

Schematheoretische Begründungen für die Ordnung unsicheren Wissens

A. Zimmer / H. Körndle

Die Wissenspsychologie seit Anfang der 70er Jahre steht in einem seltsamen Spannungsverhältnis zu den sich wandelnden Architekturen und Grundkonzeptionen von Computern. Zunächst von SELFRIDGE (1960) und anderen als Metapher für Informationsverarbeitungsprozesse genutzt und damit von einem unzweifelhaften erkenntnistheoretischen Wert (s. FARRELL 1986), entwickelten sich die im von NEUMANNschen Computer angelegten Strukturen zu „constraints on human cognition“. So hat sich die lange Auseinandersetzung um serielle vs. parallele Verarbeitung in vielen Aspekten als Scheinfrage entpuppt, seitdem massive parallele und verteilte Verarbeitung technisch realisierbar ist (s. MCCLELLAND, RUMELHART & the PDP Group 1986). Aber auch die Unterscheidung von deklarativem und prozeduralem Wissen ist eine direkte Konsequenz der üblichen Programmiersprachen, wobei zweifellos Rückschlüsse von der Art ihrer Konstruktion auf die Wissensstrukturierung ihrer Konstrukteure möglich sind. Das oben angesprochene Spannungsverhältnis wirkt jedoch nicht nur in eine Richtung: Genuin menschliche Phänomene der Wissensverarbeitung, speziell die sogenannten Formate der Wissensrepräsentation werden seitens der Psychologie (und Philosophie) an die Computerwissenschaft als einzulösende und zu realisierende Forderungen herangetragen. Dem liegt u.E. eine Fehlbeurteilung über das Verhältnis zwischen dem funktionalen Wert einer Repräsentationsform und ihrer materiellen Realisierung zugrunde. Zweifellos gibt es prinzipielle Unterschiede zwischen einer semantischen und einer imaginalen Repräsentation, aber die letztere läßt sich prinzipiell im Rahmen der ersteren für jede vorgegebene empirische Untersuchungsgenauigkeit nachbilden; dies ist die eigentlich interessante Konsequenz von ANDERSONs (1976) Mimikry-Theorem. Ähnliches gilt von der Unterscheidung episodischer vs. propositionaler Repräsentation (TULVING 1984); auch hier läßt sich in einem hinreichend allgemeinen propositionalen Kalkül (einschließlich der Modal- und „tense“ Logik) jede episodische Struktur nachbilden. Allgemein läßt sich festhalten, daß es, zumindest auf der Ebene der mit herkömmlichen psychologischen Methoden erreichbaren Genauigkeit, nicht möglich ist, zwischen nicht-symbolischen und propositionalen Formaten hinreichend genau zu unterscheiden. Von größerer Bedeutung sind zwei Aspekte des Wissens, die bei der Diskussion um die Repräsentationsformate in den Hintergrund getreten sind:

- 1) Menschliches Wissen (im folgenden wird unter Wissen stets empirisches Wissen verstanden) ist prinzipiell unsicher, sei es infolge mangelnder Kenntnis, dem Zugrundeliegen stochastischer Prozesse in der beobachtbaren Welt oder der prinzipiellen oder praktischen Unmöglichkeit der Beobachtung.
- 2) Das Reden über bzw. die Kommunikation von menschlichem Wissen ist prinzipiell kontextabhängig; d.h. „wörtliche“ Bedeutungen im strengen Sinn treten nicht auf.

Sollen nun, wie z.B. in Expertensystemen, Vernetzungen zwischen menschlichem Wissen und der Symbolverarbeitungskapazität von Computern optimal gestaltet werden, müssen diese prinzipielle Unsicherheit und die Kontextabhängigkeit in nicht-trivialer Weise in den Symbolverarbeitungsprozeß eingearbeitet werden (eine triviale, aber nicht ausreichende Behandlung von Unsicherheit besteht z.B. in Mittelwertsbildungen, aber auch einigen Techniken der sog. „second-order statistics“).

In den hier vorgestellten Überlegungen soll gezeigt werden, daß eine sinnvolle Bearbeitung von Unsicherheit nur im Rahmen eines Ansatzes zur kontextuellen Bedeutungsbestimmung geschehen kann. Als formales Modell für Kontexteinflüsse wird eine interaktive Schema-Hierarchie vorgestellt, wobei unter Schemata nach CASSIRER (1944) in Anlehnung an KANTS Schemabegriff ein „kognitiver Mechanismus zur Zuordnung von sinnlich Wahrgenommenem zu Begriffen“ verstanden wird. Ein solcher Mechanismus (s. ZIMMER 1984, 1985, 1986 c) besteht aus Primitiva (die durchaus Schemata niedriger Hierarchieebenen sein könnten), Zuordnungsregeln (wie z.B. die sog. Gestaltgesetze oder die Diskursregel der klassischen Rhetorik (s. VUKOVICH, in Vorbereitung)) und die Menge zulässiger Transformationen, durch welche die Invarianzbedingungen gegeben sind. Diese ursprünglich an der Gruppentheorie orientierte Definition des Schemas bei CASSIRER trägt nicht nur bei den von ihm analysierten Phänomenen der Konstanz, sondern auch in der Raumwahrnehmung (ZIMMER 1986 c) und in der Linguistik. Es ist jedoch fraglich, ob die Anforderungen für algebraische Strukturen in jedem Fall einlösbar und auch sinnvoll sind.

Zur Illustration, wie Kontextualität und Auflösung von Unsicherheit (hier als Verwackelung i.S. von ZAUS (1984)) zusammenhängen, kann Abbildung 1 dienen.

