1994; Schwarz & Bless, 1992). Aus einer Erweiterung des Selective Accessibility Models auf soziale Vergleichsurteile (Mussweiler, 2003; Mussweiler & Strack, 2000a) lassen sich sowohl Assimilations- als auch Kontrasteffekte erklären. Jedes Urteil ist danach ein Vergleich von dem Beurteiltem (Target) mit einer abstrakten oder Konkreten Vergleichskategorie (Standard). Moderierende Variable von Assimilations- oder Kontrasteffekten ist hierbei die spontan eingeschätzte Ähnlichkeit (Initialurteil) zwischen Target und Standard. Nach dem Modell wird die im Initialurteil gefällte Hypothese überprüft (Ähnlichkeits-Hypothese vs. Unähnlichkeits-Hypothese). Dazu werden Informationen gesammelt, die später bei der Beurteilung genutzt werden. Die Auswahl der aktivierten und genutzten Informationen ist aber abhängig von der Ausgangshypothese, d.h. hauptsächlich konfirmierende Informationen werden aktiviert und damit zugänglich (z.B. Trope & Liberman, 1996). Beim Urteil werden dann die aktivierten Informationen genutzt. Die Anfangshypothese wird eher bestätigt und es kommt zu Assimilations- (Ähnlichkeit) bzw. Kontrast-Effekten (Unähnlichkeit).

*Sampling Modelle.* Das Selective Accessibility Model befindet sich bereits in der theoretischen Nähe einer ganzen Klasse sogenannter Sampling-Ansätze, die induktive Schlüsse und Fehlschlüsse durch die Stichprobe an Stimulusinformation erklären, welche den Input für den kognitiven Prozess bildet. Diese Sampling-Ansätze (cf. Fiedler & Juslin, 2005) verweisen auf die Erkenntnis, dass induktive Urteile oft die gegebene Stichprobe von Stimulusdaten sehr genau widerspiegeln, der kognitive Prozess der Kalkulation einer Wahrscheinlichkeit oder Entscheidung aus dem Stimulusdaten also kaum verzerrt ist. Schwerwiegende Fehlertendenzen ergeben sich jedoch daraus, dass die gegebene Stichprobe an Stimulusinformation verzerrt ist und die Urteiler diese Fehlerquellen im Input nicht erkennen. Solche Sampling-Modelle betonen mithin die Rolle ökologischer Input-Faktoren bei der Entstehung von kognitiven Täuschungen (Fiedler, 2000a).

Eine Untersuchung von Fiedler et al. (2000) verdeutlicht dies am Beispiel der sogenannten Basisraten-Täuschung. Die Aufgabe bestand darin zu schätzen, wie wahrscheinlich es ist, dass eine Frau entweder Brustkrebs hat oder keinen Brustkrebs hat, wenn bei ihr ein positives Mammogramm festgestellt wurde. Tatsächlich wird die bedingte Wahrscheinlichkeit  $p(\text{Brustkrebs} / \text{positives Mammogramm})$  meist stark überschätzt (Eddy, 1982; Gigerenzer & Hoffrage, 1995). Denn obwohl die umgekehrte bedingte Wahrscheinlichkeit, dass eine Frau mit Brustkrebs ein positives Mammogramm hat, recht hoch ist (ca. 80%), impliziert die geringe Basisrate von Brustkrebs in der Bevölkerung (etwa 1% je nach Bezugsgruppe im Vergleich zu einer etwa 10 mal höheren Basisrate von positiven Mammogrammen), dass die meisten positiven Testbefunde ohne Brustkrebs zustande kommen. Die Überschätzung von  $p(\text{Brustkrebs} / \text{positives Mammogramm})$  bedeutet, dass dieses Verhältnis der Basisraten vernachlässigt wird.

In dieser Untersuchung konnten die Teilnehmer selbst eine beliebig große Stichprobe von Daten über die Mammographie-Ergebnisse von Frauen mit und ohne Brustkrebs sammeln, so lange, bis sie meinten, genügend Information zu haben, um die fragliche Wahrscheinlichkeit schätzen zu können. Der Sampling-Prozess wurde somit durch freie Informationssuche sichtbar gemacht, und es konnte gezeigt werden, dass die Urteile recht genau den Anteil der Brustkrebsfälle unter allen Frauen mit positivem Mammogramm in der gezogenen Stichprobe widerspiegelten. Dennoch konnten die Urteile dann extrem weit vom richtigen Wert abweichen, wenn die gezogene Stichprobe verzerrt war. Viele Teilnehmer nahmen etwa in ihre Stichprobe alle oder die meisten Frauen mit Brustkrebs auf, aber nur relativ wenige ohne Brustkrebs. In der selbst gezogenen Stichprobe waren dann die Basisrate  $p(\text{Brustkrebs})$  weit übertrieben. Dies spiegelte sich dann ganz folgerichtig in einer starken Überschätzung von  $p(\text{Brustkrebs} / \text{positives Mammogramm})$  wider. Wurde ein solcher Sampling-Effekt indessen verhindert – indem es nicht möglich war, Fälle mit oder ohne

Brustkrebs selektiv zu suchen, sondern nur Fälle mit positivem und negativem

Mammogramm – dann näherten sich die Urteile sehr genau an die korrekten Werte an.

Es liegt nahe, dass bei vielen realen Urteilen, auch wenn der Sampling-Prozess nicht beobachtbar ist, dasselbe passiert: Durch selektive Suche im Gedächtnis oder in externen Datenquellen wird eine verzerrte Stichprobe gezogen, die dann ein falsches Urteil informiert. Bittet man beispielsweise einen Mediziner, die Wahrscheinlichkeit  $p(\text{Brustkrebs} / \text{positives Mammogramm})$  aufgrund seiner Erfahrung zu schätzen, dann dürfte er besonders an Erfahrungen mit denjenigen Patienten denken, die tatsächlich Brustkrebs haben. Sofern dies geschieht – dass sich die Informationssuche an demselben Kriteriumsereignis orientiert, dessen Wahrscheinlichkeit gefragt ist – dann sind starke Überschätzungen zu erwarten.

*Diskrimination und Antworttendenzen.* Sampling-Modelle erklären datengetriebene Bottom-up-Einflüsse auf Kategorisierungen und Diskriminierungen. Aber auch strategische Top-Down-Einflüsse können eine große Rolle spielen. Bei gerichtlichen Entscheidungen geht es beispielsweise um die Diskriminierung zwischen Schuld und Unschuld. Ein Augenzeuge muss bei einer Gegenüberstellung unter Unsicherheit entscheiden, ob ein Verdächtiger identisch mit oder verschieden von dem Täter. Bei konstant gehaltener Information des Augenzeugen über den Tathergang kann seine Entscheidung erheblich davon abhängen, wie eng er die Kategorie „identisch“ fasst bzw. wie strikt das gewählte Entscheidungskriterium ist. Hier spielen motivationale Einflüsse eine große Rolle. Aus der forensischen Forschung ist bekannt, dass Augenzeugen normalerweise eine starke Antworttendenz haben, sich zu leicht für „identisch“ zu entscheiden. Daher gibt es neben vielen Treffern (richtigen Identifikationen von Schuldigen) auch relativ viele falsche Alarme (Falschidentifikationen von Unschuldigen), besonders wenn Gegenüberstellungen simultan erfolgen. Diese Antworttendenz, die dem sozialen Druck auf den Zeugen entspringt, zur Identifikation des Täters beizutragen, lässt sich vermeiden, indem Gegenüberstellungen sukzessive durchgeführt und Zeugen eigens daran erinnert werden, dass der Verdächtige nicht der Täter sein muss. Durch solche Instruktionen

wird die Kategorie „identisch“ restriktiver gefasst bzw. ein strengeres Kriterium für die Identifikation gesetzt.

### **Lernen und Entdecken von Kontingenzen**

Die Fähigkeit, Zusammenhänge zwischen Ereignisklassen zu erkennen, zählt zu den zentralen Merkmalen adaptiver Intelligenz (Inhelder & Piaget, 1958). Sie erlaubt es uns, vom Auftreten eines bestimmten Ereignisses mit einer entsprechenden Wahrscheinlichkeit auf das Auftreten eines anderen zu schließen. So erwerben wir beispielsweise auf der elementaren Ebene assoziativen Lernens Wissen über die Zusammenhänge von Signalen und ihrer Bedeutung, Verhaltensweisen und Belohnung vs. Bestrafung, auf komplexerer Ebene Annahmen über Zusammenhänge von sozialen Gruppen und bestimmten Eigenschaften (Stereotype) oder zwischen Symptomklassen und bestimmten Krankheiten. Mit diesem Wissen ist es uns möglich, Gefahren zu vermeiden, Sicherheit und Belohnungen anzustreben, und allgemein unsere Verhaltensweisen situativen Gegebenheiten anzupassen. Systematische Fehleinschätzungen von tatsächlichen Zusammenhängen können entsprechend zu falschen Vorhersagen und Entscheidungen führen, die jede Art von schmerzhaften Konsequenzen nach sich ziehen können, beispielsweise kann die fehlerhafte Wahrnehmung eines Zusammenhangs von bestimmten Symptomen und einer Krankheit zu Fehldiagnosen und Anwendung unzureichender Behandlungsmethoden führen. Wenn eine subjektive Korrelationsschätzung von einer objektiven Korrelation abweicht, spricht man im allgemeinen von einer illusorischen Korrelation (Chapman & Chapman, 1967; Crocker, 1981; Fiedler, 2000b). Meistens geht es dabei um das Erkennen einer Kontingenz bei tatsächlicher Nullkorrelation. Es wird angenommen, dass illusorischen Korrelationen unter anderem bei der Entstehung und Aufrechterhaltung von Stereotypen eine bedeutende Rolle zukommt (Hamilton & Rose, 1980). Dementsprechend können Stereotype als erwartete Korrelation zwischen einer Gruppenzugehörigkeit und bestimmten Eigenschaften definiert werden (McCauley & Stitt, 1978).

### *Theoretische Fragen*

*Integrationsregeln für korrelierte Beobachtungen.* Obwohl sich das Erkennen von Kontingenzen im Prinzip auf jede Art von Zusammenhängen zwischen zwei Variablen beziehen kann, basieren die meisten Arbeiten in diesem Bereich auf den vergleichsweise einfach darzustellenden Fall des Zusammenhangs zwischen zwei dichotomen Variablen. Die Kombination von zwei Variablen mit den jeweiligen Ausprägungen + und – ergibt vier beobachtbare Ereignisklassen, deren Häufigkeiten sich in einer Vierfeldertafel darstellen lassen (Abbildung 3). Aus dem Verhältnis des gemeinsamen Auftretens und Nichtauftretens von positiven und negativen Ausprägungen zweier Variablen lässt sich der Zusammenhang der beiden Variablen erschließen. Übliche Maße zu Bestimmung einer objektiven Korrelation sind der Phi-Koeffizient mit  $\Phi = (ad - bc) / \sqrt{ab + cd + ac + bd}$  und Delta mit  $\Delta = (a / a + b) - (c / c + d)$ . Die leichte Unterschiedlichkeit dieser Maße in ihrer Sensibilität für schiefe Häufigkeitsverteilungen ist für die meisten Studien zur Erfassung von Kontingenzen kaum von Belang, beide Maße zeigen sich für die meisten Anwendungen hinlänglich robust. Systematische Abweichungen von den mit diesen Maßen bestimmten Korrelationskoeffizienten können jeweils als illusorische Korrelationen betrachtet werden.

*Verzerrte Informationsverarbeitung.* Die Entstehung von illusorischen Korrelationen steht im Mittelpunkt der Untersuchung von menschlichen Kontingenzurteilen. Wie generell für die Untersuchung von kognitiven Täuschungen, so verspricht man sich auch hier von der Analyse systematischer Abweichungen von objektiven Korrelationen einen Aufschluss über die Prozesse, die Kontingenzurteilen zu Grunde liegen. Die meisten Erklärungsansätze fokussieren dabei auf Aspekte verzerrter Informationsverarbeitung. So werden illusorische Korrelationen beispielsweise auf den generell verzerrenden Einfluss von Erwartungen attribuiert, die auf alle Stufen der Informationsverarbeitung von der Wahrnehmung bis zur Urteilsintegration einwirken können (Miller & Turnbull, 1986). Die generelle Idee ist dabei, dass Erwartungen und andere Einflussfaktoren aus verschiedenen Gründen zu einer

Ungleichgewichtung der vier Zellen einer Kontingenztafel in einem entsprechenden Urteilsprozess führen, also nicht alle Beobachtungen gleichermaßen beachtet und verarbeitet werden.

Mindestens drei Varianten von illusorischen Korrelationen lassen sich unterscheiden (Fiedler, 2000b):

(1) Erwartungsbasierte Illusionen entstehen, wenn erwartete Ereignisse stärker gewichtet oder kognitiv tiefer erlaboriert werden als unerwartete. Dies liegt vielen sozialen Stereotypen zugrunde. So werden typisch weibliche Verhaltensweisen auch dann häufiger Frauen zugeschrieben als Männern, wenn die objektive Häufigkeit exakt gleich ist.

(2) Eine andere Klasse von illusorischen Korrelationen entsteht durch die ungleiche Verarbeitung von gegebenen und fehlenden Merkmalen. Da die Anwesenheit eines Symptoms oder einer Krankheit augenfälliger ist als die Abwesenheit, hängt die Beurteilung der Korrelation vor allem davon ab, wie häufig Symptom und Krankheit gemeinsam auftreten. Eine gleich große Korrelation, die dadurch entsteht, dass Symptom und Krankheit häufig gemeinsam nicht auftreten, wird deutlich geringer eingeschätzt.

