In M. Hahn and S.C. Stoness (Eds.) Proceedings of the 21st Conference of the Cognitive Science Society. 1999. pp.185-190.

Understanding probability words by constructing concrete mental models

David W. Glasspool (dg@acl.icnet.uk) and John Fox (jf@acl.icnet.uk)
Advanced Computation Laboratory, Imperial Cancer Research Fund,
61 Linclon's Inn Fields, London, England.

Abstract

We propose a model of the representation and processing of uncertainty and use it to account for data from an experimental study of the use of probability words. Given two sentences, one using a probability word and the other phrased in terms of reasons-to-believe, subjects were asked to judge if the second was an acceptable paraphrase for the first. For certain word/paraphrase pairs there was a high degree of consensus about acceptability, for others the subjects were divided. We model the decision process as involving two stages. First, a concrete "mental" model is constructed which is consistent with the first phrase. The second phrase is then tested for compatibility with this model. In simulations two different representations for the meanings of phrases were tested, one based on probability intervals, and one based on qualitative argument structures. Both versions of the model give a good account for the data, both in terms of which paraphrases are judged to be acceptable and the relative proportions of subjects agreeing or disagreeing.

Introduction

What is the meaning of probability terms (such as "probable" and "possible") as used in everyday language, and how are such concepts used in cognitive processing? The second of these questions - how such terms are used in processes such as decision making - has generally been seen as closely related to the first - their underlying cognitive representation. Historically, such terms have often been taken as conveying intervals of confidence or probability over some analogue scale, such as a probability scale or fuzzy membership functions (Wallsten & Budescu 1995). An alternative possibility is based on the view that human reasoning under uncertainty involves a process of logical argumentation, in which qualitative arguments for or against (or reasons to believe or doubt) a proposition are as important, or possibly more so, than representation in terms of quantitative values (Fox, 1994). On this view probability words may convey qualitative structures of such arguments rather than numerical degrees of belief. For example, the word "probable" might mean something more like "there are better reasons to believe this is true than to doubt it" than "the probability of this being true is greater than 0.5".

In this paper we present a model of a decision process which is applied to decisions about paraphrases for common probability words. On the model, probability terms are used by constructing a concrete (internal, or "mental") model of the world that is compatible with the term. Some more abstract representation of the meaning of the phrase - be it in terms of probability intervals, argument structures or some other formalism - is used to construct this world model, but it is the model that is the basis for the decision itself.

Before discussing the model and simulation work in more detail we first describe the experimental data on which it is based.

Experiment

The experiment reported here is designed to investigate the relationship between probability words and sets of arguments for or against propositions, with the aim of establishing a consistent set of terms for reports from risk-assessment software.

Subjects were presented with stimuli of the form "If (statement 1) then (statement 2)", as in the following examples:

If
it is TRUE that smoking causes cancer
then
There are better reasons to believe that
smoking causes cancer than to deny it

If
It can be ruled out that
benzoate derivatives are carcinogenic
then
it is PROBABLE that benzoate derivatives
are carcinogenic

They were asked to judge in each case if they agreed with the "then" statement given the "if" statement. Stimuli were presented on a computer display, and subjects responded by pressing one of three buttons marked "Agree", "Disagree" and "Unclear". They were asked to disregard any opinions they might have on the truth of the statements and to concentrate only on the consistency of the second with the first. In every stimulus one of the statements used a probability word, and the other was phrased in terms of "reasons to believe". Both words and phrases were selected from a set of five possibilities (Table 1), giving 25 possible combinations. Additionally each pair of statements was presented in both orders - with the probability word first and "reasons to believe" phrase second, and vice-versa, yielding a total of 50 stimuli. These were all presented, in random order, to 33 undergraduate students who participated for credit on a psychology course at

Table 1: The probability words and "reasons to believe" phrases used in the experiment. The latter are followed by the acronyms used to identify them hereafter.

Probability Words	"Reasons to believe"	
True	The reasons to believe (p) are totally convincing	(RBTC)
Probable	There are better reasons to believe (p) than to doubt it	(BRTB)
Possible	There is no reason to doubt (p)	(NRTD)
Improbable	There are better reasons to doubt (p) than to believe it	(BRTD)
False	It can be ruled out that (p)	(CBRO)

City University, London.

