Ergonomics of Hybrid Automated Systems I, edited by W. Karwowski, H.R. Parsaei and M.R. Wilhelm
Elsevier Science Publishers B.V., Amsterdam, 1988 — Printed in The Netherlands

DECISION SUPPORT USING QUALITATIVE EVIDENCE

Alf C. Zimmer
University of Regensburg, F.R.G.

1 ABSTRACT

On the background of possibility theory, a common framework for different expressions for modal qualification and uncertainty is developed. The uncertainty can be attributed to different factors (e.g. partial stochastic events, partial knowledge, or generalizations to novel situations). It can be qualitatively expressed by means of linguistic and logic terms: uncertainty judgments (e.g. probable, unlikely), colloquial quantifiers (e.g. few, most), temporal expressions (e.g. often, always) etc. These modal qualification are shown to play a major role in argumentative chains or more complex schemes. Finally, the results are applied to the evaluation of circumstantial evidence. It turns out that linguistic judgments lead to more realistic results them subjective numerical estimates. Furthermore, it is shown that human decision makers are quite good in evaluating elementary or simple conditioned proposition but in evaluating the total evidence they rely on biasing heuristics, therefore, the latter part should be processed by decision support systems.

2 INTRODUCTION

The history of decision aiding is a story of mutual disappointments: Decision makers have been upset about the inapplicability of results from decision science for real-world decision problems, and decision theorists bemoan the fact that decisions are made usually haphazardly and without regard for optimization algorithms and the like.

Behavioral decision theory (Edwards 1961, Einhorn & Hogarth 1981) has attempted to bring together decision makers and decision specialists by integrating subjective judgments about probability and value into models of decision making, the most prominent being the model of subjective expected utility (SEU model). On the other hand, starting with moral philosophy and ending with self-help books, there abound recepes for restructuring problems in such a way that decisions become easy. During the last years consultants or computerized decision support systems (see Hollnagel, Mancini & Woods 1986, Humphreys, Svenson, Vari 1983) have begun to integrate these complementary approaches. One major consequence of this has been a tendency away from normative frameworks as, for instance, the von Neumann & Morgenstern (1944) approach.

The applicability of normative decision theory (see e.g. v. Neumann & Morgenstern 1944) as developed in the framework of econometrics depends on (i) the knowledge of the probabilities of outcomes, (ii) the comparability of gains and losses and their integration into a one-dimensional value scale, (iii) the choice of a decision criterion.

The fact that prerequisites (i) and (ii) are usually not met and have to be replaced by subjective estimates has led to the development of the subjective-expected-utility (SEU) model (Edwards 1954) which has later on been generalized to Multi-Attribute-Utility Theory (MAUT, see Edwards & Newman 1982). Both models have been proved as valuable instruments in economic decision making (see Edwards' 1971 paper with the provocative title: "Don't waste an executive's time on decision making!").

However, humans tend to deviate from these normative models, for instance, they tend to stick to their initial estimates even if additional information makes a revision of the estimate mandatory. This so-called 'conservatism' effect plus the evidence of interactions between subjective probabilities and subjective utilities especially in betting have cast doubt on the basic assumption of econometrics, namely, that of 'rational man'. Therefore, the consequence for decision aiding systems is that they have to deal with the psychological mechanism that keep man from being rational.

There are two main possibilities for attacking this problem: one is to investigate the "cognitive illusions" deemed to be the cause of non-rational decision behavior. This is the way Kahneman and Tversky (for an overview see the collected articles in Kahneman, Slovic, Tversky 1982) have taken. They assume that usually valuable heuristics for coping with complexity lead the decision maker astray if they are overgeneralized. The two major heuristics are: (i) judging probabilities of events by assessing how available they are in one's memory (Availability), and (ii) judging probabilities of events by determining how similar these events are to representative models (Representativeness). Making people (especially, planners) conscious of these cognitive fallacies should enable them to improve the precision of their estimates or the consistency of their choices (Henrion 1982).

An alternative approach taken by Wallsten and his cooperators (s.u.) or Zimmer (1983, 1984, 1986) starts from the assumption that people, especially experts, accumulate knowledge about the frequency of events quite consistently *and* correctly, but that the internal representation of this knowledge about the uncertainty of events is *qualitative* and not quantitative as implicitly assumed in the SEU models and its derivatives. If the subjective knowledge about uncertainty is by and large qualitative, then verbal categories like 'very probable', 'practically impossible', or 'likely' have to be integrated into a behaviorally oriented decision model. Zimmer (1986) has shown that deviations from optimal information processing (e.g. Conservatism) are minimized if individually calibrated verbal categories are used instead of numerical estimates.