(3) Schließlich entstehen eingebildete Korrelationen durch schiefe Häufigkeitsverteilungen. Wenn in zwei Gruppen die gleiche Tendenz (z.B. mehr positives als negatives Verhalten) herrscht, über die eine Gruppe jedoch mehr Beobachtungen vorliegen, dann erscheint die Tendenz in der größeren Gruppe deutlicher.

Alle drei Varianten können sehr gut durch ein konnektionistisches Computer-Modell BIAS („Brunswikian Induction Algorithm for Social Inference“; Fiedler, 1996) erklärt werden, welches nichts als einfache assoziative Lernannahmen zugrunde legt.

Eine aktuelle theoretische Frage bezieht sich auf die optimale Stichprobengröße, die sich für das Erkennen von Kontingenzen benötigt wird. Wie alle statistischen Inferenzschlüsse werden quantitative Schätzungen von Kontingenzen mit wachsender Zahl von Beobachtungen immer genauer. Wenn es jedoch nicht um die quantitative Schätzung,

sondern um das möglichst sensible Entdecken von existierenden Kontingenzen geht, dann kann es in der Tat sein, dass kleine Stichproben überlegen sind. Dieser scheinbar paradoxe Befund wurde von Kareev (2000) aufgezeigt und damit erklärt, dass – wegen der Schiefe der Stichprobenverteilung von Korrelationen – real existierende Korrelationen besonders in kleinen Stichproben meist überschätzt werden. So wird es verständlich, dass Menschen mit geringerer Gedächtnisspanne – die nur eine begrenzte Stichprobengröße im Arbeitsgedächtnis zulässt – Kontingenzen besser erkennen können als Menschen mit höherer Kapazität des Arbeitsgedächtnisses (Kareev, 1995).

### *Empirische Phänomene*

*Leistung und Genauigkeit beim Erkennen von Korrelationen.* Da der Großteil der Literatur zum Erkennen von Kontingenzen auf systematische Abweichungen von objektiv erfassbaren Korrelationen fokussiert, den illusorischen Korrelationen, könnte der Eindruck entstehen, dass Menschen mit dieser Aufgabe hoffnungslos überfordert sind. Demgegenüber stehen Arbeiten, die zeigen, dass Menschen durchaus in der Lage sind, die genauen Zusammenhänge zwischen zwei oder mehr Variablen auch in großen Stichproben von sequentiellen Beobachtungen akkurat zu erkennen (z.B. Fiedler, Walther, Freytag, & Plessner, 2002; Plessner, Hartmann, Hohmann & Zimmermann, 2001). Einschränkungen ergeben sich hier in der Regel aus der Repräsentativität der jeweils verwendeten Stichproben für ein bestimmtes Urteilsproblem und nicht aus der verzerrten Wahrnehmung von Kontingenzen (Fiedler, 2000a, 2000b).

McKenzie (1994) konnte anhand von Monte Carlo Simulationen zeigen, dass einfache Algorithmen zur Einschätzung von Korrelationen für eine große Anzahl von wahrscheinlichen Verteilungen ziemlich genaue Urteile ermöglichen. Beispielsweise korreliert die Summe der diagonalen Zellhäufigkeiten ( $a + d$  in Abbildung 3) unter den meisten Bedingungen hoch mit dem tatsächlichen Phi-Koeffizienten. Eine einfache Einschätzung der relativen Zellhäufigkeiten  $a$  und  $d$  kann insofern zu sehr genauen

Kontingenzwahrnehmungen führen. Auch hier ist aber bisher ungeklärt, unter welchen Bedingungen Menschen solche vereinfachten Urteilsstrategien tatsächlich verwenden.

*Illusorische Korrelationen durch Erwartung.* Studien zum Einfluss von Erwartungen auf die soziale Informationsverarbeitung zeigen entsprechende Effekte auf nahezu allen Stufen der Verarbeitung, von der Wahrnehmung bis zur Informationsintegration beim Urteilen. Darüber hinaus können Erwartungen durch das eigene Verhalten in Sinne von sich selbst erfüllenden Prophezeiungen auch einen Einfluss auf der Verhalten von anderen Personen nehmen (Miller & Turnbull, 1986; Rosenthal & Jacobson, 1968). Im wesentlichen zeigt sich in diesen Studien eine Tendenz von Erwartungen zur ihrer Selbstbestätigung. Dementsprechend wurde auch für das Erkennen von Kontingenzen vermutet, dass Menschen eine Neigung haben, Zusammenhänge besonders dann wahrzunehmen, wenn sie diese erwarten. Tatsächlich konnte in zahlreichen Experimenten gezeigt werden, dass Kontingenzurteile häufig einem Kompromiss zwischen vorherigen Erwartungen und tatsächlichen Beobachtungen entsprechen (z.B., Berndsen, Van der Pligt, Spears, & McGarty, 1996; Chapman & Chapman, 1967, 1969; Fiedler, Hemmeter, & Hofmann, 1984; Hamilton & Rose, 1980; Kim & Baron, 1988; Mackie, et. al., 1989; Plessner, Freytag & Fiedler, 2000; Spears, Eiser, & Van der Pligt, 1987). In den meisten Studien haben die Versuchspersonen eine positive Korrelation zwischen zwei objektiv unkorrelierten Stimulusklassen wahrgenommen. Beispielsweise legten Spears, Eiser und Van der Pligt (1987) ihren Versuchsteilnehmern eine Reihe von Aussagen pro oder contra Atomkraft vor, die entweder von Einwohnern einer großen oder einer kleinen Gemeinde gemacht wurden. Tatsächlich gab es keinen Zusammenhang zwischen der Größe des Herkunftsortes und der Anzahl von pro und contra Aussagen. Die Erwartungen der Versuchspersonen, dass gerade Einwohner kleiner Gemeinden besonders gegen Atomkraft eingestellt sind, führten jedoch dazu, dass sie einen entsprechenden Zusammenhang wahrnahmen, der faktisch nicht gegeben war. Die typische Erklärung für diese Art erwartungsbasierter illusorischer Korrelation lautet, dass

erwartungskongruente Informationen einen Verarbeitungsvorteil gegenüber erwartungsinkongruenten Informationen hätten (Hamilton, 1981). Diese Annahme erscheint allerdings wenig plausibel, da es in der Literatur auch zahlreiche Belege für einen Gedächtnisvorteil von erwartungsinkongruenten Informationen gibt (Stangor & McMillan, 1992). Dieser tritt vor allem dann auf, wenn Versuchspersonen nicht ausreichend Informationsverarbeitungskapazität auf die Enkodierung von Stimulusinformationen verwenden (Macrae, Hewstone & Griffiths, 1993). Erwartungsbasierte illusorische Korrelationen scheinen daher generell eher einen oberflächlicheren Verarbeitungsprozess wiederzuspiegeln. Zudem zeigte sich, dass sie empirisch nur sehr schwer von solchen illusorischen Korrelationen zu trennen sind, die durch die Ähnlichkeit von Attributsdimensionen entstehen können (Fiedler, 2000b; Plessner, Freytag & Fiedler, 2001). Unter Ähnlichkeit ist dabei gemeint, dass die Cue-Systeme, aus denen das Vorhandensein der Attribute erschlossen wird, sich überlappen (Medin, Goldstone & Genter, 1993; Tversky, 1977). In der Untersuchung von Plessner, Freytag und Fiedler (2001) lernten die Versuchspersonen, abstrakten Bildern anhand von multiprobabilistischen Cues die Eigenschaften eines Maler („Greve“) und eines Malstils („Wenturalismus“) zuzuschreiben. Für eine Hälfte der Versuchspersonen überlappten sich die Cue-Systeme für Maler und Malstil, für die andere Hälfte nicht. Nach der Lernphase des Experiments wurde den Versuchspersonen eine Serie von 40 abstrakten Bildern gezeigt, die sich für beide Bedingungen in gleicher Weise mit jeweils 10 Bildern auf die vier Zellen der Kontingenztafel aufteilten. In der Versuchsbedingung ohne Überlappung in den Cue-Systemen erkannten die Versuchspersonen korrekterweise eine Nullkorrelation zwischen Maler und Malstil. Die Versuchspersonen mit überlappenden Cue-Systemen erkannten im Mittel jedoch einen positiven Zusammenhang, sie nahmen an, dass Greve überzufällig häufig im Wenturalismus malt. Zusätzlich konnte gezeigt werden, dass dieser Effekt unabhängig von den Erwartungen

der Versuchspersonen auftritt, wobei diese in diesem Paradigma vergleichbar große illusorische Korrelationen erzeugen.

*Illusorische Korrelationen durch Ungleichheit von positiver und negativer Information.*

Ein anderer Beleg, dass illusorische Korrelationen auch ohne vorheriges Wissen und Erwartungen entstehen können, sind Beobachtungen, die zeigen, dass Menschen die Häufigkeiten in den vier Zellen der Vierfeldertafel (Abbildung 3) unterschiedlich in ihren Urteilen gewichten (z.B. Allan & Jenkins, 1983, Jenkins & Ward, 1965; Wasserman, Elek, Chatlosh, & Baker, 1993). In einem Experiment von Kao und Wassermann (1993) sollten die Versuchspersonen den Wert eines Düngemittels in Bezug auf seine Fähigkeit, die unbekannte Pflanze Lanyu zum blühen zu bringen, einschätzen. Sie erhielten Informationen über die Häufigkeit des Blühens von Lanyu in Abhängigkeit von der Benutzung des Düngemittels in allen vier Kombinationen. Es zeigte sich, dass selbst bei gleicher relativer Häufigkeit von Blühen/Düngen und Nicht-Blühen/Düngen eine positive Korrelation zwischen Düngen und Blühen angenommen wurde, wenn die absolute Häufigkeit der Kombination Blühen/Düngen die von Nicht-Blühen/Düngen übertraf. Bei einer Umkehr dieser absoluten Häufigkeiten wurde ein negativer Zusammenhang zwischen Düngen und Blühen angenommen. Außerdem wurde eine positive Korrelation wahrgenommen, wenn die gleiche relative Häufigkeit von Blühen zu Nicht-Blühen beim Düngen absolut höher war als beim Nicht-Düngen. Ein umgekehrtes Verhältnis führte wiederum zu negativen Korrelationsschätzungen. Als Erklärung für diese Befunde wird im Allgemeinen eine unterschiedliche Gewichtung (Asymmetrie) von positiven und negativen Effekten herangezogen (z.B. feature-positve-Effekt; Newman, Wolff, & Hearst, 1980). Beispielsweise erzeugt die Anwesenheit eines Verkehrszeichen mehr Aufmerksamkeit als das Nichtvorhandensein. Nach Wasserman, Dornier und Kao (1990) wird dementsprechend bei Korrelationsschätzungen der Zelle a (beide Effekte sind positiv/anwesend) am meisten Gewicht gegeben und der Zelle d (beide Effekte sind negativ/abwesend) am wenigsten. Diese ungleiche Gewichtung wird vor allem einer

frühen Stufe der Informationsverarbeitung zugeschrieben, der Wahrnehmung oder Enkodierung von Stimulusinformationen. Theoretisch kann sie sich aber auf die Integration der Beobachtungen über die vier Zellen in einem Gesamturteil niederschlagen.

*Illusorische Korrelationen aufgrund schiefer Häufigkeitsverteilungen.* Nicht nur positive Effekte können im Vergleich zu negativen Effekten besondere Aufmerksamkeit und Gewichtung erfahren, sondern generell solche Ereignisse, die herausstechen, salient oder distinkt sind (Taylor & Fiske, 1978; von Restorff, 1933). In einer klassischen Studie aus dem Bereich der Stereotypenbildung (Hamilton & Gifford, 1976) erhielten die Versuchspersonen 26 Verhaltensbeschreibungen einer Gruppe A (der Majorität) und 13 Verhaltensbeschreibungen einer Gruppe B (der Minorität). Die Verhaltensbeschreibungen enthielten für beide Gruppe die gleiche relative Häufigkeit von wünschenswerten (positiven) Verhaltensweisen und weniger wünschenswerten (negativen) Verhaltensweisen (18 positive zu 8 negativen bei Gruppe A und 9 positive zu 4 negativen bei Gruppe B). Trotz der offensichtlichen Null-Korrelation zwischen Verhalten und Gruppenzugehörigkeit wurde die Gruppe A von den Versuchspersonen auf verschiedenen abhängigen Variablen als positiver beschrieben als die Gruppe B. Dieser Effekt hat sich über zahlreiche Studien als sehr robust erwiesen (Mullen & Johnson, 1990). Eine lange vorherrschende Erklärung für diesen Effekt liefert der so genannte distinctiveness Ansatz (Hamilton & Sherman, 1989). Danach erzeugt die Kombination der beiden seltensten Ereignisse (Minderheit und unerwünschtes Verhalten) besondere Aufmerksamkeit und verschafft der Zelle d (Abbildung 3) einen Gedächtnisvorteil. Dieser Ansatz hat eine große Relevanz für das Verständnis von Minderheitendiskrimination, da Minderheiten per Definition seltener sind als Majoritäten und negatives Verhalten als normabweichendes Verhalten seltener Auftritt als positives Verhalten. Allerdings lassen sich Befunde wie die von Hamilton und Gifford (1976) auch ohne die Annahme eines Gedächtnisvorteils für seltene Ereignisse erklären. Im Sinne eines allgemeinen Lernprinzips kann, wie es schon für einfache Kategorierungen beschrieben wurde, angenommen werden,

dass Urteile über kleinere Stichproben regressiver ausfallen als über größere Stichproben (Fiedler, 1996). Daher fällt es Personen leichter, in der größeren Stichprobe der Majorität das wahre Verhältnis von positiven zu negativen Verhaltensweisen, bzw. die Mehrheit von positiven Verhaltensweisen zu erkennen, als in der kleineren Stichprobe der Minorität. Demnach ist in diesem Paradigma die Schiefe der Häufigkeitsverteilung in der Vierfeldertafel verantwortlich für die Entstehung von illusorischen Korrelationen und nicht der vermeintliche Gedächtnisvorteil der seltensten Kombination in Zelle d (Fiedler, 1991; Fiedler, Russer & Gramm, 1993).