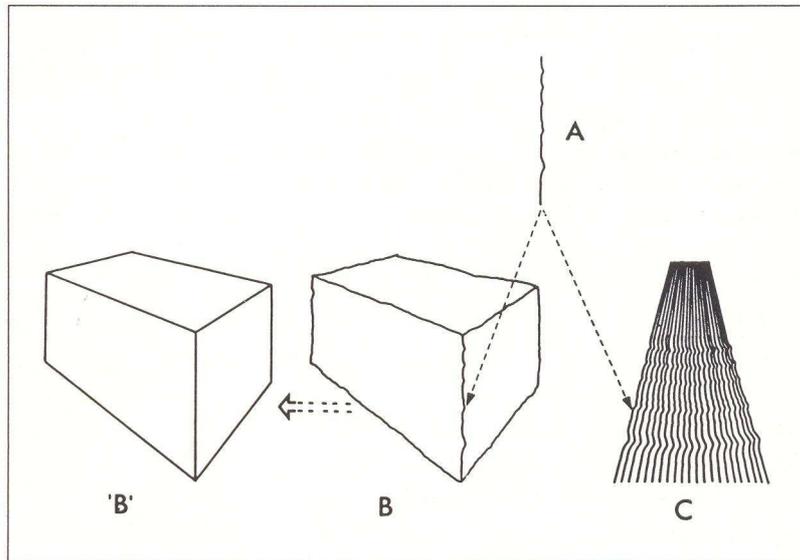


Abb. 1: Linienzug, der in B integriert als „verwackelte“ Gerade und in A als Teil einer gewellten räumlichen Textur i.S. GIBSONS (1979) zu interpretieren ist.

In Mustererkennungsalgorithmen (JAIN & HAYNES 1982) würde der Linienzug bei rein lokalem Vorgehen durch Filterung als unvollkommen realisierte oder fehlerhaft übertragene Gerade interpretiert und weiterverarbeitet, ähnlich würden interaktive Vorgehensweisen von „graceful degradation“ und Filterung nicht zur Texturdeutung in Abb. 1C führen, die durchaus auch ein praktisches Problem der Mustererkennung darstellt, z.B. bei Situationen wie einem Landeanflug auf unbekanntem Gebiet. Ohne an dieser Stelle auf Vor- oder Nachteile verschiedener Filtertypen und hybrider Mustererkennungsmodelle weiter einzugehen, sollte durch die Abbildung deutlich geworden sein, daß

was als wegzufilternde Ungenauigkeit oder Unsicherheit interpretiert wird, von der Kontexteinbettung abhängt, aber auch, daß der einbettende Kontext bei „früher Filterung“ möglicherweise als solcher gar nicht identifiziert werden kann. Dieses Problem ist nur bei paralleler Verarbeitung von Unsicherheit und Kontext zu lösen.

Im folgenden wird im Rahmen der Possibility-Theorie von ZADEH (1983, 1984) ein gemeinsamer Bezugsrahmen für Unsicherheitsausdrücke entwickelt, die auf unterschiedliche Faktoren zurückgehen (partielle Unwissenheit, stochastischen Gegebenheiten, Verallgemeinerungen auf neue Sachverhalte) und in unterschiedlichen sprachlichen bzw. logischen Ausdrücken artikuliert werden: Wahrscheinlichkeits- bzw. Unsicherheitsbegriffe (wie z.B. „sehr wahrscheinlich“, „praktisch sicher“), Quantoren (wie z.B. „fast alle“, „manche“) und temporale Terme (wie z.B. „häufig“, „selten“). Abschließend wird am Beispiel der Beurteilung von Unsicherheiten bei komplexen Gegebenheiten gezeigt, wie menschliche Unsicherheitsurteile und computergestützte Entscheidungshilfe optimal kombiniert werden können.

Ausgangspunkt der Untersuchungen von ZIMMER (1980, 1984, 1986 a & b) war die Frage nach dem Informationsgehalt verbaler Unsicherheitsausdrücke; hierbei ergab sich eine hohe **intra**individuelle Konsistenz der Urteile, aber gravierende **inter**individuelle Unterschiede hinsichtlich der Anzahl und zugrundeliegender Bedeutung der spontan verwendeten Ausdrücke.

Die individuell kalibrierten Bedeutungen qualitativer Unsicherheitsausdrücke lieferte uns ein Kalibrierungsverfahren, das aus zwei Schritten besteht.

Im ersten Schritt müssen die Beurteiler einen Fragenkatalog zu unsicheren Alltagsereignissen (z.B.: Sie kommen 1 Minute zu spät zum Bahnhof. Erreichen Sie dennoch den Zug?) spontan mit ihren eigenen Unsicherheitsausdrücken beantworten.

Im zweiten Schritt werden diese individuellen Begriffe mit Hilfe eines Computers kalibriert. WALLSTEN (1986) verwendet dazu eine Paarvergleichs-Anordnung, bei der die Probanden auf einer Skala angeben müssen, welches der paarweise auf einem Computerbildschirm präsentierten Kreissegmente besser auf den zu kalibrierenden Begriff paßt (s. Abb. 2).

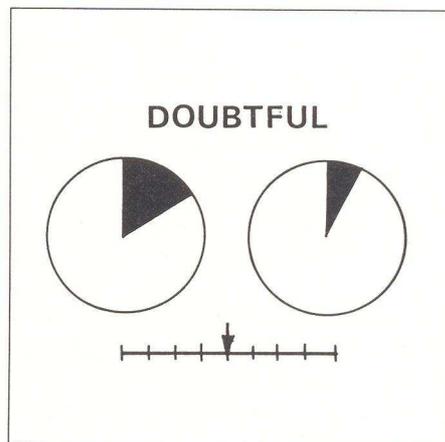


Abb. 2: Bildschirmdarstellung zur Paarvergleichsmethodik von WALLSTEN (1986)

Wir verwenden für diesen zweiten Schritt ein adaptives sequentielles Verfahren, das statt Kreissegmenten Zufallsmuster benutzt. Diese Kombination erwies sich als gleichzeitig effizient und stabil in der Schätzung, besonders wenn das zuletzt beurteilte Muster zugleich mit dem jeweils neuen Muster gezeigt wird.