Results

Of the 1650 responses of 33 subjects on 50 stimuli, only 111 (6.7%) were "unclear". In the current work we focus on the "agree" and "disagree" responses only. Table 2 shows the proportion of such responses to each stimulus which were "agree" rather than "disagree".

Table 2: Experimental results: Proportion of subjects responding "agree" in (a) the "Term then Phrase" condition and (b) the "Phrase then Term" condition. Entries show the quantity (agree responses) / (agree + disagree responses). Any "unclear" responses are not counted.

(a)

	RBTC	BRTB	NRTD	BRTD	CBRO
True	0.94	0.88	0.88	0.06	0.03
Probable	0.26	0.84	0.29	0.09	0.03
Possible	0.14	0.71	0.17	0.2	0
Improbable	0	0.22	0.03	0.87	0.19
False	0.06	0.07	0.09	0.84	0.84

(b)

	RBTC	BRTB	NRTD	BRTD	CBRO
True	0.9	0.37	0.91	0	0.12
Probable	0.79	0.97	0.79	0.26	0.03
Possible	0.69	0.96	0.68	0.66	0.13
Improbable	0.03	0.17	0.19	0.81	0.57
False	0	0.06	0.15	0.34	0.93

A number of interesting features emerge. On some stimuli the subjects were quite consistent in their responses, on others they were clearly divided with as much as a 60%:40% split in opinion. The ordering of the statements in the stimuli is important - for example, 20% agreed that if something is "possible" then there are better reasons to believe it than to doubt it, whereas 66% agreed that if there are better reasons to believe than to doubt then it is still "possible". This asymmetry makes sense given the usual intuitive meanings for these phrases, but it is interesting to note that those stimuli which show such asymmetry do so to differing degrees.

Finally, we note that the phrase "There is no reason to doubt" is treated as very similar to "The reasons to believe are totally convincing". It seems that the experiment is not sensitive to the differences in meaning between the two phrases, which intuitively seem clear. We hope to probe such differences further in future experiments.

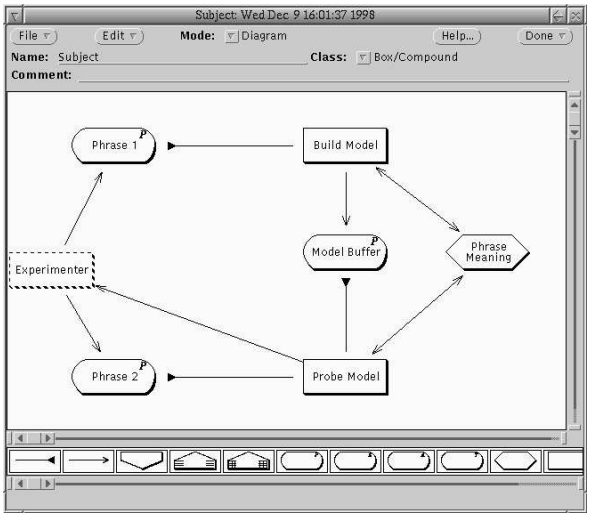

Figure 1: Overall structure of the decision model, as simulated in the COGENT modelling package.

A computational model

How might one characterise the cognitive processes involved in deciding whether one phrase is an acceptable paraphrase of another? Our approach is based on two central hypotheses:

1. The task is carried out by forming a symbolic internal ("mental") model of the first phrase, then testing the second phrase against it. If the second phrase is consistent with the established model then the paraphrase is accepted.

2. The model of a phrase employed in this process takes the form of a set of alternative possible situations in which the phrase would hold.

These assumptions are embodied in the decision-making model shown in Figure 1. The model is implemented using the COGENT cognitive modelling system (Cooper & Fox, 1998), which allows the components of the model to be fully specified so that its operation can be simulated.