Part 2.1 of this articles concentrates on the further development of a computer-assisted calibration procedure for individual verbal categories for uncertainty of single events (two-alternative choice). In the part 2.2 this procedure is generalized to uncertainty judgments for multiple events that are case mutually exclusive in the simplest. More complex situations with dependent events or conjunctions of events will be analyzed in order to assess how complex a situation can be that is still reliably represented by verbal judgments.

While parts 2.1 and 2.2 concentrate on situations where the frequency of events is fairly high and therefore the verbal judgments rely on enough observations to form a stable frame of reference, part 2.3 explores how people estimate the *probability of scenarios,* that is, *complex systems of events*. These scenarios plays an important role in risk analysis and environmental impact analysis where frequentist information is either totally lacking or only known for isolated parts of the scenario.

3 CALIBRATION OF QUALITATIVE PROBABILITIES

The starting point of the experiments of Zimmer (1983, 1984) as well as those of Wallsten and his collaborators (e.g. Wallsten & Budescu 1983) was the question about the informativeness of verbal expressions of uncertainty. The analysis of the conditions under which these expressions are spontaneously uttered revealed that there is a high <u>intraindividual</u> consistency regarding the number and the meaning of the expressions but marked <u>interindividual</u> differences. Subjects used between 4 and 7 categories and split evenly about the question if 'probable' means more than 'likely' or vice versa.

3.1 Two-Alternative Situations

In order to arrive at individually calibrated meanings for qualitative expression of uncertainty, in the first step the spontaneous usage of such expressions is elicited by means of a questionaire about everyday uncertain events (e.g. "You are one minute late at the station. Will you catch the train nevertheless?).

In the second step, the sets of expressions of each subjects are calibrated in the following way: The subjects are seated in front of a computer terminal. On the screen an upright rectangle is shown with 50 % of its area covered by random dots. The subjects are asked to estimate the probability of hitting a dot when blindly pointing at the rectangle. Then the display is shifted to the left part of the screen and is labelled with the given expression. On the right side a new rectangle is shown with random dots covering 55 percent of the rectangle area. In the following trials the rectangles and the expressions of the last (i-th) trial are always shown on the left side of the screen and the rectangles of the new (i+1 st) trial on the right side. The number of dots in the i+1 st rectangle is determined by an adaptive procedure (stochastic approximation, Wetherill 1963):

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

where

$$u(t_i) = \begin{cases} +1 \text{ iff } u(t_i) = u(t_{i-1}) \\ -2 \text{ iff } u(t_i) = u(t_{i-1}) \end{cases}$$

$d(t_i)$: = number of dots in trial i
$u(t_i)$: = response in trial i
c: = constant, $0 < c \leq 1$.

By changing the value of the constant c the procedure can be speeded up for fast and coarse estimates of the meanings or slowed down in order get more fine-tuned estimates. After 9 trials with $u(t_{i+i}) = u(t_i)$ the procedure jumps to a new starting position, for instance, to $d(t_i) = 45$ % with decreasing steps.

The starting positions and directions are always changed when the criterion of 9 changes in one direction ($u(t_i) = u(t_{i-1})$) is reached. This is repeated until for each qualitative expression one decreasing and one increasing sequence of trials his been run except for the two most extreme expressions because their meaning is trivially limited by either 1 or 0. From the results of the trials the underlying meaning of the expressions can be determined as fuzzy probabilities (Zadeh 1968, Yager 1983).

In order to test the assumption that all subject have the same notion of uncertainty but differ only in the expressions they use for uncertainty, the results of all subjects with the same number of expressions are pooled and compared with the results of the probabilities for correct answers in an almanac test. The results in Figure 2 reveal that there is no bias in the probability estimates and that the different individual meanings vary only randomly.

These results (for details of this study see Zimmer 1983) indicate that the underlying notion of uncertainty is indeed the same for all subjects and furthermore that the meanings are unbiased as compared to the meanings derived from numerical estimates which consistently show 'overconfidence' (Lichtenstein, Fischhoff, Phillips 1982).