*Pseudokontingenzen.* Eine gänzlich anderer Prozess, der das Auftreten von eingebildeten Korrelationen erklären kann, wurde kürzlich von Fiedler und Freytag (2004) aufgezeigt. Statt wirklich die Kontingenz von zwei Merkmalen X und Y zu erfassen – also durch Beobachten oder Nachdenken zu ermitteln, ob die Wahrscheinlichkeit eines hohen (vs. niedrigen) Y-Wertes sich bei hohen und niedrigen X-Werten unterscheidet – werden einfach die beiden Einzel-Verteilungen der X- und Y-Werte verglichen. Sind in einem bestimmten Realitätsbereich etwa die meisten X-Werte hoch sind und die meisten Y-Werte ebenfalls hoch, dann wird darauf geschlossen, dass X und Y positiv korrelieren. Wenn indessen meist hohe X-Werte mit meist niedrigen Y-Werten auftreten, dann wird auf eine negative Korrelation geschlossen. Dieser Schluss wurde Pseudo-Kontingenz genannt, weil er einem Kategorienfehler entspringt; die vorhandenen Daten sagen über eine Kontingenz gar nichts aus. Wenn in einem Supermarkt die meisten Preise hoch sind und die meisten Waren gute Qualität haben, besagt das noch nichts über den Zusammenhang von Preis und Qualität; es kann sein, dass die wenigen billigen Artikel gleich gute oder sogar bessere Qualität haben als die vielen teuren Artikel. Gezielte Experimente, in denen die Kontingenz zwischen zwei Merkmalen X und Y,  $\Delta = p(Y_{hoch}/X_{hoch}) - p(Y_{hoch}/X_{niedrig})$ , bei null oder gar bei einem negativen Wert kontrolliert wird und nur die Basisraten oder Auftretenshäufigkeiten der beiden Merkmale  $p(X_{hoch}), p(X_{niedrig})$  und  $p(Y_{hoch}), p(Y_{niedrig})$  variiert werden, zeugen von

teilweise sehr starken Pseudo-Kontingenz-Illusionen (Fiedler & Freytag, 2004; Meiser, in press).

Obwohl Pseudo-Kontingenzen eine schwerwiegende, logisch kaum nachvollziehbare Täuschung widerspiegeln, liefern sie wie viele Illusionen unter bestimmten Randbedingungen recht genaue Urteile. Wenn immer das Zusammentreffen zweier Basisraten-Trends (z.B. meist teure Preise und meist gute Waren) in einer Umgebung (z.B. Supermarkt) eine gemeinsame Ursache hat (z.B. aufwendige Qualitätskontrollen), dann kann man von der Pseudo-Kontingenz tatsächlich als Proxy für eine echte Kontingenz ausgehen. Pseudo-Kontingenzen sind daher ein potentiell sehr nützliches Instrument für induktive Schlüsse in einer Welt, in der korrelierte Merkmale (Preis und Qualität, Erreger und Krankheiten) oftmals nicht in derselben Situation beobachtet werden können.

### **Hypothesentesten in komplexen Umwelten**

#### *Theoretische Fragen*

Viele Probleme des induktiven Denkens und Schließens unter realen Bedingungen sind ihrer Struktur nach noch komplexer als einfache Kontingenzaufgaben. Sie erfordern den Vergleich von zwei oder mehr Kontingenzen, die als alternative Erklärungen eines komplexen Problems herangezogen werden können. Ein solches mehrfaktorielles Kausalitätsproblem beispielsweise liegt dann vor, wenn ein Effekt (Verkehrsunfall) nicht nur mit einer möglichen Ursache kovariiert (Geschwindigkeit des Fahrzeuges), sondern gleichzeitig auch mit anderen möglichen Ursachen (Straßenverhältnisse; technischer Zustand des Fahrzeuges; Ablenkungen des Fahrers). Die Lösung des induktiven Problems besteht darin, die Kontingenzen zwischen dem Effekt und den verschiedenen Ursachen zu bewerten und zu vergleichen. Ähnlich verlangen diagnostische Probleme, die Kontingenzen von Symptomen mit verschiedenen Krankheitskategorien abzuwägen und zu vergleichen. Risiko- und Investitionsentscheidungen in Politik und Wirtschaft erfordern es, neben der Kontingenz zwischen Maßnahmen und den

beabsichtigten Nutzen/Kosten auch die Kontingenzen mit möglichen Nebenwirkungen zu berücksichtigen.

*Vereinfachung mehrfaktorieller Probleme: Discounting.* Eine zwar trügerische, aber sowohl im alltäglichen Denken wie in der Wissenschaft weit verbreitete Strategie im Umgang mit komplexen Problemen ist Vereinfachung. Obwohl sich mehre Ursachen zur Erklärung eines multikausalen Problems anbieten, wird ein monokausales (oder zumindest minderkausales) Modell gesucht, um Entscheidungen treffen und handeln zu können. Diese Tendenz zur Vereinfachung kausaler Strukturen – man könnte sie auch bezeichnen als bevorzugte Suche nach einer einzigen wahren Ursache statt nach einem Ursachengeflecht – ist unter dem Begriff „Discounting“ wohl bekannt und gilt in der Attributionforschung gar als plausibles Prinzip des kausalen Schließens. Discounting besagt: Der Beitrag einer bestimmten möglichen Ursache zur Erklärung eines Effektes nimmt ab, wenn eine zweite mögliche Ursache bekannt wird, die ebenfalls den Effekt vorhersagt (Downing, Sternberg & Ross, 1985; Kelley, 1973; McClure, 1998; Morris & Larrick, 1995; Schustack & Sternberg, 1981). Wenn ein Sündenbock für ein Unglück gefunden ist, können andere mögliche Sünder ungestraft davon kommen. Aber nicht nur naives Denken und Aberglaube folgen diesem Prinzip, sondern wie gesagt auch die (psychologische) Wissenschaft, wenn etwa angenommen wird, dass ein Befund entweder die eine oder die andere Theorie bestätigt, obwohl sich die Theorien gar nicht logisch ausschließen. In dem Masse wie solche Discounting-Effekte auftreten – das heißt, die Vernachlässigung weiterer Ursachen, wenn primäre Ursachen schon bekannt sind – werden Kausalprobleme notwendigerweise vereinfacht.

*Auflösung mehrfaktorieller Probleme.* Statt eine multikausale Problemstruktur einfach zu ignorieren oder (durch Discounting von alternativen Ursachen) auf vereinfachende, monokausale Modelle zu reduzieren, kann man auch versuchen, das Zusammenwirken verschiedener Faktoren aufzulösen. Dies kann sich besonders dann als sehr schwierig erweisen, wenn zwei Faktoren nicht unabhängig, sondern miteinander korreliert sind, so dass

die Korrelation eines Effekts mit einer Ursache durch die Hinzunahme einer zweiten, konfundierten Ursache alternativ erklärt und als unechte Korrelation entlarvt werden kann. Induktiv-statistische Probleme mit der Problemstruktur einer unechten Korrelation (spurious correlation) werden auch als *Simpson's (1950) Paradox* bezeichnet. Obwohl zwischen zwei Variablen eine klare Kontingenz besteht, verschwindet diese oder wird sogar ins Gegenteil verkehrt, wenn eine Drittvariable berücksichtigt wird.

Das "klassische" Beispiel hierfür betrifft die scheinbare Bevorzugung männlicher Bewerber auf ein Graduiertenstudium. Obwohl es in der Tat zutrifft, dass mehr weibliche Bewerberinnen als männliche Bewerber abgelehnt werden, ergibt sich ein umgekehrtes Bild, wenn man den Unterschied zwischen zwei Graduiertenprogrammen mit in Betracht zieht (vgl. Abbildung 4). Nun stellt sich heraus, dass innerhalb beider Programme eigentlich die weiblichen Bewerberinnen besser abschneiden. Die Aufklärung des Paradoxons liegt also in der Erkenntnis, dass sich weitaus mehr weibliche Bewerberinnen auf das anspruchsvollere Graduiertenprogramm bewerben, in dem absolut gesehen viel mehr Bewerber abgelehnt werden. Die relative Ablehnungsrate liegt jedoch in beiden Programmen für Frauen niedriger als für Männer. Die Lösung eines solchen Problems entspricht dem statistischen Verfahren der Partialkorrelation bzw. der Kovarianzanalyse: Die Korrelation zwischen (weiblichem) Geschlecht und Ablehnung wird aufgehoben und sogar leicht umgekehrt, wenn der Einfluss einer dritten Variablen, Graduiertenprogramme, auspartialisiert wird.

Solche induktiv-statistischen Probleme sind typisch für viele Kontroversen, die sich um unterschiedliche Interpretationen desselben Sachverhalts drehen. Was besagt die Tatsache, dass mehr weibliche Bewerberinnen abgelehnt werden? Handelt es sich um sexistische Diskrimination oder gibt es eine andere Erklärung für denselben Sachverhalt, der selbst nicht in Frage gestellt wird? Ähnliche Probleme mit derselben unterliegenden Struktur lauten etwa: Zeugen die unterdurchschnittlichen PISA Leistungen deutscher Schüler von schlechtem Unterricht in deutschen Schulen oder verschwinden die Leistungsunterschiede, wenn die

ungünstige Zusammensetzung deutscher Schulklassen auspartialisiert wird? Spiegelt das gute oder schlechte Abschneiden eines Sportlers seine tatsächliche Form wider oder aber die Gunst der situativen Umstände?

Wichtig für das Verständnis dieser induktiven Denkprobleme ist die Einsicht, dass eine eindeutig richtige Lösung allein auf statistischem Wege nicht bestimmt werden kann. Es ist logisch nicht gerechtfertigt anzunehmen, dass die Partialkorrelation näher an der Wahrheit ist als die undifferenzierte Ausgangskorrelation. Wenn sich in dem Beispiel des Graduiertenproblems herausstellt, dass das scheinbar schlechtere Abschneiden der weiblichen Bewerberinnen verschwindet, sobald die beiden Programme berücksichtigt werden, so besagt dies keineswegs, dass die neue Interpretation (weibliche Überlegenheit) die richtige ist. Es ist zwar in der Tat möglich, dass Frauen in Wirklichkeit nicht unterlegen sind, sondern sich nur vermehrt auf ein Programm mit einer sehr hohen Zurückweisungsrate bewerben. Doch dieselben Daten sind auch mit der alternativen Interpretation vereinbar, dass das scheinbar schwierigere Programm deshalb eine so hohe Zurückweisungsrate hat, weil sich besonders viele unterlegene Frauen bewerben. Keine der beiden Interpretationen ist a priori korrekt, und allein aufgrund der Korrelationsstatistik kann eine Entscheidung nicht getroffen werden.

Simpson's Paradox ist nur ein spezieller Fall einer größeren Vielfalt von trivariaten Problemen, in denen eine Kontingenz zwischen zwei Variablen, X and Y, durch eine dritte Variable Z moderiert wird (Fiedler & Walther, 2004). Abbildung 5 unterscheidet vier Formen solcher komplexen Probleme gemäß der Beziehung, welche die Drittvariable zu den anderen beiden Variablen aufweist. Angenommen X sei eine unabhängige Variable oder Ursache und Y eine abhängige Variable bzw. ein Effekt, dann kann Z die Rolle einer weiteren orthogonalen Ursache spielen, wenn Z mit X unkorreliert ist. Wenn Z nicht-orthogonal ist, sondern redundant, dann kann Z in der Weise mit X korrelieren, dass die Korrelation zwischen X and Y zum Teil aufgeklärt wird (d.h. dass nach Auspartialisieren von Z die Korrelation geringer wird:  $r_{XY.Z} < r_{XY}$ ). Dies ist der Fall bei der unechten Korrelation im Sinne

von Simpson's Paradox. Oder er kann sein, dass die Korrelation zwischen X und Y ansteigt, wenn die Rolle von Z berücksichtigt wird ( $r_{XY.Z} > r_{XY}$ ). Diesen Fall bezeichnet man als Suppressor-Effekt (Conger & Jackson, 1972), weil bei solchen Problem die Drittvariable Fehlervarianz unterdrückt oder an sich bindet, so dass die Kontingenz zwischen X und Y deutlicher sichtbar wird.

### *Empirische Phänomene*

Obwohl trivariate Strukturen dieser Art repräsentativ sind für viele Probleme in der Realität, steckt die Forschung über die kognitive Verarbeitung solcher Probleme noch in den Anfängen. Die wenigen einschlägigen Studien stimmen weitgehend in der Erkenntnis überein, dass induktive Schlüsse bei derartigen Problemen sehr schwer fallen (Fiedler, Walther, Freytag, & Stryczek, 2002; Schaller & O'Brien, 1992; Waldmann & Hagmayer, 2001). Andererseits besteht aber ebenfalls weitgehend Einigkeit darin, dass diese Schwierigkeit nicht einfach an fehlender kognitiver Kapazität liegt. Vielmehr reicht zumindest bei dichotomen Variablen die kognitive Kapazität aus, um trivariate statistische Beziehungen im Gedächtnis zu repräsentieren und darüber nachdenken und argumentieren zu können (Downing, Sternberg & Ross, 1985; Fiedler & Walther, 2002; Meiser, in press; Spellman et al., 2001). Sogar Tiere im Konditionierungs-Experiment sind in der Lage, korrekt zu erfassen, dass sich die Kontingenz zwischen konditioniertem und unkonditioniertem Reiz mit einem Kontextreiz verändern oder umkehren kann (Kimmel & Gardner, 1981).