The overall simulation contains models of both the task environment (labelled "experimenter") and the subject, although only the subject model is shown here. The task environment model is responsible for presenting pairs of phras-

Table 3: The set of possible situation models which could be produced under each representational scheme.

Probabilistic	Reason-based
0.0	
0.1	Confirm
0.2	Exclude
0.3	Support
0.4	Oppose
0.5	(Support, Oppose)
0.6	(Support, Support, Oppose)
0.7	(Support, Oppose, Oppose)
0.8	
0.9	
1.0	

Table 4: Meanings assigned to each phrase under the different representational schemes. * "NRTD" is equated with "True" here.

Phrase	Probabilistic	Reason-based
True	p = 1.0	Confirmed
Probable	p > 0.5	Support > Opposition
Possible	p > 0	Not excluded
Improbable	p < 0.5	Opposition > Support
False	p = 0.0	Excluded
RBTC	p = 1.0	Confirmed
BRTB	p > 0.5	Support > Opposition
NRTD	p = 1.0*	Confirmed*
BRTD	p < 0.5	Opposition > Support
CBRO	p = 0.0	Excluded

es equivalent to those used in the real experiment, and for recording results.

The layout of the subject model is intended to reflect hypothesis 1. The pair of phrases presented on a particular trial are placed in two storage buffers. Phrase 1 is the operative phrase from the first statement presented on that trial (for example *probable* in the sentence "It is probable that benzoate derivatives are toxic"). Phrase 2 is the operative phrase from the second statement. The process "Build Model" implements hypothesis 2 by constructing a set of situational models, all compatible with Phrase 1. To do this, "Build Model" communicates with another process, "Phrase Meaning", to check whether each of a standard set of candidate models is compatible with the phrase. Any which are compatible accumulate in a "Model Buffer". The final set of models in this buffer is taken to conceptualise Phrase 1. The approach has similarities with the "mental models" approach to reasoning (Johnson-Laird, 1983) to the extent that a formal proposition is represented by a set of concrete world models with which it is compatible.

The process "Probe Model" checks each situational model in the buffer for compatibility with Phrase 2. Again the "Phrase Meaning" process is used to perform the compatibility check. "Probe Model" responds to the experimenter according to the number of models in the buffer that are compatible with Phrase 2. If all models are compatible the paraphrase is accepted (the response "agree" is sent to the experimenter). If all are incompatible it is rejected (the response is "disagree"). If some are compatible and some incompatible the situation is unclear. Under either of the representational schemes discussed below 20% of stimuli result in this situation, whereas only 6.7% of subjects' responses are "unclear". It is therefore not obvious that such cases should be interpreted as "unclear" responses. We return to this point below.

To implement the situational models themselves two different representational schemes were investigated, which we label "probabilistic" and "reason-based". Table 3 shows the set of candidate situational models which could be considered under each representational scheme. Each situational model is intended to represent a single possible state of the world (we refer to these as "possible worlds", although the term is not used in its formal sense). The set of models accumulated in the Model Buffer represents a set of possible worlds in each of which Phrase 1 would be true.

We will first consider the probabilistic scheme, which is perhaps the more compatible with traditional ideas about the meaning of probability words. Here each situational model comprises a single quantity, which is taken to represent the probability that an event will occur in a particular possible world. We will describe the operation of the model and the results obtained using this representational scheme, returning to the reason-based scheme later.

The "Phrase Meaning" process checks the compatibility of a particular situational model with a particular phrase. It contains a definition of the meaning of each phrase under the appropriate representational scheme, as shown in Table 4. For the probabilistic scheme the meanings of phrases are defined in terms of probability intervals, and a straightforward set of intervals is chosen for the five probability words. For consistency the "reasons-to-believe" phrases are also assigned probability interval meanings. The meaning assigned to "No Reason to Doubt" is the same as that for "Reasons to Believe are Totally Convincing", in response to the clear tendency of subjects to treat both phrases as equivalent in the experiment. We return to this point in the discussion section.