Das adaptive sequentielle Verfahren läuft dabei folgendermaßen ab: Auf dem Bildschirm wird rechts ein Zufallspunktemuster gezeigt, das flächenmäßig zu 50 % mit Punkten gefüllt ist. Der Beurteiler hat die Aufgabe, die Wahrscheinlichkeit einzuschätzen, beim blinden Zeigen auf das Muster einen der Punkte zu treffen. Dann wandert dieses Muster, versehen mit dem Unsicherheitsausdruck nach links, auf der rechten Seite erscheint ein neues Muster mit zu 55 % flächendeckenden Punkten. In den folgenden Durchgängen werden die Muster und Begriffe des i -ten (letzten) Durchgangs auf der linken, die des $(i + 1)$ -ten Durchgangs immer auf der rechten Seite gezeigt. Die Anzahl der Punkte im $(i + 1)$ -ten Durchgang wird dabei durch folgendes adaptive Verfahren bestimmt (stochastische Approximation, WETHERILL 1963):

$$d(t_{i+1}) = d(t_i) + \frac{u(t_i)}{c \cdot i} ,$$

wobei

$$u(t_i) = \begin{cases} +1, & \text{wenn } u(t_i) = u(t_{i-1}) \\ -2, & \text{wenn } u(t_i) \neq u(t_{i-1}) \end{cases}$$

mit

$$\begin{aligned} d(t_i) &= \text{Anzahl der Punkte im Durchgang } i \\ u(t_i) &= \text{Beurteilung im Durchgang } i \\ c &= \text{Konstante, } 0 < c \leq 1. \end{aligned}$$

Durch entsprechende Wahl der Konstanten c kann das Verfahren für eine grobe Beurteilung beschleunigt bzw. für fein abgestimmte Antworten verlangsamt werden. Nach 9 Durchgängen mit $u(t_{i+1}) \neq u(t_i)$ folgt eine neue Startposition z.B. mit $d(t_i) = 45\%$ und abnehmenden Punktezahlen, wobei

$$u(t_i) = \begin{cases} -1, & \text{wenn } u(t_i) = u(t_{i-1}) \\ +2, & \text{wenn } u(t_i) \neq u(t_{i-1}). \end{cases}$$

Startposition und Richtung (zunehmend/abnehmend) ändern sich immer dann, wenn das Kriterium von 9 Beurteilungswechseln $u(t_i) \neq u(t_{i+1})$ erreicht wurde.

Das Verfahren wird solange durchgeführt, bis für jeden Unsicherheitsausdruck je eine zunehmende und abnehmende Folge von Beurteilungen erfaßt wurde. Eine Ausnahme stellen natürlich die Extremkategorien dar, für die eine Folge genügt.

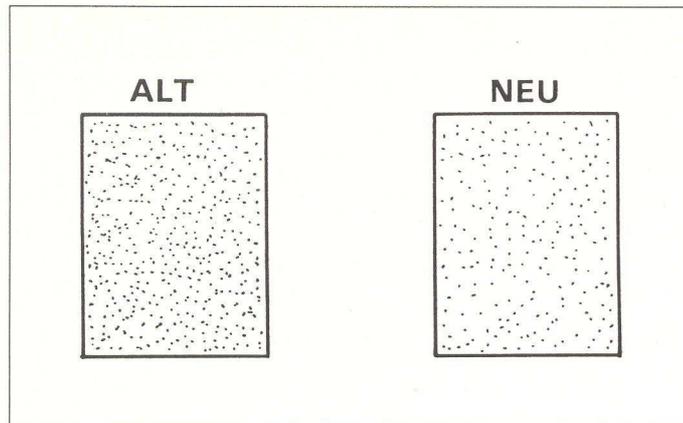


Abb. 3: Zufallsmuster für das Staircase-Verfahren zur Kalibrierung von Unsicherheitsausdrücken von ZIMMER (1986 a)

Ergebnis dieser Kalibrierungen ist die Zuweisung einer „fuzzy number“ (DUBOIS & PRADE 1980) zu den verwendeten Unsicherheitsbegriffen; diese können dann algorithmisch weiterverarbeitet werden.

Bei der Erhebung von Unsicherheitsausdrücken wurden von den Versuchspersonen spontan verschiedene Quantoren („fast alle“, „viele“, „praktisch nie“, „sehr oft“ u.ä.) genannt, speziell bei der Begründung von Unsicherheitsurteilen, z.B.: „Es ist sehr wahrscheinlich, daß R. REAGAN wiedergewählt wird, da fast immer Amtsinhaber von den Wählern bevorzugt werden“. Diese Begründung ist einem syllogistischen Schluß in der folgenden Form äquivalent:

$$\begin{array}{l}
 P_1 \text{ Fast immer } (A \text{ ist } W) \\
 P_2 \text{ } R \text{ ist } A \\
 \hline
 C \text{ sehr wahrscheinlich } (R \text{ ist } W)
 \end{array}$$

Da für nicht-klassische Quantoren bzw. speziell die Mischung von Quantoren und Wahrscheinlichkeiten keine Auflösungsregeln bestehen, ist von ZADEH (1984) und ZIMMER (1984) vorgeschlagen worden, sie als „fuzzy numbers“ zu interpretieren und die entsprechenden Operatoren zur Generierung der Konklusionen zu benutzen. Die empirischen Untersuchungen von ZIMMER (1984) zu unscharfen Quantoren haben gezeigt, daß einerseits das von GOGUEN (1969) vorgeschlagene Modell der Superpositionen von Unschärfe über die Bedeutung von klassischen Quantoren nicht zutrifft, andererseits aber auch das von ZIMMER (1980) vorgeschlagene Modell für linguistische Variablen (gleiche Informativität aller Elemente) auf Quantoren nur dann anwendbar ist, wenn kontextuelle Einflüsse ausgeschlossen werden können. In Kontexten wie „alltägliche Begebenheiten“, „Sachverhalte aus Sozialwissenschaften“ und „Sachverhalte aus Naturwissenschaften“ stellt sich heraus, daß die resultierende Bedeutung der Quantoren zum einen auf das „ideale“ (i.S. von GRICE 1975) Modell gleicher Informativität und zum anderen auf die kontextabhängigen Scope-Funktionen zurückzuführen sind (Abb. 4 a, b, c).