*Simpson's Paradox.* Schaller (1992a, 1992b, 1994) und Kollegen (Schaller, Boyd, Yohannes, & O'Brien, 1995; Schaller & O'Brien, 1992). führten eine Serie von Experimenten mit Aufgaben vom Typ des Simpson-Paradox durch. Den Teilnehmern bei Schaller (1992) wurde beispielsweise eine Kontingenz derart präsentiert, dass Frauen schlechtere Arbeitsleistungen zeigen als Männer, wobei sich als Drittvariable anbot zu berücksichtigen, dass Frauen unter ungünstigeren Bedingungen arbeiten müssen als Männer. Ob diese alternative Interpretation der Leistungsunterschiede berücksichtigt wurde oder nicht, hing in

erster Linie von motivationalen Bedingungen ab. Solche Teilnehmer, die ausreichend motiviert waren, eine für Frauen ungünstige Interpretation zu vermeiden, gelangten angesichts der mehrdeutigen Information zu einer relativ positiven Bewertung der Leistung der Frauen. Weniger motivierte Teilnehmer lösten das Problem hingegen im Sinne der scheinbaren globalen Überlegenheit der Männer auf.

Im übrigen konnten Schaller und Kollegen (Schaller, 1992a, 1992b, 1994; Schaller, Boyd, Yohannes & O'Brian, 1995; Schaller & O'Brian, 1992) diese Befunde sowohl mit tabellarisch dargebotenen Statistiken zeigen als auch mit sequentiellen Beobachtungen der Leistungen von Frauen und Männern – oder etwa der Leistung von Tennisspielern mit unterschiedlichen Erfolgsstatistiken von gewonnenen und verlorenen Spielen, die jedoch gegen unterschiedlich schwierige Gegner errungen worden waren (bspw. könnte der bessere von zwei Tennisspielern weniger Spiele gewonnen haben, weil er gegen mehr erstklassige Gegner gespielt hat).

Theoretisch vertreten Schaller und Kollegen die Ansicht, dass im Falle zu geringer Motivation das trivariate Problem kognitiv vereinfacht – also bivariat – repräsentiert wird. Die Annahme lautet, dass nur die Beziehung zwischen den beiden fokalen Variablen (z.B. Geschlecht und Leistung) repräsentiert wird, so als ob die Drittvariable (Arbeitsbedingungen) gar nicht existieren würde. Letztere Variable wird nur dann berücksichtigt, wenn die Motivation dazu stark genug ausgeprägt ist.

Diese Ansicht kann aber im Lichte neuerer Untersuchungen kaum aufrecht erhalten werden (Fiedler et al., 2002; Fiedler et al., 2003; Meiser, in press; Spellman et al., 2001), die beweisen, dass die volle dreidimensionale Verteilung erkannt und recht genau wiedergegeben wird. Das Problem besteht vielmehr darin, mit dem Konflikt umzugehen, dass eine Kontingenz auf verschiedenen Ebenen der Analyse unterschiedlich oder gar gegensätzlich ausfallen kann. Dies konnten Schaller und Kollegen allerdings nicht feststellen, weil sie als abhängige Messung nur den globalen Zusammenhang (z.B. zwischen Geschlecht und

Leistung) erfassten und keine spezifischen Maße für die Wahrnehmung der Partialkorrelation einbezogen.

Die Befunde von Waldmann und Hagmayer (2001) besagen, dass die erfolgreiche Lösung von Simpson's Paradox davon abhängt, ob die Teilnehmer ein geeignetes mentales Modell zur Verfügung haben. Die von diesen Autoren verwendete experimentelle Aufgabe bestand darin, den Einfluss von Bestrahlung ( $X$ ) auf die Qualität von Obst ( $Y$ ) zu erkennen und zu beurteilen, wobei zwei verschiedene Obstsorten ( $Z$ ) zu unterschieden waren. Über beide Obstsorten hinweg führte Bestrahlung zu erhöhter Qualität, aber dieser Vorteil erwies sich als nichtig, sobald der Einfluss auf die beiden Obstsorten getrennt betrachtet wurde. Dann zeigte sich, dass Obstsorte  $Z_1$  im allgemeinen von besserer Qualität war als Obstsorte  $Z_2$ , und der scheinbare Vorteil der Bestrahlung ließ sich darauf reduzieren, dass das bestrahlte Obst hauptsächlich von der Sorte  $Z_1$  war. Die Teilnehmer erkannten dies tendenziell, wenn die Unterscheidung der Obstsorten kausal relevant war (genetisch unterschiedliche Obstsorten), aber sie fanden die Lösung des trivariaten Problems nicht, wenn die Unterscheidung ursächlich sinnlos war (Obst, das in verschiedenen Labors untersucht worden war).

Allerdings stellt die von Waldmann und Hagmayer (2001) aufgezeigte Randbedingung – Vorliegen eines relevanten kausalen Modells – nicht die einzige und nicht unbedingt eine hinreichende Bedingung für die Lösung des induktiven Denkproblems dar. Denn auch wenn ein offenkundiges kausales Modell vorliegt, muss das Paradox keineswegs aufgelöst werden. Dies zeigen neuere Experimente von Fiedler et al. (2003). Obwohl sich ein kausales Modell anbot und durch explizite Instruktion offenkundig gemacht wurde – unterschiedliche Standards zweier Universitäten, um aufzuklären, warum insgesamt mehr weibliche Bewerberinnen abgelehnt werden – und obwohl die bivariaten Zusammenhänge durchweg erkannt wurden – mehr Frauen abgelehnt, mehr Bewerber von Uni A abgelehnt, mehr Frauen

bewerben sich bei Uni A – blieb die durchgängige Illusion erhalten, dass Frauen insgesamt eher abgelehnt werden, und zwar auch innerhalb der beiden Universitäten, A und B.

Tatsächlich waren weibliche Bewerberinnen innerhalb beider Universitäten gleich erfolgreich oder gar leicht erfolgreicher. Aufgehoben wurde die Illusion nur durch eine Manipulation der zeitlichen Abfolge der Stimulusdarbietung in jedem Durchgang. Wenn zuerst die Information über die Universität, wo eine Bewerbung eingegangen ist, dargeboten wurde und dann erst das Bewerber-Geschlecht, dann wurde die erhöhte Ablehnungsquote eher dem Standard der Universität statt dem Geschlecht der BewerberInnen zugeschrieben. Ein solcher Einfluss der zeitlichen Darbietung konnte auch in anderen Experimenten bestätigt werden (Fiedler, Walther, Freytag & Nickel, 2003; Fiedler, Walther, Freytag & Stryczek, 2002).

Andere Anwendungen von Simpson's Paradox finden sich beim induktiven Schließen im Kontext von Organisationen und wirtschaftlichen Entscheidungen (Curley & Browne, 2001), bei der Bewertung der Effizienz von Psychotherapie (Hsu, 1989) und nicht zuletzt auch bei induktiven Schlüssen von Wissenschaftlern, zum Beispiel in der Gedächtnisforschung, wo eine längere Debatte darüber ausgetragen wurde, ob bestimmte Kontingenzen (z.B. zwischen Wiedererkennen und Freier Wiedergabe; Hintzman, 1980, 1993; Tulving & Flexser, 1992) nur durch das Zusammenwerfen von nicht-vergleichbaren Item-Teilmengen entstehen (vgl. auch Flexser, 1991; Nyberg, 1993).

*Kompetition verschiedener Cues.* Sofern bei Simpson's Paradox zwei Faktoren, eine Ursache und eine Kovariate, um die Erklärung eines Effektes streiten, deckt sich die Struktur des Problems weitgehend mit der Struktur der in dem Abschnitt über einfache Kategorisierung behandelten Experimente mit mehr als einem Cue. In der Tat sind die Unterschiede fließend. Während dort die Frage im Mittelpunkt stand, ob Kategorisierungen auf einzelnen oder mehreren Cues beruhen, steht hier der Konflikt im Vordergrund, der entsteht, wenn verschiedene Prediktoren oder Cues um die Erklärung eines Effekts

konkurrieren (Edgell, 1978, 1980; Koehler, 2000). So verliert ein Cue an prediktiver Wirkung, wenn er als Compound mit einem anderen Cue dargeboten wird, der große prediktive Kraft hat und somit die Erklärung eines Effektes an sich bindet (Chapman & Robbins, 1990). Dieser Befund erinnert an das aus der Konditionierung bekannte Phänomen des compound conditioning (Kamin, 1968). Auch das Phänomen des conditioned inhibition lässt sich beim induktiven Schließen wiederfinden, wenn ein Cue negative prediktive Bedeutung bekommt, nachdem er zusammen mit einem positiv-prediktiven Cue dargeboten wird und der erwartete Effekt nicht eintritt, so dass das Ausbleiben des Effektes das zu erklärende Ereignis ist (Chapman & Robbins, 1990). So gibt es im Wettbewerb verschiedener Cues sowohl Befunde von „Discounting“ als auch „Augmentation“. Ersteres bedeutet, dass ein Cue weniger Gewicht bekommt, wenn bereits ein anderer Cue bekannt ist, der das Auftreten eines Effektes ganz oder teilweise erklärt. Letzteres bedeutet, dass das Gewicht eines Cue beim induktiven Schließen zunimmt, wenn ein Effekt trotz des hemmenden Einflusses eines anderen, inhibitorischen Cues auftritt.

*Kooperation verschiedener Cues.* Zwei Cues, die verschiedene Hypothesen zur Erklärung desselben Ereignisses anbieten, verhalten sich nicht immer kompetitiv zueinander. Unter bestimmten Bedingungen kann das Gewicht eines primären Cue sogar zunehmen, wenn ein zweiter Cue, der mit dem ersten redundant ist, dasselbe Ereignis voraussagt. Wie dies möglich ist, zeigt ein Zwei-Stufen-Modell der Attribution von Trope (1986), das empirisch mehrfach bestätigt wurde. Wurde beispielsweise beobachtet, dass ein bestimmtes Individuum ein aggressives Verhalten zeigt, dann bietet sich als primäre Erklärung eine entsprechende Disposition bzw. ein Persönlichkeitsmerkmal (feindselig) an. Diese vorherrschende Tendenz, zur Erklärung von Verhalten auf interne Dispositionen zu schließen, ist gemeinhin als correspondence bias (Gilbert & Malone, 1995; Jones & Davis, 1965) oder als fundamentale Attributionstendenz (Ross, 1977) bekannt. Wird nun außerdem bekannt, dass die beobachtete Aggression in einer typisch aggressiven (z.B. tumultartigen, emotional aufgeheizten)

Situation stattfand, dann sollte der primäre Personen-Cue eigentlich durch den hinzu gekommenen Situations-Cue, der eine alternative Interpretation anbietet, an Bedeutung verlieren.

Wie Trope (1986) und zeigen konnte, kann unter Umständen jedoch das Gegenteil eintreten: In einer typisch aggressiven Situation wird erst richtig deutlich, dass es sich um Aggression (statt um einen anderen Aspekt, wie z.B. Ungeschicklichkeit oder Spaß) gehandelt hat, so dass infolge des zweiten Cues die Kategorisierung der Person als aggressiv noch zusätzlich aufgewertet wird. Nach dem Modell geschieht dies in einer ersten, frühen Phase der Identifikation des Verhaltens, das noch nicht eindeutig als Aggression erkannt ist. In einer frühen Phase des Schlussprozesses, besonders wenn die Problemsituation ambivalent ist, kann ein zusätzlicher, redundanter Cue die erlebte Bedeutung eines anderen, primären Cue unterstützen. In einer späteren Phase des kompetitiven Hypothesentestens, kehrt sich die Art der Cue-Interaktion indessen um. Dann stellt sich die Frage, welche von zwei rivalisierenden Hypothesen – eine feindselige internale Disposition der Person oder aber externalen Umstände der Situation – die angemessenere Erklärung darstellt. Und in dieser kompetitiven Phase führt ein sekundärer Cue dazu, die Bedeutung des primären Cues in Frage zu stellen und abzuschwächen.

Solch kooperatives Zusammenwirken zweier redundanter Cues ist besonders dann von funktionellem Wert, wenn die Aufgabe komplex und die Gedächtnislast sehr groß ist, so dass Redundanz gefragt ist, um mit der großen Menge an Informationen umgehen zu können. Eine recht komplexe Aufgabe wurde etwa in einer Untersuchung von Fiedler, Walther, und Nickel (1999) verwendet, in der die Teilnehmer die Rolle eines Varieté-Direktors zu spielen hatten und die Leistung (Erfolg vs. Misserfolg) von fünf Entertainern bei fünf verschiedenen Disziplinen erfassen und vergleichend bewerten sollten. Die Leistung war stets mit Personen korreliert, indem die fünf Entertainer bei einer, zwei, drei, vier bzw. fünf Disziplinen erfolgreich waren. In zwei verschiedenen experimentellen Bedingungen wurde die Rolle des

zweiten Cue, Disziplinen, manipuliert. In der nicht-redundanten Bedingung bestand zwischen Leistung und Disziplinen keine Korrelation, da in allen fünf Disziplinen die von drei Entertainern gut war und die Leistung von zweien schlecht. In der redundanten Bedingung kovarierte die Leistung indessen ebenfalls mit dem zweiten Cue, so dass die Leistung in einer, zwei, drei, vier bzw. fünf Disziplinen gut ausfiel. In dieser Bedingung bot sich grundsätzlich die alternative Interpretation an, dass die Leistungsunterschiede nicht durch Personen, sondern durch die Disziplinen verursacht werden.