Suppose Phrase 1 is "Probable". Under the probabilistic representational scheme the "Build Model" process would build a concept for the phrase "Probable" comprising the set of probability values {0.6, 0.7, 0.8, 0.9, 1.0}. That is, all candidate models with value > 0.5. If Phrase 2 is "Possible" then the "Probe Model" process will test each of these values against the meaning for "Possible" (> 0), and find them all to be compatible. The response "agree" will be sent to the "experimenter".

Study 1

Using the probabilistic representational scheme the full set of 50 experimental stimuli was presented to the model. For an initial comparison with the subject data we examine those stimuli for which 90% or more of the subjects either agreed or disagreed with the paraphrase. There are 21 such stimuli (6 with most subjects agreeing and 15 with most disagreeing). In every one of these cases the model gives an agree response (where 90% or more of subjects agree) or a disagree response

(where 90% or more disagree). This confirms that the model predicts the decisions made by subjects in those cases where the subjects themselves agree on the response.

Many of the remaining stimuli result in a mixture of compatible and incompatible situation models in the model buffer. As mentioned earlier it is not obvious how these cases should be handled. In order to investigate this further the criteria for "agree" or "disagree" responses from the model were relaxed. An "agree" response was made if 50% or more of the situational models in the model buffer were compatible with Phrase 2, a "disagree" response otherwise. This allows all stimuli to produce a response. Under these conditions, every "agree" response from the model corresponds to a stimulus for which more than 50% of subjects agree, and every "disagree" response to a stimulus for which more than 50% disagree (In only one case does the model produce exactly 50% compatible models. This is in response to a stimulus to which more than 50% of subjects responded "agree").

Clearly the proportion of "compatible" models is an excellent predictor of subjects' responses at this coarse level of analysis. This result suggests a possible interpretation for cases with both compatible and incompatible models in the Model Buffer: The ratio of compatible to incompatible models may correspond to the ratio of subjects agreeing to those disagreeing. To test this idea Table 5 shows, for each of the 50 stimuli, the proportion of "compatible" situational models produced by the simulation. The proportions do indeed correlate strongly with the proportions of "agree" responses in the experimental data of Table 2 (Spearman's rho = 0.91, p<.001, one-tailed).

The average absolute difference between simulated and actual proportions of "agree" responses over the table as a whole is 0.12, and the maximum is 0.43. If we look only at stimuli which result in both compatible and incompatible models ("mixed" responses) in the simulation, the fit appears more uniform, with an average is 0.13 and a maximum of 0.2.

Study 2

Study 1 used a representational scheme based on simple probability intervals, which is compatible with established ideas about the meaning of probability phrases. To what extent are the results of the simulation dependent on the use of a quantitative representational scheme? In this study we adopt a different approach, based on qualitative "reasons to believe" or arguments for and against a proposition. This approach is based on the idea of logical *argumentation* as a model for reasoning under uncertainty (Fox et al 1992; Fox, 1994; Krause et al, 1994), a process in which qualitative arguments for or against (or reasons to believe or doubt) a proposition are weighed up in order to make a decision.

We classify the reasons one might have for believing or disbelieving a proposition into four classes, following the type of classification common in argumentation theory (Fox, 1994). *Confirming* and *excluding* reasons are those which establish beyond doubt that a proposition is true or false, respectively. *Supporting* and *opposing* reasons provide qualitative but inconclusive evidence for or against the proposition, respectively. Table 3 shows the candidate situational models which can be chosen from by the "Build Model" process using the reason-based representational scheme. This set of candidates

Table 5: Proportion of "compatible" models produced for each stimulus under the "probabilistic" representational scheme.

(a) "Term then Phrase" condition.

	RBTC	BRTB	NRTD	BRTD	CBRO
True	1	1	1	0	0
Probable	0.2	1	0.2	0	0
Possible	0.1	0.5	0.1	0.4	0
Improbable	0	0	0	1	0.2
False	0	0	0	1	1

(b) "Phrase then Term" condition.