3.2 N-Alternative Situations

The usage of uncertainty expressions in 2-alternative situations appears natural and - at least in the analyzed experimental situations - leads to consistent estimates for categories of responses (r_{ik}). If, however, there are more than 2 alternatives, subjects do feel no longer comfortable with uncertainty expressions but turn to colloquial quantifiers like many, most, few, practically none etc. and change the form of their propositions accordingly: "I am moderately certain that there is X" is replaced by "There are probably few X". The uncertainty expression 'probably' is now only used for hedging but the major part of the uncertainty information is now conveyed by the colloquial quantifier 'few'. Colloquial quantifiers can be expressed as fuzzy numbers as well as uncertainty expressions (see Zimmer in press) but the meaning of these quantifiers varies systematically with contexts (see Figure 2). For well circumscribed contexts, scope functions can be determined which capture how the context modifies the meaning of quantifiers.

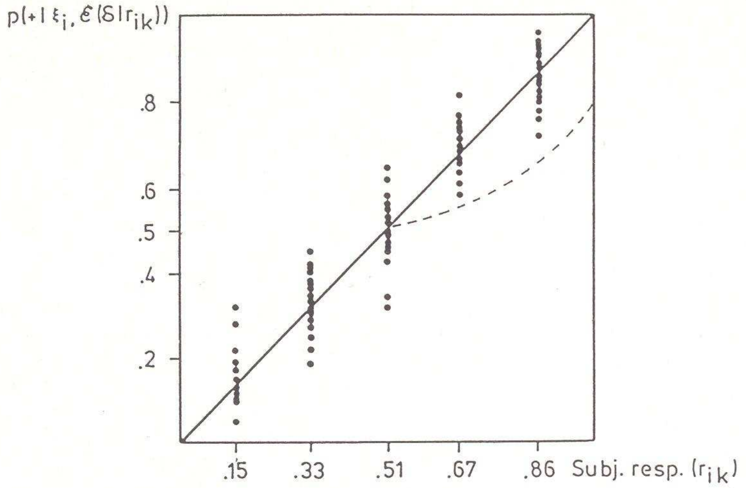

Figure 1: The probabilities of successes for given ability and difficulty levels conditioned on the applied varbal expressions plotted against the median values of the possibility functions for these verbal expressions. The parameters ξ and δ are the individual and the item parameters of the Rasch model. The dashed line indicates the bias of overconfidence from studies with numerical estimates of successes (Lichtenstein, Fischoff & Phillips 1982). These studies have only investigated 2-alternative responses, therefore the overconfidence function is bounded by 0.5 and 1.

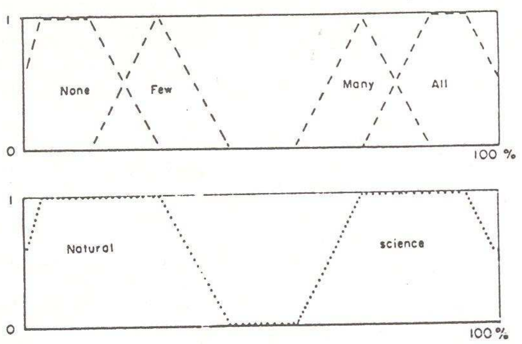

Figure 2: Context-dependent quantifiers and the scope function for the context of 'natural science'. (Zimmer 1984)

In complex reasoning schemes uncertainty expressions as well as explicit or implicit quantifiers are applied in order to evaluate claims (see Figure 3).

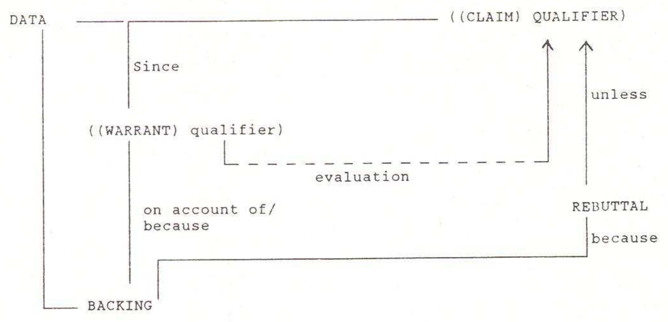

Figure 3: A modified version of Toulmin's (1964) model for syllogisms in argumentation. Toulmin's original model is indicated by upper-case letters and bold lines.

Although the meaning of simple quantified propositions and the claims of prototypical reasoning schemes (e.g. the syllogism modus Barbara) can be evaluated consistently and unbiased, this is no longer the case if the expressions become more complex or the chain of reasoning consists of more than about 5 propositions. In these cases experts as well as naive subjects tend to use simplifying and often biased heuristics in order to cope with the complexity (see part 3).