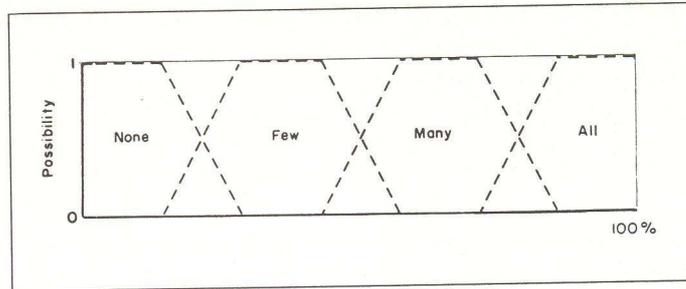


Abb. 4a: Bedeutung der Quantoren im Modell gleicher Informativität (ZIMMER 1980)

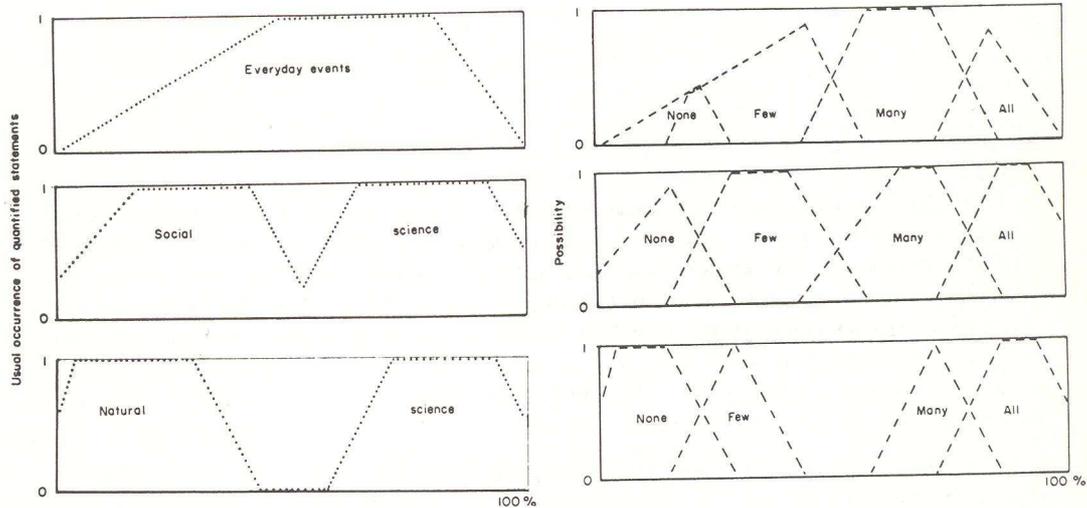


Abb. 4b: Scope-Funktionen für Kontexte (Wie häufig wird im Kontext x über Ereignisse mit Häufigkeit ... % geredet?)

Abb. 4c: Resultierende kontextunabhängige unscharfe Quantoren

Scope-Funktionen geben an, inwieweit Aussagen, die einen bestimmten Prozentsatz von Präferenzobjekten betreffen, für den jeweiligen Kontext typisch sind, wobei hier Typikalität im Sinne von ROSCH (1975) verstanden wird. Der Kontext determiniert also nicht nur die Verwendung von Quantoren (dies drückt sich darin aus, daß z.B. im Kontext „alltägliche Begebenheiten“ die extremen Quantoren nicht „Possibility“-Werte von 1 erreichen), sondern auch ihre Bedeutung. In den Scope-Funktionen spiegelt sich darüber hinaus das allgemeine Verständnis darüber wider, welche Referenzklassen für die verschiedenen Wissensbereiche von Interesse sind; kurz ausgedrückt:

Je „härter“ das Wissen ist, desto weniger sind Ereignisse eines mittleren Wahrscheinlichkeitsgrades von Bedeutung. Es kann gezeigt werden, daß bei Beschränkung auf die Quantoren „alle“ und „kein“ sowie ihrer unscharfen Negationen und der Faltung mit einer geeigneten Scope-Funktion die klassische Bedeutung als Sonderfall der unscharfen, d.h. nicht-klassischen Bedeutung aufgefaßt werden kann. Der in Abbildung 1 angedeutete Sachverhalt der kontextuellen Bedeutungszuweisung bzw. Trennung von Information und Rauschen trifft also auch auf Quantoren in eher umgangssprachlicher Bedeutung zu.

Zusammenfassend ist festzuhalten, daß bei der Interpretation von Quantoren als „fuzzy numbers“ im Intervall $[0,1]$, deren Werte durch die Scope-Funktionen kontextuell modifiziert worden sind, diese mit subjektiven Wahrscheinlichkeiten (verbale Unsicherheitsausdrücke als „fuzzy numbers“) kombiniert werden können und damit die Auswertung von Evidenzketten oder -netzen möglich wird, deren Propositionen durch Quantoren bzw. Wahrscheinlichkeiten gewichtet sind. Als allgemeiner Auswertungsalgorithmus für akkumulierte Evidenz bietet sich also ein „fuzzy number“-Algorithmus an, da dieser nicht nur Quantoren und Wahrscheinlichkeiten vergleichbar repräsentiert, sondern auch die Kontextualisierung aller Bedeutungsvarianten von der klassischen (scharfen) Interpretation bis zur umgangssprachlichen Verwendung darzustellen und auszuwerten vermag.