Tatsächlich wurde der Unterschied zwischen den fünf Entertainern in der letzteren Bedingung keineswegs geringer gesehen, sondern noch verstärkt. Das heißt, obwohl die Leistung mit einer alternativen Ursache korrelierte, wurde die Korrelation der Leistung mit den Entertainern deutlicher erkannt. Denn die redundante Bedingung war nur durch ein überaus konsistentes Muster möglich, welches die Komplexität der  $5(\text{Entertainer}) \times 5(\text{Disziplinen}) \times 2$  (gute vs. schlechte Leistung) Kontingenz entscheidend reduziert: Der beste Entertainer war bei allen Disziplinen erfolgreich, der zweitbeste bei allen vier außer der schwersten Disziplin und so weiter bis zum schlechtesten Entertainer, der nur bei der leichtesten Disziplin erfolgreich war. Ein solches Muster bildet sowohl die Fähigkeit der Entertainer wie auch die Schwierigkeit der Disziplinen auf derselben (Guttman-)Skala ab. Hieran sollte deutlich werden, warum Cue-Redundanz sich dann nicht kompetitiv, sondern kooperativ auswirkt, wenn komplexe induktive Probleme nur dadurch gelöst werden können, dass eine vereinfachende oder erhellende innere Struktur entdeckt wird.

*Additive Kombination verschiedener Einflüsse.* Während bei den bisher referierten Untersuchungen zwei oder mehr Ursachen in multiplikativer Weise interagieren, stellt die Support Theory (Tversky & Koehler, 1994) einen Theorierahmen dar, in dem das additive Zusammenwirken verschiedener Faktoren untersucht wird. Angenommen, es soll die Wahrscheinlichkeit ermittelt werden, dass ein europäisches Land die nächste Fussball-Weltmeisterschaft gewinnt. Hierfür bieten sich verschiedene Einzelländer – wie Italien,

Frankreich, Portugal, England, Spanien etc. – die sich additiv zu der Gesamtkategorie europäische Länder zusammensetzen. Verschiedene Hypothesen darüber, welches Land Europa zum Weltmeistertitel verhilft, streiten sich nicht miteinander, sondern ergänzen sich, indem verschiedene – durch ein logisches „oder“ verbundene – Wege zu demselben Ziel führen. Für derartige Probleme, bei denen eine Gesamtkategorie in mehrere (additive) Teilkategorien „aufgespalten“ (Fiedler & Armbruster, 1994) oder „entpackt“ (Rottenstreich & Tversky, 1997; Tversky & Koehler, 1994) werden kann, sagt die Support Theory zutreffend voraus, dass Schätzungen der Gesamt-Wahrscheinlichkeit *sub-additiv* sind. Das heißt, die für Europa insgesamt geschätzte Wahrscheinlichkeit ist geringer als die Summe der Wahrscheinlichkeiten für die einzelnen europäischen Länder, die zusammen Europa ausmachen. Nach Koehler (2000) kann dieser regelmäßig erzielte Befund der Sub-Additivität von Wahrscheinlichkeitsurteilen weitgehend durch die Annahme einer Support-Funktion erklärt werden, die einen linearen Discounting-Einfluss vorhersagt. Support ist eine kognitive Funktion des Ausmaßes von subjektiver Evidenz, die psychologisch eine Kategorisierung unterstützt. Sub-Additivität entsteht, weil der Gesamt-Support durch die gleichzeitige Betrachtung mehrerer Teilkategorien im Sinne eines „Discounting“ abnimmt.

Warum die Teilkategorien sich gegenseitig Support wegnehmen, so dass die Gesamturteile sub-additiv werden, lässt Support Theory weitgehend offen. Eine mögliche Erklärung liefert das generelle Prinzip, dass Wahrscheinlichkeitsurteile unter Unsicherheit regressiv sind (vgl. Abschnitt über einfache Kategorisierung). In dem Maße, wie Urteile ungenau und fehlerbehaftet sind, werden große Wahrscheinlichkeiten unterschätzt und kleine überschätzt. Und das Ausmaß dieser Regression steigt mit der Extremität an (d.h., ist besonders groß für extrem wahrscheinliche und unwahrscheinliche Ereignisse). Da durch Aufspalten einer Kategorie in der Regel kleinere Teilkategorien entstehen, die für sich genommen weniger wahrscheinlich sind, werden sie infolge Regression eher überschätzt als die große Gesamtkategorie, für die der Regressionseffekt eher eine relative Unterschätzung

bedeutet. Eine solche Regressions-Erklärung der Sub-additivität (Fiedler, 2002; Fiedler & Armbruster, 1994; Parducci & Wedell, 1986) ist mit vielen bekannten empirischen Befunden vereinbar und hat darüber hinaus eine ganze Reihe von neuen empirisch prüfbareren Implikationen. Vor allem sollte die Sub-Additivität verstärkt werden, wenn Wissen abnimmt und Unsicherheit zunimmt, weil dann das Ausmaß der Regression ebenfalls zunimmt.

*Induktive Schlüsse bei komplexen realistischen Problemen.*

Viele realistische Probleme erreichen einen noch höheren Grad an Komplexität als bisher in diesem Abschnitt behandelt. Es müssen oft mehr als zwei oder drei Dimensionen gleichzeitig bedacht werden, und innerhalb der Dimensionen müssen oft viele qualitative oder quantitative Abstufungen unterschieden werden. Ein Lehrer in der Grundschule oder in der Sekundarstufe beispielsweise, muss auf induktivem Wege herausfinden, welche Leistungsstufe viele verschiedene Schüler in mehreren Fächern kennzeichnen, und er/sie muss dabei sogar verschiedene Aspekte der Leistung wie Motivation, Wissen, Kreativität, Genauigkeit etc. unterscheiden. Ein Makler oder Anleger an der Börse muss die Kurse von vielfältigen Aktien, geschachtelt nach Ländern und Portfolios, über viele Zeitpunkte hinweg beobachten und mit betrieblichen und konjunkturellen Indikatoren in Beziehung setzen. Oder ein junger Mensch, der eine Berufs- oder Laufbahn-Entscheidung trifft, muss sich über die Vorteile und Nachteile von vielen Berufen in zahlreichen Dimensionen (Interesse, Einkommen, Freizeit, persönliche Entfaltung, Kollegialität) auf einmal kundig machen.

Es überrascht nicht, dass bisher nur wenige kontrollierte Untersuchungen über induktives Schließen bei solch komplexen Problemen vorliegen – allein wegen des großen Aufwandes und der Schwierigkeit, solche Probleme im Rahmen einzelner experimenteller Sitzungen zu bearbeiten. Die wenigen Untersuchungen, die in der Literatur zu finden sind, zeugen indessen von einer ganz erstaunlichen Leistung bei derartig komplexen Urteilen und Entscheidungen. Die Genauigkeit der Leistungsbewertung von Lehrern erwies sich in mehreren Studien als beeindruckend genau (Fiedler, Walther, Freytag & Plessner, 2002;

Jussim & Eccles, 1992). Im allgemeinen erreichen komplexe diagnostische Erkenntnisse im allgemeinen in kürzester Zeit mit einem Minimum an Daten erstaunliche Genauigkeit (Ambady & Rosenthal, 1992).

Trotz dieses optimistischen Befunds bleibt genügend Spielraum für Fehler und Illusionen, besonders weil die Nebenwirkungen von Entscheidungen in komplexen Systemen oft nicht ausreichend berücksichtigt werden (Funke, 1999). Eine ausführlichere Darstellung dieses zunehmend an Bedeutung gewinnenden Forschungsfeldes findet sich in Kapitel xx zum komplexen Problemlösen.

Wie können Menschen mit einer solchen Komplexität umgehen? Wie können sie versuchen, die Komplexität zu reduzieren und durch Einsatz nützlicher Heuristiken dennoch angepasste und korrekte Entscheidungen zu treffen? – Wie schon betont, existieren zu wenige unmittelbar relevante empirische Erkenntnisse, um eine informierte Antwort geben zu können. Wenn man jedoch die wenigen verfügbaren Befunde betrachtet und die Erkenntnisse von weniger komplexen Paradigmen extrapoliert, dann hat die momentane Forschung folgende Antwortalternativen anzubieten.

*Vereinfachende additive Schlussregeln.* Eine Möglichkeit besteht darin, vereinfachende lineare Näherungsmodelle auf komplexe, nicht-lineare Probleme anzuwenden. Seit langem ist bekannt, dass lineare Näherungen in vielen Kontexten recht erfolgreich sein können (Dawes, 1979). Obwohl beispielsweise ein Experte, der eine Prognose abgibt, viele Wechselwirkungen zwischen den beteiligten Cues berücksichtigen müsste, liefert ein einfaches lineares Modell, bei dem jeder Cue gemäß seiner Einzelkorrelation mit der Zielvariablen gewichtet wird, meist eine erstaunlich genaue Vorhersage. Dies ist selbst dann der Fall, wenn alle Cues gleich gewichtet werden und nur das richtige Vorzeichen berücksichtigt wird (vgl. Dawes, 1979; Gigerenzer, Todd & the ABC Group, 1999). Vereinfachende lineare oder monotone Näherungen sind nicht nur meist recht effektiv, sondern haben darüber hinaus den

pragmatischen Vorteil, kommunizierbar (Coombs & Avrunin, 1977) und leicht erlernbar zu sein (Brehmer, 1974; Slovic & Lichtenstein, 1971).

*Analogien und konfigurale Schlussregeln.* Eine andere Möglichkeit der Bewältigung komplexer induktiver Probleme besteht darin, differenzierte Muster von Informationen wie in der Wahrnehmung oder bei der Mustererkennung zu erfassen. Anstatt den Versuch zu unternehmen, Regeln zu abstrahieren, besteht die Strategie hier im Erfassen von Analogien und Gestalten (Gentner, 1983; Gick & Holyoak, 1980). Diese Art der ganzheitlichen, konfiguralen Informationsverarbeitung ist am ehesten im Zusammenhang mit Hypergedächtnis bei Schachspielern untersucht worden, die sich auf einen einzigen Blick hin ganze Schachbrett-Konstellationen merken können. Vielleicht gehen Lehrer, Börsenmakler oder Fußballtrainer mit der Komplexität ihres Berufes ähnlich um wie ein Schachgroßmeister, der sich viele räumliche Konfigurationen auf einmal merkt. In jedem Falle werden Analogieschlüsse bei derartigen Lösungen induktiver Probleme eine bedeutende Rolle spielen (vgl. Kapitel X).

*Reduktion der Komplexität.* Eine weitere Möglichkeit besteht darin, Informationen auf einem so hohen Niveau der Aggregation zu betrachten, dass die Komplexität drastisch reduziert wird. Dies ist nebenbei bemerkt die vorherrschende Strategie in den empirischen Sozialwissenschaften, wo man die Vielfalt idiomatischer Gesetze durch einfachere universelle Gesetze annähert, die durch Mittelung oder Aggregation über viele individuelle Versuchsteilnehmer hinweg entstehen. Für einen typischen Versuchsplan mit je zwei Stufen auf jeder von zwei unabhängigen Variablen, aber 50 Stufen auf dem Faktor Versuchspersonen bedeutet dies eine immense Vereinfachung.

Ein von der Methodik der Psychologie und ihrer Nachbarwissenschaften weitgehend vernachlässigtes Problem besteht darin, dass die auf diesem Wege erzielten universellen Gesetze oft grundverschieden sind von denjenigen, die auf der Ebene einzelner Individuen gelten. Ein eindrucksvolles Beispiel bietet das aus der Soziologie lange bekannte Phänomen

der ökologischen Korrelation (Hammond, 1973; Hannon, 1970; Robinson, 1950). Eine Korrelation, die auf der Ebene individueller Personen vernachlässigbar gering ist (wie die Korrelation von  $r = .20$  zwischen Hautfarbe und Lesenkoennen; cf. Robinson, 1950) kann durch Aggregation inflationär hoch erscheinen ( $r = .91$  zwischen dem Anteil Schwarzer mit dem Anteil Analphabeten ueber 9 grosse Umfragegebiete der USA). Die im Abschnitt über Kontingenzschlüsse behandelte Pseudo-Kontingenz-Täuschung (Fiedler & Freytag, 2004) stellt einen der ökologischen Korrelation entsprechenden kognitiven Prozess dar, bei dem nicht die Korrelation der Einzelwerte, sondern die Beziehung der Basisraten in verschiedenen Gruppen oder Kategorien zugrunde gelegt wird.

*Top-Down-Einflüsse von Theorien und Wissensstrukturen.* Schließlich besteht die wohl wichtigste, fast überall verwendete Methode, mit Komplexität umzugehen, darin, bewährte Schemata oder Theorien auf neue Probleme anzuwenden. Solche Top-Down-Strategien sind der eigentliche Gradmesser für intelligentes Handeln, das nicht allein von den aktuellen Stimuli gesteuert wird, sondern auch bewährtes Wissen und frühere Lernprozesse heranzieht. Allerdings markieren diese Top-Down-Einflüsse die Schnittstelle zwischen den eigentlichen induktiven Bottom-up-Prozessen und den deduktiven Prozessen, die in Kapitel X eingehender behandelt werden. Insofern markieren sie auch ein natürliches Ende dieses Kapitels, das sich auf induktiv-statistisches Denken und Schließen beschränkt.