3.3 Situations with Circumstantial Evidence in Scenarios

A special case for reasoning is the evaluation of circumstantial evidence in scenarios as paradigmatically shown in Conan Doyle's detective stories where the evaluation of every single piece of evidence is straightforward but only Sherlock Holmes is able to put together all pieces in a cogent fashion. The reason why most people have the same problems as Dr. Watson with scenarios, seems to be that in these cases it is not possible to build up mental models in the sense of Johnson-Laird (1984) in which direct causal or set-mapping relationships lead to a unique solution that is compelling but sometimes wrong. For circumstantial-evidence situations such solutions are missing even if the plausibility (or evidential weight) for every piece of evidence can be judged without any problems.

In order to handle these situations, a number of schemes for the evaluation of circumstantial evidence have been proposed. The most straightforward solutions have been developed in the framework of Bayesian partitioning of evidence (Shafer & Tversky 1985). In order to show the strengths as well as the limits of this approach, we have experimentally investigated a criminal-case scenario constructed according to Shafer & Tversky's (1985) case of Gracchus and Maevius:

A gardener is suspected to have killed the butler. He seems to be the only one who had the opportunity and the butler has definitely not committed suicide. The gardener has also a motive, some years ago the butler has abandoned his sister and he is said to bear grudges very long and if angered tends to be quite brutal. However, the gardener used to be good friend with the butler and greed cannot be a motive.

One possibility to analyze such a scenario is by asking subjects for an overall judgment about the possibility that the gardener is the murderer if nobody else did it and suicide can be excluded as a possibility. In doing this, we have elicited numerical estimates with a mean of p (the gardener is the murderer) = .2. The equivalent verbal judgments, calibrated as described in part 2.1 and re-translated, can be summarized by "it is rather unlikely that the gardener has killed the butler".

An alternative approach consists in the systematic partitioning of the scenario into many simple singular propositions containing conditional and unconditional evidence (Shafer & Tversky 1985). The probabilities of these propositions can be judged in a piecemeal fashion and these probability jugdments can then be accumulated according to the Bayesian algorithm into the overall evaluation. In a first experiment (for details see Zimmer & Körndle 1988), numerical estimates have been elicited and accumulated, with a Bayesian algorithm. This results in: p(the gardener is the murderer/nobody else killed the butler) =

$$\frac{p(\text{G is M})}{p(\text{G is M}) + p(\text{somebody else is M})} = \frac{0.3838}{0.3838 + 0.00078} = .98$$

This extremely high probability is mostly due to the very low probability for "somebody else is the murderer" which is computed from estimates in the range of 0.0 to 0.4. If however qualitative judgments are elicited, calibrated, and accumulated, the final result is "(probable) to (very probable)" for (G is M). This result includes the numerical value of .98 as a special case but it includes also more conservative estimates that take better into account the unreliability of the empirical estimates. What is most striking, however, is the huge discrepancy between the direct overall estimates (0.2 and "rather unlikely) and the accumulated estimates (.98 and "(probable) to (very probable").

4 Heuristics and Biases in Accumulating Evidence

If one probes in detail how many propositions are taken into account by subjects for their overall judgment, it turns out that there exist at least three distinct heuristics for the selection and the weighting of pieces of evidence.

(i) <u>The "odds" heuristic:</u> Here the proportion of propositions in favor and against a culprit is determined and the final judgment depends directly on this proportion. Variants of this heuristic are taking into account the evidential weights of the propositions or using only propositions with an above-threshold evidential weight.

(ii) <u>The "frame of reference" heuristic:</u> In this heuristic the first (about 3) substantial propositions from the scenario are used to build up a reference system into which all the later propositions are fit. Substantial in this sense are only those propositions containing strong positive or negative evidence. Neutral information does not influence this heuristic, therefore it seems to be different from mere memory effects.

(iii) <u>The "maximum" heuristic:</u> Here, the final overall judgment is derived from the proportion of the maximal evidential weights of the pro and con propositions. This heuristic uses least information and tends to lead to extremely unstable judgments.

The advantages and disadvantages of these heuristics are listed in Table 1.

If one varies the complexity of scenarios, it turns out that subjects start to apply these simplifying heuristics if propositions are conditioned on more than two statements or if more than 5 simple propositions have to be evaluated for the overall judgment.