	RBTC	BRTB	NRTD	BRTD	CBRO
True	1	0.2	1	0	0
Probable	1	1	1	0	0
Possible	1	1	1	0.8	0
Improbable	0	0	0	1	1
False	0	0	0	0.2	1

was chosen to give the full range of qualitatively different structures using the four classes of reason - those which simply confirm, exclude, support or oppose the proposition, those which offer qualitatively balancing support and opposition, and those with more supporting than opposing arguments or vice-versa. Table 4 shows the meanings assigned to phrases under this scheme. In this case the reason-based phrases are more obvious in their meaning than the probability words, but note again that "NRTD" is assigned the same meaning as "RBTC". We have attempted to assign reasonable meanings to probability words.

Table 6 shows predicted proportion of "agree" responses using this representational scheme. The correlation (Spearman's rho) with the experimental data (Table 2) is again 0.91 (p<.001, one-tailed). The table differs from Table 5 only for the 10 "mixed" responses, for which the average absolute difference when compared with the experimental data is lower than for study 1, at 0.07, still with a maximum of 0.21. The overall average difference is 0.11.

Discussion

How are we to interpret the fit between the proportion of "compatible" models produced by the simulation and the proportion of "agree" responses from subjects? This would make sense under the assumption that the concept for phrase 1 comprises not the full set of "possible worlds" in which that phrase would be true, but only one such world model (the actual choice might be influenced by factors such as the availability, simplicity or concreteness of models, for example). In other words subjects tend to represent a probability phrase with the first appropriate situational model which comes to mind - a process of "satisficing" consistent with ideas of bounded rationality (Simon, 1956, 1982, Gigerenzer and Goldstein, 1996). The important feature of the model which allows this fit to the data seems to be the representation of the first phrase as a *concrete example* of a situation compatible with the phrase, rather than a more abstract representation capable of capturing the full range of meaning of

Table 6: Proportion of "compatible" models produced for each stimulus under the "reason-based" representational scheme.

(a) "Term then Phrase" condition.

	RBTC	BRTB	NRTD	BRTD	CBRO
True	1	1	1	0	0
Probable	0.33	1	0.33	0	0
Possible	0.17	0.5	0.17	0.33	0
Improbable	0	0	0	1	0.33
False	0	0	0	1	1

(b) "Phrase then Term" condition.

	RBTC	BRTB	NRTD	BRTD	CBRO
True	1	0.33	1	0	0
Probable	1	1	1	0	0
Possible	1	1	1	0.67	0
Improbable	0	0	0	1	1
False	0	0	0	0.33	1

the phrase. We assume that some such abstract meaning is nonetheless available at some level to allow the selection of a representative situation in the first place, and to test the second phrase against it.

An important issue in the simulation is the choice of candidate models and meanings assigned to phrases. Both clearly influence the number of compatible and incompatible models accumulated in the model buffer, which in turn determines the predicted proportions of "agree" and "disagree" responses from the simulation. As far as candidate models are concerned, the choice for the probabilistic representational scheme appears reasonable and the only real degree of freedom here would be to alter the grain size of the point probabilities available (giving a distribution in the limit), not their range. This should have no effect on the predicted proportions. The choice of candidates for the reason-based scheme was intended to capture the minimum set of qualitatively different argument structures. Various alternatives could be considered, for example including *excluding* or *confirming* arguments in the same models as *supporting* or *opposing* arguments, and this might alter the predicted proportions. In the absence of a principled procedure for selecting argument structures, however, it seems unreasonable to depart from the minimal set we have used.

For the probabilistic scheme the set of phrase meanings we have used are very simplistic probability intervals. These were chosen to give a neutral first approximation to the intended meanings of the phrases rather than with any empirical evidence in mind. There are however empirical results concerning subjects' willingness to assign particular probability values or intervals to various phrases, and this evidence could be used in a more principled version of the model. Changing the intervals would undoubtedly change the resulting predicted proportions. There would appear to be far less latitude possible in the selection of reason-based meaning definitions, which are essentially qualitative in nature. The results from study 2 can accordingly be considered more robust than those from study 1.