Ausgangspunkt ist die Frage, unter welchen Bedingungen menschliche Beurteiler die vorhandene Information (Evidenz) optimal auswerten. Das Standardkriterium für optimale Informationsauswertung besteht darin, daß geprüft wird, ob neue Information zu einer entsprechenden Revision des vorherigen Urteils führt ('Bayesian information updating', da ihm die Anwendung des Bayes-Theorems für bedingte Wahrscheinlichkeiten zugrundeliegt). Um diese Frage zu untersuchen, haben wir mit komplexen Geschichten gearbeitet und nicht, wie sonst üblich, mit der Abschätzung von einzelnen Ereignissen. Wie sieht so eine komplexe Geschichte aus? Ein Beispiel ist ein Kriminalfall [analog zu SHAFER & TVERSKY (1985) Fall von GRACCHUS & MAEVIUS]: Ein Gärtner steht im Verdacht, den Butler des gleichen Hauses umgebracht zu haben. Dagegen spricht, daß beide eigentlich sehr gute Freunde waren. Es spricht ebenfalls dagegen, daß das Standardmotiv für Mord, nämlich Geld, ausgeschlossen werden kann. Aber der Butler hat irgendwann einmal die Schwester des Gärtners sitzen lassen. Der Gärtner gilt als nachtragend und, wenn er tatsächlich von Haßanfällen überkommen wird, auch als gewalttätig.

Eine Geschichte dieser Art kann man in zweierlei Weise analysieren. Die eine Vorgehensweise wäre, der Versuchsperson ein Gesamturteil abzufordern, nämlich: Wie wahrscheinlich ist es, daß der Gärtner den Butler umgebracht hat, wenn kein anderer die Gelegenheit hatte und es sich ganz sicher nicht um einen Selbstmord gehandelt hat?

Die Alternative dazu besteht in der von SHAFER & TVERSKY (1985) vorgeschlagenen Partitionierung der Geschichte in viele Einzelpropositionen, die jeweils bedingte oder unbedingte Evidenzen enthalten. Läßt man diese Einzelpropositionen beurteilen, dann lassen sich mit dem Bayes'schen Kalkül diese zu dem Gesamturteil akkumulieren. Der Vergleich des globalen Gesamturteils und des akkumulierten lassen den Schluß zu, wie bayesianisch der einzelne Globalbeurteiler tatsächlich ist.

Wie man mit Hilfe der Partitionierung die in dem Szenario enthaltene Evidenz auswerten kann, wird in den Tabellen 2 und 3 dargestellt. In Tabelle 1 finden sich die Abkürzungen für elementare Propositionen, die bei der Evidenzbewertung in den Tabellen 2 und 3 benutzt werden.

Tabelle 1: Abkürzungen der unbedingten elementaren Propositionen des Szenarios

H	:=	Der Gärtner haßte den Butler
GI	:=	Der Gärtner beabsichtigte, den Butler zu töten
SI	:=	Jemand anderer beabsichtigte, den Butler zu töten
GM	:=	Der Gärtner hat den Butler umgebracht
SM	:=	Jemand anderer hat den Butler umgebracht
NM	:=	Niemand hat den Butler umgebracht

In Tabelle 2 ist die vollständige Partitionierung analog der Vorgehensweise von SHAFER & TVERSKY (1985) wiedergegeben, wobei auch die numerischen Werte übernommen worden sind.

Tabelle 2: Auswertung der Evidenz mit numerischen Urteilen

Zuweisung numerischer Werte (Shafer & Tversky 1985) zu den Propositionen, die für die Schuld des Gärtners sprechen:

$$p(H) = 0.2, p(GI|H) = 0.2, p(GI|\neg H) = 0.01$$

$$p(SI) = 0.001$$

SI und GI sind unabhängig

$$p(GM|GI \& SI) = 0.4$$

$$p(SM|GI \& SI) = 0.4$$

$$p(NM|GI \& SI) = 0.2$$

$$p(GM|GI \& \neg SI) = 0.8$$

$$p(NM|GI \& \neg SI) = 0.2$$

$$p(SM|SI \& \neg GI) = 0.8$$

$$p(NM|SI \& \neg GI) = 0.2$$

$$p(NM|\neg GI \& \neg SI) = 1$$

$$\begin{aligned} p(GI) &= p(GI|H)p(H) + p(GI|\neg H)p(\neg H) \\ &= 0.2 \cdot 0.2 + 0.1 \cdot 0.8 = \underline{0.48} \end{aligned}$$

$$\begin{aligned} p(GM) &= p(GM|\neg GI)p(\neg GI) + p(GM|GI \& SI)p(GI)p(SI) \\ &\quad + p(GM|GI \& \neg SI)p(GI)p(\neg SI) \\ &= 0 \cdot 0.952 + 0.4 \cdot 0.048 \cdot 0.001 \\ &\quad + 0.8 \cdot 0.048 \cdot 0.999 = \underline{0.3838} \end{aligned}$$

$$p(SM) = 0.00078$$

$$1 - p(NM) = 0.96084$$

$$p(GM|\neg NM) = \frac{0.3838}{0.3838 + 0.00078} = 0.98$$

Die von SHAFER & TVERSKY (1985) angenommenen Wahrscheinlichkeiten für die elementaren und einfach bedingten Propositionen erscheinen plausibel und stimmen auch sehr gut mit Werten überein, die wir in eigenen Experimenten erhoben haben. Im Gegensatz dazu ist die Gesamtbewertung der akkumulierten Evidenz, daß nämlich mit einer Wahrscheinlichkeit von 0.98 der Gärtner

den Butler getötet hat, unrealistisch hoch. Dies wird besonders deutlich, wenn man in Betracht zieht, daß diese Gesamturteilung darauf basiert, daß der Wahrscheinlichkeit, jemand anderer habe den Butler umgebracht, ein Wert von weniger als einem Zehntausendstel zugewiesen wird; dies impliziert eine Reliabilität der empirischen Daten, wie sie praktisch nicht erreichbar ist.