Table 1:

Features of information processing / Heuristic	Quantity of propositions in the evaluations	Taking into account the weight of evidence	Unbiasedness
"odds"	high	low (assumption of equal weight)	+ high
"Maximum"	extremely low	low (only extreme values are taken into account)	high
"referential system"	low	moderate (only the first and the last propositions are taken into account)	extremely low

5 CONCLUSION

The failure of decision makers to take advantage of decision support systems can be traced back to the fact that DSSs demand information from the experts and DMs in a form that differs qualitatively from the form they normally use in argumentative discourse. Suggestions are made together with preliminary empirical results how this problem can be overcome. The main aspects are

(i) improvement of discourse design especially, error tolerance, and taking into account the inherent vagueness of propositions

(ii) adaptation of information demands to the knowledge of the DM (processing of verbal judgments of uncertainty, verbal feedback)

(iii) modelling of argumentative structure of the DM by the DSS.

If these aspects are implemented into an decision support system, this results in a specific task distribution between decision makers and decision support systems. Decision makers are very efficient in the evaluation of probabilities and costs for simple events and conditional events provided they can express their knowledge verbally and according to their normal argumentative structure. If these evaluations can be calibrated interactively, the decision support systems should accumulate and process the provided evidence or expertise according to Bayesian evaluation schemes, because the decision makers tends to apply simplifying or even biased heuristics if the complexity is high.

6 REFERENCES

Edwards, W. (1954) The theory of decision making. Psychological Bulletin, 41, 380-417.

Edwards, W. & Newman, J.R. (1982) Multiattribute evaluation. Beverly Hills, CA: Sage.

Einhorn, H. & Hogarth, R.M. (1981) Behavioral decision theory: Processes of judgment and choice. Annual Review of Psychology, 32, 53-88.

Henrion, M. (1982) The value of knowing how little you know: The advantages of a probabilistic approach to uncertainty in policy analysis. PhD Thesis, School of Urban and Public Affairs, Carnegie Mellon University.

Hollnagel E., Mancini, G., Woods, D.D. (Eds.)(1986) Intelligent Decision Support in Process Environments. Berlin: Springer.

Humphreys, P., Svenson, D., Vari, A. (Eds.) (1983) Analysing and aiding decision processes. Amsterdam: North-Holland.

Johnson-Laird, P.N. (1983) Mental models: Towards a cognitive science of language, inference, and consciousness. Cambridge: Cambridge University Press.

Kahneman, D., Slovic, P., Tversky, A. (Eds.) (1982) Judgment under uncertainty: Heuristics and biases. Cambridge: Cambridge University Press.

Lichtenstein, S., Fischhoff, B., Phillips, L.D. (1982) Calibration of probabilities: The state of the art to 1980. In: D. Kahneman, P. Slovic, A. Tversky (Eds.) Judgment under uncertainty. Heuristics and Biases. Cambridge: Cambridge University Press.

Neumann, J. von, Morgenstern, O. (1944) Theory of games and economic behavior. Princeton: Princeton University Press.

Shafer, G., Tversky, A. (1985) Languages and designs for probability judgment. Cognitive Science, 9, 309-339.

Toulmin, S.E. (1964) The uses of argument. Cambridge: The University Press.

Wallsten, T.S., Budescu, D.V. (1983) Encoding subjective probabilities: A psychological and psychometric review. Management Science 29, 151-172.

Wetherill, G.B. (1963) Sequential Estimation of Quantal Response Curves. J. Roy. Statist. Soc., Series B, 25, 1-48.

Yager, R.R. (1983) Entropy and specificity in a mathematical theory of evidence. Int. J. of General Sys., 9, 249-260.

Zadeh, L.A. (1968) Probability measures of fuzzy events. Journal of Math. Analysis and Applications, 23, 421-427.

Zimmer, A. (1983) Verbal vs. Numerical processing of Subjective probabilities. In: R.W. Scholz (Ed.) Decision Making Under Uncertainty. Amsterdam: North-Holland.

Zimmer, A. (1984) A model for the interpretation of verbal predictions. Int. J. Man-Machine Studies, 20, 121-134.

Zimmer, A. (1986) What Uncertainty Judgments Can Tell About the Underlying Subjective Probabilities. In: L.N. Kanal & J. Lemmer (Eds.) Uncertainty in Artificial Intelligence. Amsterdam: North-Holland.

Zimmer, A., Körndle, H. (1988) Schematheoretische Begründungen für die Ordnung unsicheren Wissens. In: J. Krems & G. Heyer (Hrsg.) Wissensarten und ihre Darstellung. Ergebnisse eines Workshops des Arbeitskreises Kognition im Fachausschuß 1.2 der GI. Berlin/Heidelberg: Springer.