#### ABSCHLIESSENDE BEMERKUNG

Grundsätzlich stimmen Wissenschaftstheoretiker und empirische Forscher in dem Schluss überein, dass Experten und auch Wissenschaftler im großen und ganzen denselben Beschränkungen und Verzerrungen des Denkens und Urteilens unterliegen wie Laien und Versuchspersonen, die an psychologischen Experimenten teilnehmen (Chalmers, 1976; Mitroff, 1974). Statt den dezidierten Versuch zu wagen, ihre eigenen Theorien wirklich streng zu prüfen und zu falsifizieren (Platt, 1964; Popper, 1959), suchen sie einseitig nach

Konfirmation ihrer Hypothesen und unterliegen ähnlichen Verlockungen des Wunschdenkens und oftmals unlogischen Vorgehens wie naive Menschen (Greenwald, 1975). Insbesondere zeigen Experten die gleichen Illusionen und Fehlertendenzen und folgen ähnlichen kognitiven Prozessen wie Laien im Bereich des induktiv-statistischen Denkens. Aber der wissenschaftliche Ansatz unterscheidet sich in einer kritischen Hinsicht von der naiven Herangehensweise an induktive Probleme. Wissenschaftler haben – obwohl sie in denselben Bahnen denken wie Laien – eine Reihe von Modellen und methodischen Hilfsmitteln entwickelt, die es ihnen ermöglichen, zumindest dann mit den zahlreichen Fallen und Wirrungen des induktiven Denkens umzugehen, wenn die Probleme erkannt werden. Ein vornehmliches Ziel des vorliegenden Kapitels besteht darin, das Bewusstsein für diese Gelegenheiten zu schärfen, wo der gesunde Menschenverstand allein nicht ausreicht, sondern technische und methodische Hilfsmittel notwendig sind, um in einer modernen Informationsgesellschaft entscheiden und handeln zu können.

## Literatur

- Allan, L. G. & Jenkins, H. M. (1983). The effect of representations of binary variables on judgment of influence. *Learning and Motivation*, 14, 381-405.
- Alloy, L.B., & Abramson, L.Y. (1979). Judgment of contingency in depressed and nondepressed students: Sadder but wiser? *Journal of Experimental Psychology: General*, 108, 441-485.
- Alloy, L.B., & Tabachnik, N. (1984). Assessment of covariation by humans and animals: the joint influence of prior expectations and current situational information. *Psychological Review*, 91, 112-149.
- Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, 111, 256-274.
- Ayton, P., & Fischer, I. (in press). The Hot-Hand Fallacy and the Gambler's Fallacy: Two faces of Subjective Randomness? *Memory & Cognition*.
- Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review*, 97, 548-564.
- Becker, G. M, Degroot, M. H., & Maschak, J. (1963). Stochastic models of choice behavior. *Behavioral Science*, 8, 41-55.
- Berndsen, M., Van der Pligt, J., Spears, R., & McGarty, C. (1996). Expectation-based and data-based illusory correlation: the effects of confirming versus disconfirming evidence. *European Journal of Social Psychology*, 17, 899-913.
- Betsch, T., Haberstroh, S., & Hhle, C. (2002). Explaining and predicting routinized decision making: A review of theories. *Theory and Psychology*, 12, 453-488.
- Betsch, T., Hoffmann, K., Hoffrage, U., & Plessner, H. (2003). Intuition beyond recognition: When less familiar events are liked better. *Experimental Psychology*, 50, 49-54.

- Betsch, T., & Pohl, D. (2002). Tversky and Kahneman's availability approach to frequency judgement: A critical analysis. In P. Seldmeier & T. Betsch (Eds.), *ETC. Frequency processing and cognition* (pp. 109-119). London: Oxford University Press.
- Betsch, T., Siebler, F., Marz, P., Hormuth, S., & Dickenberger, D. (1999). The moderating role of category salience and category focus in judgments of set size and frequency of occurrence. *Personality and Social Psychology Bulletin*, 25, 463-481.
- Bodenhausen, G.V. (1990). Stereotypes as judgmental heuristics. Evidence of circadian variations in discrimination. *Psychological Science*, 1, 319-322.
- Bröder, A. (2001). Die eingeschränkte Fruchtbarkeit eines unvollständigen Forschungsprogramms: Kommentar zu Hertwig und Hoffrage (2001). *Psychologische Rundschau*, 52, 159-162.
- Bröder, A. (2003). Decision making with the “adaptive toolbox”: Influence of environmental structure, intelligence, and working memory load. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 611–625.
- Brown, N. R. (1995). Estimation strategies and the judgment of event frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 21, 1539-1553.
- Brunswik, E. (1956). *Perception and the representative design of experiments*. Berkeley: University of California Press.
- Chalmers, A.F. (1976). *What is this thing called science? An assessment of the nature and status of science and its methods*. St. Lucia, Queensland: University of Queensland Press.
- Chapman, L. J., & Chapman, J. P. (1967). Genesis of popular but erroneous diagnostic observations. *Journal of Abnormal Psychology*, 72, 193-204.
- Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, 74, 271-280.

- Chapman, G. B., & Johnson, E. J. (1999). Anchoring, activation, and the construction of values. *Organizational Behavior and Human Decision Processes*, 79, 115-153.
- Chapman, G.B., & Robbins, S.J. (1990). Cue interaction in human contingency judgment. *Memory & Cognition*, 18, 537-545.
- Cheng, P.W. (1997). From covariation to causation: A causal power model. *Psychological Review*, 104, 367-405.
- Combs, B. & Slovic, P. (1979). Causes of death: Biased newspaper coverage and biased judgments. *Journalism Quarterly*, 56, 837-843, 849.
- Conger, A. J. & Jackson, D. N. (1972). Suppressor variables, prediction, and the interpretation of psychological relationships. *Educational & Psychological Measurement*, 32, 579-599.
- Coombs, C. H. & Avrunin, G. S. (1977). Single-peaked functions and the theory of preference. *Psychological Review*, 84, 216-230.
- Christianson, S. (1992). Emotional stress and eyewitness memory: A critical review. *Psychological Bulletin*, 112, 284-309.
- Crocker, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin*, 90, 272-292.
- Curley, S.P., & Browne, G.J. (2001). Normative and descriptive analysis of Simpson's Paradox in decision making. *Organizational Behavior and Human Decision Processes*, 84, 308-333.
- Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, 34, 571-582.
- Downing, C.J., Sternberg, R.J., Ross, B.H. (1985). Multicausal inference: Evaluation of evidence in causally complex situations. *Journal of Experimental Psychology: General*, 114, 239-263.

- Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 249-267). Cambridge: University Press.
- Edgell, S. E. (1978). Configural information processing in two-cue probability learning. *Organizational Behavior & Human Decision Processes*, 22, 404-416.
- Edgell, S. E. (1980). Higher order configural information processing in nonmetric multiple-cue probability learning. *Organizational Behavior & Human Decision Processes*, 25, 1-14.
- Englich, B., & Mussweiler, T. (2001). Sentencing under uncertainty: Anchoring effects in the courtroom. *Journal of Applied Social Psychology*, 31, 1535-1551.
- Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). Simultaneous over- and underconfidence: The role of error in judgment processes. *Psychological Review*, 101, 519-527.
- Evans, J.St.B.T., Clibbens, J., Cattani, A., Harris, A. & Dennis, I. (2003). Explicit and implicit processes in multi-cue judgment. *Memory & Cognition*, 31, 608-618.
- Fiedler, K. (1991). The tricky nature of skewed frequency tables: An information loss account of distinctiveness-based illusory correlations. *Journal of Personality and Social Psychology*, 60, 24-36.
- Fiedler, K. (1993). Kognitive Täuschungen: Faszination eines modernen Forschungsprogramms. In W. Hell, K. Fiedler & G. Gigerenzer (Hrsg.), *Kognitive Täuschungen* (S. 7-12). Heidelberg: Spektrum Akademischer Verlag.
- Fiedler, K. (1996). Explaining and simulating judgment biases as an aggregation phenomenon in probabilistic, multiple-cue environments. *Psychological Review*, 103, 193-214.
- Fiedler, K. (2000a). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological Review*, 107, 659-676.

- Fiedler, K. (2000b). Illusory correlations: A simple associative algorithm provides a convergent account of seemingly divergent paradigms. *Review of General Psychology*, 4, 25-58.
- Fiedler, K. (2002). Frequency judgment and retrieval structures: Splitting, zooming, and merging the units of the empirical world. In P. Sedlmeier & T. Betsch (Eds.), *Frequency processing and cognition* (pp. 67–87). Oxford: Oxford University Press.
- Fiedler, K., & Armbruster, T. (1994). Two halves may be more than one whole: Category-split effects on frequency illusions. *Journal of Personality and Social Psychology*, 66, 633-645.
- Fiedler, K., Brinkmann, B., Betsch, R., & Wild, B. (2000). A sampling approach to biases in conditional probability judgments: Beyond base rate neglect and statistical format. *Journal of Experimental Psychology: General*, 129, 1-20.
- Fiedler, K., & Freytag, P. (2004). Pseudocontingencies. *Journal of Personality & Social Psychology*, 87, 453-467.
- Fiedler, K., Hemmeter, V., & Hofmann, C. (1984). On the origin of illusory correlation. *European Journal of Social Psychology*, 14, 191-201.
- Fiedler, K., Russer, S. & Gramm, K. (1993). Illusory correlations and memory performance. *Journal of Experimental Social Psychology*, 29, 111-136.
- Fiedler, K., Walther, E., Freytag, P., & Nickel, S. (2003). Inductive reasoning and judgment interference: Experiments on Simpson's paradox. *Personality and Social Psychology Bulletin*, 29, 14-27.
- Fiedler, K., Walther, E., Freytag, P., & Plessner, H. (2002). Judgment biases in a simulated classroom — A cognitive-environmental approach. *Organizational Behavior and Human Decision Processes*, 88, 527-561.

- Fiedler, K., Walther, E., Freytag, P., & Stryczek, E. (2002). Playing mating games in foreign cultures: A conceptual framework and an experimental paradigm for inductive trivariate inference. *Journal of Experimental Social Psychology, 38*, 14-30.
- Fiedler, K., Walther, E., Freytag, P., & Nickel, S. (2003) Inductive reasoning and judgment interference. Experiments on the Simpson paradox. *Personality and Social Psychology Bulletin, 29*, 14-18.
- Fiedler, K., Walther, E., & Nickel, S. (1999). The autoverification of social hypotheses: Stereotyping and the power of sample size. *Journal of Personality and Social Psychology, 77*, 5-18.
- Flexser, A.J. (1991). The implications of item differences: Commentary on Hintzman and Hartry. *Journal of Experimental Psychology: Learning, Memory & Cognition, 17*, 338-340.
- Fried, L.S., & Holyoak, K.J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory & Cognition, 10*, 234-257.
- Funke, J. (1999). Komplexes Problemlösen: Ein Blick zurück und nach vorne. *Psychologische Rundschau, 50*, 194-197.
- Furby, L. (1973). Interpreting regression toward the mean in developmental research *Developmental Psychology, 8*, 172-179.
- Galinsky, A. D., & Moskowitz, G. B. (2000). Counterfactuals as behavioral primes: Priming the simulation heuristic and consideration of alternatives. *Journal of Experimental Social Psychology, 36*, 384-409.
- Galinsky, A., & Mussweiler, T. (2001). First offers as anchors: The role of perspective-taking and negotiator focus. *Journal of Personality and Social Psychology, 81*(4), 657-669.

- Galinsky, A., Mussweiler, T., & Medvec, V. H. (2002). Disconnecting outcomes and evaluations: The role of negotiator focus. *Journal of Personality and Social Psychology*, 83(5), 1131-1140.
- Gick, M.L., & Holyoak, K.J. (1980). Analogical problem solving. *Cognitive Psychology*, 12, 306-355.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103, 650-669
- Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102, 684-704.
- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, 98, 506-528.
- Gigerenzer, G. & Murray, D. J. (1987). *Cognition as intuitive statistics*. Hillsdale, NJ: Erlbaum.
- Gigerenzer, G., & Todd, P. and the ABC Group (Eds.). (1999). *Simple heuristics that make us smart*. Oxford: Oxford University Press.
- Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, 117, 21-38.
- Gilovich, T., Griffin, D. & Kahneman, D. (Eds.) (2002). *Heuristics and biases: The psychology of intuitive judgment*. New York: Cambridge University Press.
- Gluck, M. A. & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, 117, 227-247.
- Goldberg, L. R. (1968). Simple models or simple processes? Some research on clinical judgments. *American Psychologist*, 23, 483-496.
- Goldberg, L.R. (1970). Man versus model of man: A rationale, plus some evidence, for a method of improving on clinical inferences. *Psychological Bulletin*, 73, 422-432.

- Goldstein, D. G. & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, 109, 75–90.
- Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, 82, 1-20.
- Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, 109, 3-25.
- Hamilton, D.L. (1981). Illusory correlations as a basis for stereotyping. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 115-144). Hillsdale, N.J.: Lawrence Erlbaum.
- Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology*, 12, 392-407.
- Hamilton, D. L., & Rose, T. (1980). Illusory correlation and the maintenance of stereotypic beliefs. *Journal of Personality and Social Psychology*, 39, 832-845.
- Hamilton, D. L. & Sherman, S. J. (1989). Illusory correlations: Implications for stereotype theory and research. In D. Bar-Tal, C. F. Graumann, A. W. Kruglanski, & W. Stroebe (Eds.). *Stereotype and prejudice: Changing conceptions* (pp. 59-82). New York: Springer.
- Hammond, J.L. (1973). Two sources of error in ecological correlations. *American Sociological Review*, 38, 764-777.
- Hammond, K. R. & Summers, D. A. (1972). Cognitive control. *Psychological Review*, 79, 58-67.
- Hertwig, R. Barron, G., Weber, E.U., & Erev, I. (in press). Risky Prospects: When Valued Through A Window of Sampled Experiences. In Fiedler, K & Juslin, P., *Sampling as a key to understanding adaptive cognition*. Cambridge University Press.

- Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science, 15*, 534-539.
- Hertwig, R. & Gigerenzer, G. (1999). The “conjunction fallacy” revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making, 12*, 275–305.
- Hintzman, D.L. (1980). Simpson’s paradox and the analysis of memory retrieval. *Psychological Review, 87*, 398-410.
- Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers, 16*, 96-101.
- Hoch, S. J. (1984). Availability and interference in predictive judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 649-662.
- Holland, J.H., Holyoak, K.J., Nisbett, R.E., & Thagard, P.R. (1986). *Induction: Processes of inference, learning, and discovery*. Cambridge, MA: MIT Press.
- Holyoak, K.J., & Nisbett, R.E. (1988). Induction. In R.J. Sternberg & E.E. Smith (Eds.), *The psychology of human thought* (pp. 50-91). New York, NY: Cambridge University Press.
- Hursch, C. J., Hammond, K. R. & Hursch, J. L. (1964). Some methodological considerations in multiple-cue probability studies. *Psychological Review, 71*, 42-60
- Hsu, L.M. (1989). Random sampling, randomization, and equivalence of contrasted groups in psychotherapy outcome research. *Journal of Consulting and Clinical Psychology, 57*, 131-137.
- Huttenlocher, J., Hedges, L.V., & Vevea, J.L. (2000) Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General, 129*, 220-249.
- Kao, S. -F. & Wasserman, E. A. (1993). Assessment of an information integration account of contingency judgment with examination of subjective cell importance and method of information presentation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1363-1386.

- Jacoby, L. L., Kelley, C. M., Brown, J., & Jasechko, J. (1989). Becoming famous overnight: Limits on the ability to avoid unconscious influences of the past. *Journal of Personality and Social Psychology*, *56*, 326–338.
- Jacowitz, K. E. & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin*, *21*, 1161-1166.
- Jenkins, H. M. & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs*, *79*(1, Whole No. 594).
- Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on "selective exposure to information". *Journal of Personality and Social Psychology*, *80*, 557-571.
- Jones, E. E. & Davis, K. E. (1965). From acts to dispositions: the attribution process in social psychology. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Volume 2, pp. 219-266). New York: Academic Press.
- Juslin, P. (1994). The overconfidence phenomenon as a consequence of informal experimenter - guided selection of almanac items. *Organizational Behavior and Human Decision Processes*, *57*, 226-246.
- Juslin, P., Jones, S., Olsson, H. & Winman, A. (2003). Cue abstraction and exemplar memory in categorization: Evidence for multiple representation levels. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *29*, 924-941.
- Juslin, P., & Persson, M. (2002). PROBabilities from Exemplars (PROBEX): A “lazy” algorithm for probabilistic inference from generic knowledge. *Cognitive Science*, *26*, 563-607.
- Juslin, P., Winman, A., & Olsson, H. (2000). Naive empiricism and dogmatism in confidence research: A critical examination of the hard-easy effect. *Psychological Review*, *107*, 384-396.

- Jussim, L., & Eccles, J.S. (1992). Teacher expectations II: Construction and reflection of student achievement. *Journal of Personality and Social Psychology*, 63, 947-961.
- Kahneman, D., Slovic, P., & Tversky, A. (Eds.) (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263-291.
- Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201-208). Cambridge, England: Cambridge University Press.
- Kamin, L.J. (1968). "Attention-like" processes in classical conditioning. In M.R. Jones (Ed.), *Miami symposium on the prediction of behavior 1967: Aversive stimulation* (pp. 9-31). Coral Gables, FL: University of Miami Press.
- Kareev, Y. (1995). Through a narrow window: Working memory capacity and the detection of covariation. *Cognition*, 56, 263-269.
- Kareev, Y. (2000). Seven (indeed, plus minus two) and the detection of correlation. *Psychological Review*, 107, 397-402.
- Kareev, Y., Lieberman, I., & Lev, M. (1997). Through a narrow window: Sample size and the perception of correlation. *Journal of Experimental Psychology: General*, 126, 278-287.
- Kim, H., & Baron, R. S. (1988). Exercise and the illusory correlation: Does arousal heighten stereotyping processing? *Journal of Experimental Social Psychology*, 24, 366-380.
- Kimmel, H. D. & Gardner, K. A. (1981). Transswitching as a means of studying within-subjects conditioning effects. *Bulletin of the Psychonomic Society*, 18, 315-317
- Klayman, J., Soll, J. B., González-Vallejo, C. & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational Behavior and Human Decision Processes*, 79, 216-247.

- Koehler, D. J. (1991). Explanation, imagination, and confidence in judgment. *Psychological Bulletin*, 110, 499-519.
- Koehler, D.J. (2000). Probability judgment in three-category classification learning. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 26, 28-52.
- Koehler, D. J., White, C. M., & Grondin, R (2003). An Evidential Support Accumulation Model of subjective probability. *Cognitive Psychology*, 46, 152-197.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108, 480-498.
- Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of subjective probabilities: The state of the art up to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306-334). New York: Cambridge University Press.
- Mackie, D. M., Hamilton, D. L., Schroth, H. A., Carlisle, C. J., Gersho, B. F., Meneses, L. M., Nedler, B. F., & Reichel, L. D. (1989). The effects of induced mood on expectancy-based illusory correlations. *Journal of Experimental Social Psychology*, 25, 524-544.
- Macrae, N. C., Hewstone, M., & Griffiths, R. J. (1993). Processing load and memory for stereotype-based information. *European Journal of Social Psychology*, 23, 77-87.
- Martignon, L., & Laskey, K. B. (1999). Bayesian benchmarks for fast and frugal heuristics. In G. Gigerenzer, P. M. Todd & ABC Research Group (Eds.), *Simple Heuristics That Make Us Smart* (pp. 169-188). London: Oxford University Press.
- Martignon, L., & Hoffrage, U. (1999). Why does one-reason decision making work? A case study in ecological rationality. In G. Gigerenzer, P. M. Todd & ABC Research Group (Eds.), *Simple Heuristics That Make Us Smart* (pp. 119-140). London: Oxford University Press.
- McCauley, C. & Stitt, C. L. (1978). An individual and quantitative measure of stereotypes. *Journal of Personality and Social Psychology*, 36, 929-940.

- McKenzie, C. R. M. (1994). The accuracy of intuitive judgment strategies: Covariation assessment and Bayesian inference. *Cognitive Psychology*, 26, 209-239.
- McReynolds, P. (1989). Diagnosis and clinical assessment: Current status and major issues. *Annual Review of Psychology*, 40, 83-108.
- Medin, D. L., Goldstone, R. L., & Genter, D. (1993). Respects for similarity. *Psychological Review*, 100, 254-278.
- Meiser, T. (in press). Contingency learning and biases group impressions. In K. Fiedler & P. Juslin (Eds.), *In the beginning there is a sample: Information sampling as a key to understand adaptive cognition*. New York: Cambridge University Press.
- Meiser, T., & Bröder, A. (2002). Memory for multidimensional source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 116-137.
- Miller, D. T., & Turnbull, W. (1986). Expectancies and interpersonal processes. *Annual Review of Psychology*, 37, 233-256.
- Mitroff, I. I. (1974). *The subjective side of science*. Amsterdam: Elsevier.
- Mullen, B. & Johnson, C. (1990). Distinctiveness-based illusory correlations and stereotyping: A meta-analytic integration. *British Journal of Social Psychology*, 29, 11-28.
- Mussweiler, T. (2003). Comparison Processes in Social Judgment: Mechanisms and Consequences. *Psychological Review*, 110, 472-489.
- Mussweiler, T., & Strack, F. (1999a). Comparing is believing: A selective accessibility model of judgmental anchoring. In W. Stroebe & M. Hewstone (Eds.), *European Review of Social Psychology* (Vol. 10, pp. 135-168). Chichester, England: Wiley and Sons, Inc.
- Mussweiler, T., & Strack, F. (1999b). Hypothesis-consistent testing and semantic priming in the anchoring paradigm: A selective accessibility model. *Journal of Experimental Social Psychology*, 35, 136-164.

- Mussweiler, T., & Strack, F. (2000a). The “relative self”: Informational and judgmental consequences of comparative self-evaluation. *Journal of Personality and Social Psychology, 79*(1), 23-38.
- Mussweiler, T., & Strack, F. (2000b). The use of category and exemplar knowledge in the solution of anchoring tasks. *Journal of Personality and Social Psychology, 78*, 1038-1052.
- Mussweiler, T., Strack, F., & Pfeiffer, T. (2000). Overcoming the inevitable anchoring effect: Considering the opposite compensates for selective accessibility. *Personality and Social Psychology Bulletin, 26*(9), 1142-1150.
- Newman, J., Wolff, W. T. & Hearst, E. (1980). The feature-positive effect in adult human subjects. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 630-650.
- Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review, 90*, 339-363.
- Northcraft, G. B. & Neale, M. A. (1987). Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions. *Organizational Behavior and Human Decision Processes, 39*, 84-97.
- Nyberg, L. (1993). More on Simpson’s paradox and the analysis of memory retrieval. *Bulletin of the Psychonomic Society, 31*, 326-328.
- Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review, 72*, 407-418.
- Parducci, A. & Wedell, D. H. (1986) The category effect with rating scales: Number of categories, number of stimuli, and method of presentation. *Journal of Experimental Psychology: Human Perception & Performance, 12*, 496-516.
- Petzold, P. & Georgieva, P. (1994). Kontrast und Assimilation - sich ausschließende oder sich ergänzende Prozesse? *Zeitschrift für Psychologie, 202*, 21-36.

- Platt, J.R. (1964). Strong inference. *Science*, 146, 347-353.
- Plous, S. (1987). Thinking the unthinkable: The effects of anchoring on likelihood estimates of nuclear war. *Journal of Applied Social Psychology*, 19, S. 67-91.
- Plessner, H., Freytag, P., & Fiedler, K. (2000). Expectancy-effects without expectancies: Illusory correlations based on cue-overlap. *European Journal of Social Psychology*, 30, 837-851.
- Plessner, H., Hartmann, C., Hohmann, N., & Zimmermann, I. (2001). Achtung Stichprobe! Der Einfluss der Informationsgewinnung auf die Bewertung sportlicher Leistungen. *Psychologie & Sport*, 8, 91-100.
- Popper, K.R. (1959). *The logic of scientific discovery*. London: Hutchinson.
- Pyszczynski, T. & Greenberg, J. (1987) Toward an integration of cognitive and motivational perspectives on social inference: A biased hypothesis-testing model. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 20, pp. 297-340). San Diego, CA: Academic Press.
- Reyes, R. M., Thompson, W. C., & Bower, G. H. (1980). Judgmental biases resulting from differing availabilities of arguments. *Journal of Personality and Social Psychology*, 39, 2-12.
- Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15, 351-357.
- Roe, R.M., Busemeyer, J.R., & Townsend, J.T. (2001). Multi-alternative decision field theory: A dynamic connectionist model of decision making. *Psychological Review*, 108, 370-392.
- Rosenthal, R. & Jacobson, L. (1968). *Pygmalion in the classroom: Teacher expectations and student intellectual development*. New York: Holt.
- Ross, L. (1977). The intuitive psychologist and his shortcomings. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 10, pp. 174-221). New York: Academic Press.

- Ross, M. & Sicoly, F. (1979). Egocentric biases in availability and attribution. *Journal of Personality & Social Psychology*, 37, 322-336
- Rothman, J. A., & Schwarz, N. (1998). Constructing perceptions of vulnerability: Personal relevance and the use of experiential information in health judgments. *Personality and Social Psychology Bulletin*, 24, 1053-1064.
- Rottenstreich, Y., & Tversky, A. (1997). Unpacking, repacking, and anchoring: Advances in support theory. *Psychological Review*, 104, 406-415.
- Ruder, M. & Bless, H. (2003). Mood and the reliance on the ease of retrieval heuristic. *Journal of Personality and Social Psychology*, 85, 20-32.
- Schaller, M. (1992a). In-group favoritism and statistical reasoning in social inference: Implications for formation and maintenance of group stereotypes. *Journal of Personality and Social Psychology*, 63, 61-74.
- Schaller, M. (1992b). Sample size, aggregation, and statistical reasoning in social inference. *Journal of Experimental Social Psychology*, 28, 65-85.
- Schaller, M. (1994). The role of statistical reasoning in the formation, preservation and prevention of group stereotypes. *British Journal of Social Psychology*, 33, 47-61.
- Schaller, M., Boyd, C., Yohannes, J., & O'Brien, M. (1995). The prejudiced personality revisited: Personal need for structure and formation of erroneous group stereotypes. *Journal of Personality and Social Psychology*, 68, 544-555.
- Schaller, M., & O'Brien, M. (1992). "Intuitive analysis of covariance" and group stereotype formation. *Personality and Social Psychology Bulletin*, 18, 776-785.
- Schulz-Hardt, S., Frey, D., Lüthgens, C., & Moscovici, S. (2000). Biased information search in group decision making. *Journal of Personality and Social Psychology*, 78, 655-669.
- Schustack, M.W., & Sternberg, R.J. (1981). Evaluation of evidence in causal inference. *Journal of Experimental Psychology: General*, 110, 101-120.

- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, 54, 93-105.
- Schwarz, N. & Bless, H. (1992). Constructing reality and its alternatives: An inclusion/exclusion model of assimilation and contrast effects in social judgment. In H. Martin & A. Tesser (Eds.). *The construction of social judgment* (pp. 217-245). Hillsdale, NJ: Erlbaum.
- Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology*, 61, 195-202.
- Schwarz, N., Strack, F., Hilton, D., & Naderer, G. (1991) Base rates, representativeness, and the logic of conversation: The contextual relevance of "irrelevant" information. *Social Cognition*, 9, 67-84.
- Sedlmeier, P. (1999). *Improving statistical reasoning: Theoretical models and practical implications*. Mahwah, NJ: Erlbaum.
- Sedlmeier, P. (2002). Associative learning and frequency judgments: The PASS model. In P. Sedlmeier & T. Betsch (Eds.), *Etc. Frequency processing and cognition* (pp. 137-152). Oxford: Oxford University Press.
- Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology*, 41, 1-19.
- Simon, H.A., & Kotovsky, K. (1963). Human acquisition of concepts for sequential patterns. *Psychological Review*, 70, 534-546.
- Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society, Ser. B*, 13, 238-241.
- Singer, J. L. (1990) (Ed.). *Repression and dissociation: Implications for personality theory, psychopathology, and health* (pp. 387-403). Chicago: University of Chicago Press.
- Slovic, P. (1995). The construction of preference. *American Psychologist*, 50, 364-371.

- Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavioral and Human Performance*, 6, 649-744.
- Soll, J. B., & Klayman, J. (in press). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Spears, R., Eiser, J. R., & Van der Pligt, J. (1987). Further evidence for expectation-based illusory correlations. *European Journal of Social Psychology*, 17, 253-258.
- Spellman, B. A., Price, C. M., & Logan, J. M. (2001). How two causes are different from one: The use of (un)conditional information in Simpson's Paradox. *Memory & Cognition*, 29, 193-208.
- Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures. *Psychological Bulletin*, 111, 42-61.
- Stewart, N., Chater, N., Stott, H. P., & Reimers, S. (2003). Prospect relativity: How choice options influence decision under risk. *Journal of Experimental Psychology: General*, 132, 23-46.
- Strack, F. & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social Psychology*, 73, 437-446.
- Swets, J., Dawes, R.M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, 1, Whole No. 1.
- Taylor, S. E. & Fiske, S. T. (1978). Salience, attention, and attribution: Top of the head phenomena. In L. Berkowitz (Ed.). *Advances in experimental social psychology* (Vol. 11, pp. 249-288). New York: Academic Press.
- Trope, Y. (1986). Identification and inference processes in dispositional attribution. *Psychological Review*, 93, 239-257.

- Trope, Y. & Liberman, A. (1996). Social hypothesis testing: Cognitive and motivational factors. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 239-270). New York: Guilford Press.
- Tulving, E., & Flexser, A. J. (1992) On the nature of the Tulving-Wiseman function *Psychological Review*, 99, 543-546.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327-352.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207-232.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211, 453-458
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90, 293-315.
- Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review*, 101, 547-567.
- von Restorff, H. (1933). Über die Wirkung von Bereichsbildungen im Spurenfeld. *Psychologische Forschung*, 18, 299-342.
- Wänke, M., Schwarz, N., & Bless, H. (1995). The availability heuristic revisited: Experienced ease of retrieval in mundane frequency estimates. *Acta Psychologica*, 89, 83-90.
- Wänke, M., Bless, H., & Biller, B. (1996). Subjective experience versus content of information in the construction of attitude judgments. *Personality and Social Psychology Bulletin*, 22, 1105-1113.
- Wänke, M., Bohner, G., & Jurkowitsch, A. (1997). There are many reasons to drive a BMW – Surely you know one: Ease of argument generation influences brand attitudes. *Journal of Consumer Research*, 24, 70-77.

- Waldmann, M. R., & Hagmayer, Y. (2001). Estimating causal strength: The role of structural knowledge and processing effort. *Cognition*, *82*, 27-58.
- Wasserman, E. A., Dorner, W. W., & Kao, S. F. (1990). Contributions of specific cell information to judgments of inter-event contingency. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *16*, 509-521.
- Wasserman, E. A., Elek, S. M., Chatlosh, D. L. & Baker, A. G. (1993). Rating causal relations: Role of probability in judgments of response-outcome contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 174-188.
- Weinstein, N.D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, *39*, 806-820.
- Wells, G. L. & Gavanski, I. (1989). Mental simulation of causality. *Journal of Personality and Social Psychology*, *56*, 161-169.
- Wilson, T. D., Houston, C., Etling, K. M. & Brekke, N. (1996). A new look at anchoring effects: Basic anchoring and its antecedents. *Journal of Experimental Psychology: General*, *4*, 387-402.
- Winman, A., & Juslin, P. (in press). "I'm m/n Confident that I'm Correct": Confidence in Foresight and Hindsight as a Sampling Probability. In Fiedler, K & Juslin, P., *Sampling as a key to understanding adaptive cognition*. Cambridge University Press.
- Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. *Advances in Experimental Social Psychology*, *44*, 1-57.

## Abbildungen

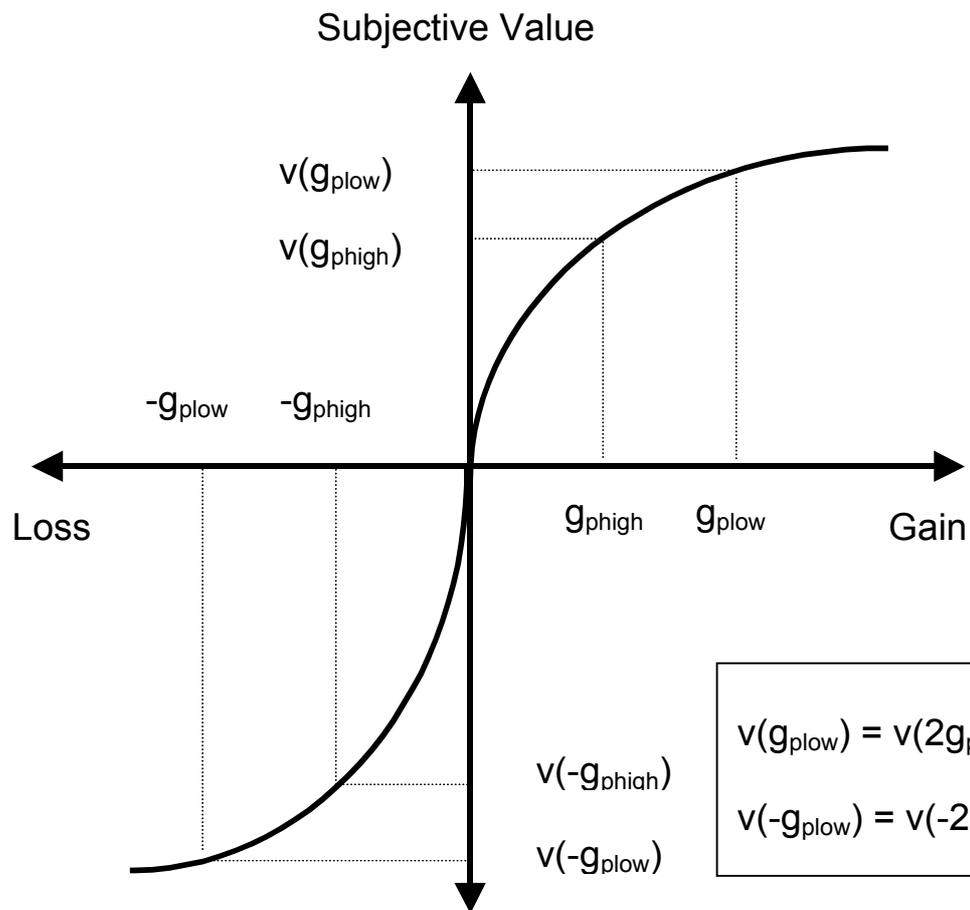
*Abbildung 1.* Die S-foermige Wertefunktion in Prospect Theory. Der subjektive Wert eines verdoppelten eines Gewinnes (g) oder Verlustes (v) ist weniger als doppelt so hoch.

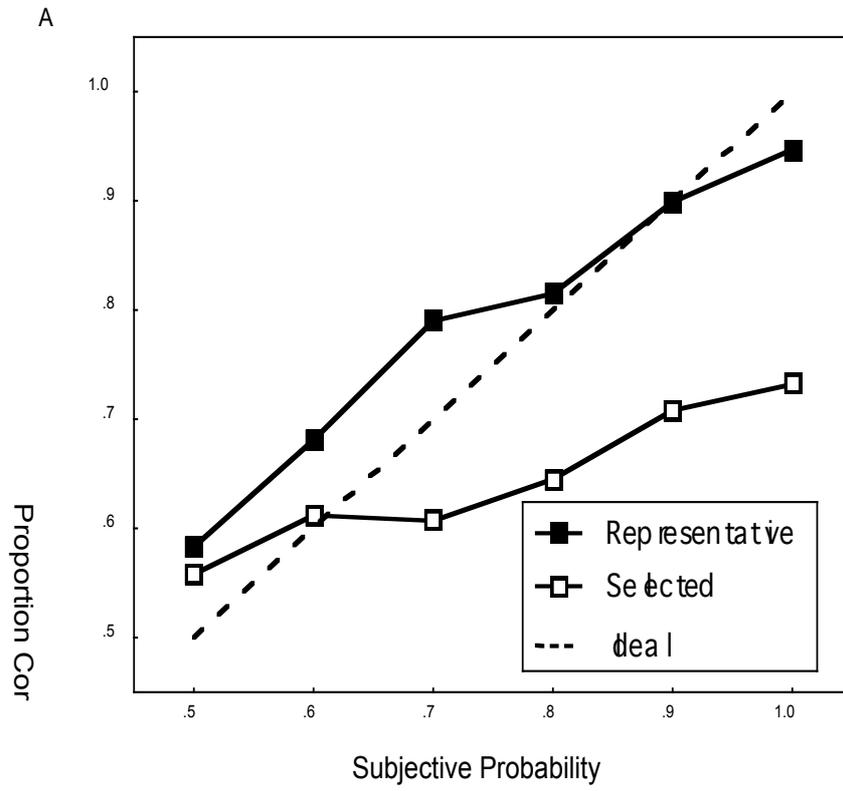
*Abbildung 2.* Kalibrierungskurve für Konfidenzurteile nach Juslin (1994)

*Abbildung 3.* Vier-Felder Kontingenztafel

*Abbildung 4.* Das Simpson-Paradox am Beispiel der Annahmeraten von zwei Graduiertenprogrammen

*Abbildung 5.* Vier mögliche Beziehungen zwischen den drei Variablen x, y und z. Der Einfluss eines weiteren Faktors z auf die fokale Beziehung  $r_{xy}$  zwischen x und y schlaegt sich in der Partialkorrelation  $r^*_{xy,z}$  nieder.





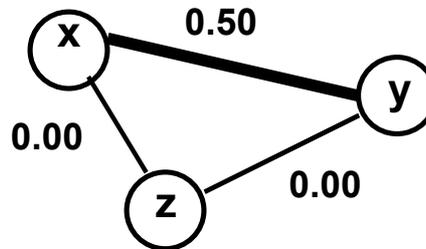
---

Attribut Y		
Attribut X	Ausprägung +	Ausprägung -
Ausprägung +	A	B
Ausprägung -	C	D

---

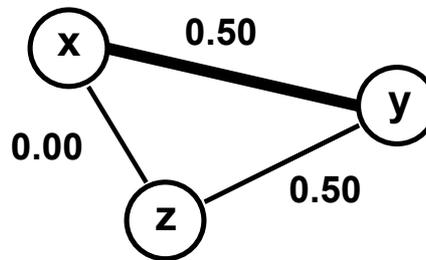
	<b>Annahme</b>	<b>Ablehnung</b>	<b>Annahmerate</b>	
<b>Männer</b>	<b>22</b>	<b>100</b>	<b>1/6</b>	<b>Insgesamt</b>
<b>Frauen</b>	<b>14</b>	<b>86</b>	<b>1/9</b>	
<b>Männer</b>	<b>20</b>	<b>80</b>	<b>1/5</b>	<b>Programm A</b>
<b>Frauen</b>	<b>4</b>	<b>6</b>	<b>2/5</b>	
<b>Männer</b>	<b>2</b>	<b>20</b>	<b>1/10</b>	<b>Programm B</b>
<b>Frauen</b>	<b>10</b>	<b>80</b>	<b>1/9</b>	

**Uni-kausaler  
Einfluss. z  
ohne Effekt**



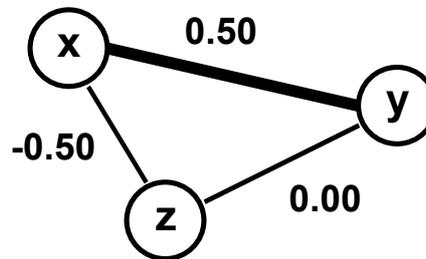
$$r_{xy.z}^* = r_{xy}$$

**Orthogonales  
Muster. z hat  
unabhäng.  
Einfluss**



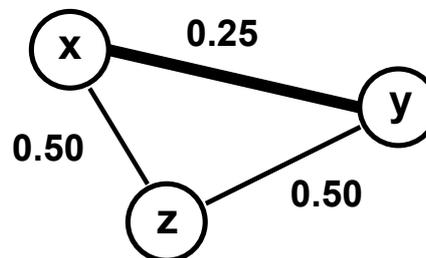
$$r_{xy.z}^* = r_{xy}$$

**Suppressor.  
z unterdrückt  
Fehlervarianz  
in x-y**



$$r_{xy.z}^* > r_{xy}$$

**Spurious  
correlation. z  
verantwortet  
x-y**



$$r_{xy.z}^* < r_{xy}$$