Eine Möglichkeit, die mangelnde Reliabilität der Einzelschätzungen zu berücksichtigen, besteht darin, die Zahlen als „fuzzy numbers“ zu interpretieren. Wie in ZIMMER (1986 a) gezeigt, wird mit dieser Vorgehensweise der mangelnden Reliabilität der Daten Rechnung getragen. Führt aber die Verwendung von Zahlen selbst zu systematischen Verfälschungen, wird dies durch ihre Interpretation als 'fuzzy numbers' nicht aufgefangen. Aus diesem Grunde sind wir bei der Beurteilung von elementaren Propositionen und einfach bedingten Propositionen dazu übergegangen, verbale Urteile zu erheben und diese mit dem schon beschriebenen Verfahren zu kalibrieren, so daß es möglich ist, verbale Urteile als „fuzzy numbers“ zu interpretieren und weiterzuverarbeiten. Für das gegebene Szenario sehen diese Beurteilungen wie folgt aus (Tabelle 3):

Tabelle 3: Auswertung der Evidenz mit verbalen Urteilen

Verbale Beurteilungen (interpretiert als "fuzzy numbers") der Propositionen:	
$p(H)$	= "eher unwahrscheinlich"
$p(GI H)$	= "eher unwahrscheinlich"
$p(GI \neg H)$	= "sehr unwahrscheinlich"
$p(SI)$	= "extrem unwahrscheinlich"
SI und GI sind unabhängig	
$p(GM GI \& SI)$	= "gut möglich"
$p(SM GI \& SI)$	= "gut möglich"
$p(NM GI \& SI)$	= "eher unwahrscheinlich"
$p(GM GI \& \neg SI)$	= "sehr wahrscheinlich"
$p(NM GI \& \neg SI)$	= "eher unwahrscheinlich"
$p(SM SI \& \neg GI)$	= "sehr wahrscheinlich"
$p(NM SI \& \neg GI)$	= "eher unwahrscheinlich"
$p(NM \neg GI \& \neg SI)$	= "sicher"
$p(GM \neg NM)$	= "(wahrscheinlich) bis (sehr wahrscheinlich)"

Das Ergebnis „(wahrscheinlich bis sehr wahrscheinlich)“ spiegelt nicht nur die inhärente Unsicherheit der Einzelurteile wieder, sondern enthält als Spezialfall auch das von SHAFER & TVERSKY (1985) erhobene Ergebnis. Graphisch läßt sich dieses Ergebnis wie in Abbildung 5 darstellen.

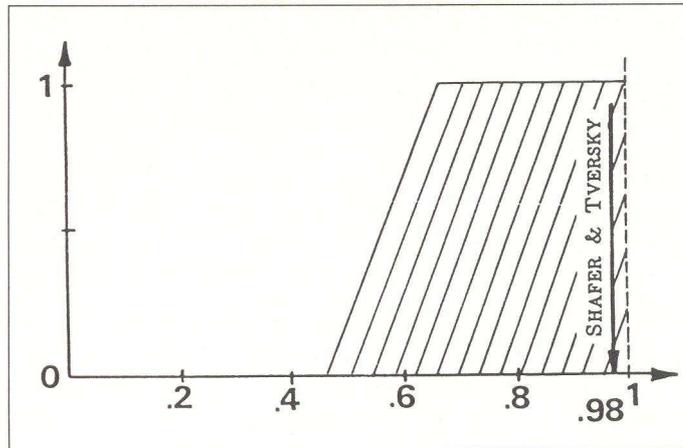


Abb. 5: Charakteristische Funktion für die „fuzzy number“ der Gesamtbeurteilung der akkumulierten Evidenz

Betrachtet man die Propositionen in den Tabellen 2 und 3, fällt auf, daß nahezu ausschließlich bedingte Propositionen auftreten. SHAFER & TVERSKY (1985) fügen selbst keine Gründe an, warum sie gerade diese Partitionierung der Evidenz vorgenommen haben und keine andere äquivalente. Aufgrund der Experimente zur kontextspezifischen Bedeutung von Quantoren (ZIMMER 1984) haben wir untersucht, inwieweit das Urteilsverhalten von der Art der zu beurteilenden Propositionen abhängt. Dabei hat sich herausgestellt, daß die geringste intraindividuelle Variabilität bei Propositionen auftrat, die einfach bedingt waren und deren bedingendes Ereignis konkret vorstellbar war. Sowohl bei der Beurteilung elementarer, d.h. nicht bedingter Propositionen als auch bei mehrfach bedingten Propositionen bzw. bei nicht konkreten bedingenden Ereignissen traten sehr häufig Inkonsistenzen auf. Diese Vorliebe für einfache Konditionierungen auf leicht vorstellbare Ereignisse könnte die Ursache für die sog. „conjunction-fallacy“ (TVERSKY & KAHNEMAN 1983) sein, die ja dann auftritt, wenn das bedingende Ereignis ein abstrakter Oberbegriff für eine ganze Ereignisklasse ist (z.B. Naturereignisse als Oberklasse zu Erdbeben). Ähnlich wie bei den Quantoren stellt es sich also auch bei den subjektiven Wahrscheinlichkeitsbegriffen heraus, daß die kontextspezifische Bedeutung die „natürliche“ Form des Umgangs mit solchen Begriffen ist. Es läßt sich zeigen (ZIMMER 1986 a), daß die Generierung von kontextspezifischen Bedeutungen für Quantoren mit Hilfe der Scope-Funktionen nichts anderes als eine spezielle Form der Konditionierung ist. Aus diesem Grunde ist es möglich, im Partitionierungsalgorithmus sowohl Propositionen zu verwenden, deren subjektive Wahrscheinlichkeit verbal beurteilt worden ist, als auch solche, die umgangssprachlich quantifiziert worden sind. Auf diese Weise lassen sich relativ leicht komplexe Evidenznetze herstellen.

Um die weiter vorn gestellte Frage zu beantworten, inwieweit und unter welchen Umständen menschliche Beurteiler bayesianisch seien, ist es notwendig, von ein und denselben Versuchspersonen einmal Urteile für die einzelnen Propositionen zu erheben und das Gesamtergebnis darauf abzuleiten und zum zweiten Globalurteile über das Gesamtszenario zu erheben. Wenn also eine Versuchsperson bayesianisch urteilt, dann sollte bei der Auswertung der akkumulierten Urteile dasselbe herauskommen, wie bei den Globalbeurteilungen.

In einer Reihe von Experimenten (ZIMMER 1987) hat sich folgendes für die Beurteilung komplexer Evidenz durch Experten bzw. Laien ergeben:

1) Experten und Laien unterscheiden sich signifikant hinsichtlich der Komplexität der Propositionen, die sie konsistent beurteilen können z.B. die Anzahl und die Art der Bedingungen. Darüberhinaus sind Experten besser in der Lage als Laien, Propositionen umzuformen, so daß am Ende alle zu beurteilenden Propositionen das gleiche Format haben. Während Propositionen, die eine Negation beinhalten, für Laien praktisch nicht konsistent bewertbar waren, zeigte sich der Effekt der Erschwerung von Informationsverarbeitung durch Negation bei Experten in geringerem Umfang; allerdings verschwand dieser Effekt auch bei ihnen nicht vollständig.

2) Wenn man unter Berücksichtigung des ersten Ergebnisses für Experten und Laien Evidenzketten erstellt, die in ihrer Komplexität äquivalent sind, dann zeigen Experten wie Laien praktisch das gleiche Urteilsverhalten, d.h. sie wenden Heuristiken an, in denen nur ein Teil der Evidenz verarbeitet wird oder in denen die strukturellen Beziehungen zwischen den einzelnen Evidenzen nicht berücksichtigt werden. Insgesamt führt dies zu einem konservativen Verhalten bei der Gesamtbeurteilung, selbst wenn bei Einzelbeurteilungen „Overconfidence“ feststellbar ist (ein ähnliches Ergebnis findet sich in May 1986).

3) Das oben beschriebene konservative Gesamtbeurteilungsverhalten läßt sich auf drei, empirisch unterscheidbare Heuristiken zurückführen:

Die „Odds“-Heuristik:

Diese Heuristik besteht darin, daß praktisch das Verhältnis von Propositionen „dafür“ und „dagegen“ gebildet wird und dieses Verhältnis am Ende das Gesamturteil bestimmt. Varianten dieser Heuristik bestehen darin, daß entweder die Stärke der Evidenz für jede einzelne Proposition als Gewichtungsfaktor einbezogen wird oder aber nur die Evidenzen, die einen bestimmten Schwellenwert überschreiten, verrechnet werden.

Die „Maximum“-Heuristik:

Diese Heuristik beinhaltet, daß die Gesamtbeurteilung durch die Höhe der größten Einzelevidenz bestimmt wird. Bei dieser Heuristik wird insgesamt am wenigsten Information verarbeitet und das Gesamtergebnis ist üblicherweise am „konservativsten“.

Die „Referenzsystem“-These:

Bei dieser Heuristik wird aufgrund der ersten (3-5) Propositionen ein Referenzsystem aufgebaut, wonach alle folgenden Evidenzen angeordnet werden. Dieses Referenzsystem könnte man umgangssprachlich als „Grundeinstellung“ gegenüber dem fraglichen Sachverhalt beschreiben. Aufgrund der Anordnung der Propositionen lassen sich eindeutig positive, eindeutig negative und gemischte Referenzsysteme erzeugen; diese führen dazu, daß entweder alle folgende Information in positive oder negative Richtung oder überhaupt nicht verfälscht wird. Bei dieser Heuristik spielen die letzten Propositionen noch eine besondere Rolle; sind sie mit dem ursprünglichen Referenzsystem konsistent, dann wird das Urteil noch stärker in dieser Richtung verfälscht, während bei Inkonsistenz das Ausmaß der Verfälschung gemildert wird. Es muß allerdings festgehalten werden, daß dieser Verfälschungseffekt nur in Relation zu den anderen Heuristiken besteht. Gegenüber der vollständigen und optimalen Evidenzauswertung ist auch diese Heuristik konservativ.

Die Vor- und Nachteile der einzelnen Heuristiken sind in der folgenden Tabelle 4 dargestellt.

Tabelle 4:

Informations- verarbeitung Heu- ristik	Anzahl berücksichtigter Propositionen	Berücksichtigung der Evidenzstärke	Stabilität gegenüber systematischer Verzerrung (Bias-Freiheit)
„odds“	hoch	niedrig (Gleichsetzung)	hoch
„Maximum“	extrem niedrig	niedrig (nur höchste Evidenz berücksichtigt)	hoch
„Referenz- system“	niedrig	mäßig (nur am Ende & Anfang der Kette)	extrem niedrig

Daß die Referenz-Heuristik nicht auf „primacy“- und „recency“-Effekte des Gedächtnisses zurückzuführen ist, läßt sich dadurch zeigen, daß relevante, aber weder eindeutig dafür noch dagegen sprechende Informationen, am Anfang nicht zu einem Bias führen; ähnliches gilt für die letzte Proposition. Unter relevanter Information wird hier verstanden, daß die entsprechenden Propositionen nicht in sich selbst als Argumente dafür oder dagegen erkennbar sind, sondern erst aufgrund der Information, die durch weitere (in diesem Fall später kommende) Propositionen geliefert wird, sich dahin entwickeln.

Aus diesem Ergebnis lassen sich leicht Konsequenzen für den Einsatz von computergestützten Entscheidungssystemen ableiten. Erstens lassen sich subjektive verbale Wahrscheinlichkeitsbeurteilungen interaktiv mit einem Rechner erheben (ZIMMER 1986 b). Diese verbalen Urteile können mit Hilfe eines adaptiven sequentiellen Verfahrens kalibriert werden und endlich kann der Rechner die Information ohne Bias akkumulieren und optimal, d.h. bayesianisch, auswerten. Praktisch bedeutet dies, daß auf den beiden ersten Stufen das Urteilsverhalten der Versuchspersonen oder des Experten entscheidend ist und nur auf der dritten der evidenzauswertende Algorithmus des Rechners.

Literaturangaben:

- ANDERSON, J.R. (1976). *Language, memory, and thought*. Hillsdale, N.J.: Erlbaum.
- CASSIRER, E. (1944). The concept of group and the theory of perception. *Philosophy and Phenomenological Research*, **5**, 1-36.
- DUBOIS, D. & PRADE, H. (1980). *Fuzzy sets and systems: Theory and applications*. New York: Academic Press.
- FARRELL, E. (1986). Metaphor and psychology: A reply to Gholson and Barker. *American Psychologist*, **6**, 719-720.
- GIBSON, J.J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin Company.
- GOGUEN, J.A. (1969). The logic of inexact concepts. *Synthèse*, **19**, 325-373.
- GRICE, H.P. (1975). Logic and conversation. In P. COLE & J.L. MORGAN (Eds.), *Syntax and semantics* (Vol. 3). New York: Academic Press.
- JAIN, R. & HAYNES, S. (1982). Imprecision in computer vision. *IEEE Transactions on Computers*, August, **18**, 39-48.
- MCCLELLAND, J.L., RUMELHART, D.E. & the PDP Research Group. (1986). *Parallel distributed processing. Explorations in the microstructure of cognition* (Vol. 1: Foundations). Cambridge, Mass.: The MIT Press.
- MCCLELLAND, J.L., RUMELHART, D.E. & the PDP Research Group. (1986). *Parallel distributed processing. Explorations in the microstructure of cognition* (Vol. 2: Psychological and biological models). Cambridge, Mass.: The MIT Press.
- ROSCH, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, **104**, 192-233.
- SELFRIDGE, O. (1959). Pandemonium: A paradigm for learning. In *Symposium on the mechanization of thought process*. London: HM Stationery Office.
- SHAFER, G. & TVERSKY, A. (1985). Languages and Designs for Probability Judgment. *Cognitive Science*, **9**, 309-339.
- TULVING, E. (1984). Précis of elements of episodic memory. *The Behavioral and Brain Science*, **2**, 223-238.
- TVERSKY, A. & KAHNEMAN, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, **90**, 293-315.
- VUKOVICH, A. (1987). *Sprache und Gestalttheorie*. Plenumsvortrag gehalten bei der 5. wissenschaftlichen Arbeitstagung der Gesellschaft für Gestalttheorie und ihre Anwendung e.V.
- WALLSTEN, T.A. (1986). *Meanings of nonnumerical probability phrases*. Research Memorandum No. 67. University of North Carolina. Chapel Hill, N.C.
- WETHERILL, G.B. (1963). Sequential estimation of points on a psychometric function. *Journal of the Royal Statistical Society*, **B25**, 1-48.
- ZADEH, L.A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, **1**, 3-38.

- ZADEH, L.A. (1983). A computational approach to fuzzy quantifiers in natural languages. *Computers and Mathematics*, **9**, 149–184.
- ZADEH, L.A. (1984). *Syllogistic reasoning in fuzzy logic and its application to reasoning with dispositions*. Berkeley: UCB/EECS. Memorandum.
- ZAUS, M. (1984). *Stochastische Meßstrukturen* (Europäische Hochschulschriften). Frankfurt am Main, Bern, New York: Peter Lang.
- ZIMMER, A. (1980). Eine Formalisierung mnestisch stabilisierter Bezugssysteme auf der Grundlage von Toleranzmengen. In A. THOMAS & R. BRACKHANE (Hrsg.), *Wahrnehmen. Urteilen. Handeln*. (159–178).
- ZIMMER, A. (1984). A model for the interpretation of verbal predictions. *International Journal of Man-Machine Studies*, **20**, 121–134.
- ZIMMER, A. (1985). A model for schema-guided reasoning. In A. DI NOLA & A.G.S. VENTRE (Eds.), *The mathematics of fuzzy systems*. Verlag TÜV Rheinland.
- ZIMMER, A. (1986a). What uncertainty judgments can tell about the underlying subjective probabilities. In L.N. KANAL & J.F. LEMMER (Eds.), *Uncertainty in artificial intelligence* (249–258). Amsterdam: Elsevier Science Publishers.
- ZIMMER, A. (1986b). A fuzzy model for the accumulation of judgments by human experts. In W. KARWOWSKI & A. MITAL (Eds.), *Applications of fuzzy set theory in human factors*. Amsterdam: Elsevier Science Publishers.
- ZIMMER, A. (1986c). What makes the eye intelligent? *Gestalt Theory*, **4**, 256–279.
- ZIMMER, A. (1987). Verbale vs. numerische Wahrscheinlichkeitsurteile in komplexen Entscheidungssituationen. Vortrag 29. TeaP. Aachen.