Table 2 suggests that subjects treat the phrase "No reason to doubt" as substantially equivalent to "The reasons to believe are totally convincing", despite the fact that intuitively the phrases do have different meanings. In both simulation studies the meanings of the two phrases have thus been made identical. It is not obvious how "no reason to doubt" would otherwise be represented on the probabilistic scheme, but there is a clear candidate meaning for the phrase under the reason-based scheme - intuitively it should correspond to an absence of opposing (and excluding) arguments. We assume that the current experimental task is insufficiently demanding to bring out any differences between the phrases, and we intend to investigate this anomaly further in subsequent studies. Another area which we will follow up in further work is the incidence of "unclear" responses from subjects. A larger study should provide a more reasonable number of these for analysis.

Study 2 shows that a representational formalism based on qualitative "argument" structures gives at least as good a fit to the data as one using quantitative probability values. This parallels findings from other modelling work in human decision making (Fox and Cooper, 1997; Cooper and Fox, 1997; Yule, Cooper and Fox, 1998) and is interesting in the light of claims that a formal theory of decision making under uncertainty based on a logic of argumentation (Fox et al 1992; Fox, 1994; Krause et al, 1994) may provide a more natural basis for understanding human decision making than traditional normative statistical approaches.

Conclusions

Both versions of the model give a good account of the data, both in terms of which paraphrases are judged to be "correct" (including the effect of order of presentation), and the relative proportions of subjects agreeing or disagreeing. The effect of the order of the two statements in the stimulus (phrases 1 and 2) can be qualitatively understood from the point of view of their logical interdependencies, but the strength of the approach presented here is that it gives a quantitative prediction for the size of the effect for different stimuli, based on the proportion of "compatible" and "incompatible" situational models generated. We conclude that the two hypotheses on which the model is based are appropriate, and we take our results as suggestive that subjects use a single concrete example to represent a probability phrase for the purposes of comparison (a result consistent with ideas of satisficing in "mental models" approaches to reasoning; Evans & Over, 1996).

The use of qualitative "argument" structures in the simulation provides at least as good a fit to the data as the use of more traditional quantitative probability values. The model is thus compatible with a view of of reasoning and decision making as involving qualitative argumentation, which we believe may provide a more natural basis for understanding these cognitive processes than quantitative statistical approaches.

Acknowledgments

The authors would like to thank David Hardman and Peter Ayton for their help in carrying out the experiment, and Andrew Coulson and Richard Cooper for their comments on the manuscript. This work was supported by an award from

the UK Economic and Social Research Council (award no. L127251011) under the Cognitive Engineering Programme.

References

Cooper, R. & Fox, J. (1997). Learning to make decisions under uncertainty: The contribution of qualitative reasoning. In Langley, P. & Shafto, M. (Eds.), *Proceedings of the 19th International Conference of the Cognitive Science Society.* Madison, WI. Cognitive Science Society Incorporated.

Cooper, R. & Fox, J. (1998). COGENT: A visual design environment for cognitive modeling. *Behavior Research Methods, Instruments & Computers, 30*, 553-564.

Evans, J.St.B.T. & Over, D.E. (1996) *Rationality and Reasoning.* Hove: Psychology Press.

Fox, J. (1994). On the necessity of probability, reasons to believe and grounds for doubt. In G Wright and P Ayton (eds) *Subjective Probability*. Chichester: John Wiley.

Fox, J. & Cooper, R. (1997): Cognitive processing and knowledge representation in decision making under uncertainty. In Scholz, R. W. & Zimmer, A. C. (eds.), *Qualitative Aspects of Decision Making.* Pabst, Lengerich, Germany.

Fox, J., Krause, P. & Ambler S. (1992). Arguments, Contradictions and Practical Reasoning. In: Neumann B (Ed) ECAI92, Vienna, Austria. *Proceedings of the 10th European Conference on AI*.

Gigerenzer, G. & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103*, 650-669.

Johnson-Laird, P. N. (1983). *Mental Models.* Cambridge: Cambridge University Press.

Krause, P., Fox, J. & Judson P. (1994). An Argumentation Based Approach to Risk Assessment. Presented at IMA conference on Risk: Analysis and Assessment, 14-15 April '94. In *IMA Journal of Mathematics Applied to Business and Industry, 5*, 249-263.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63*, 129-138.

Simon, H. A. (1982). *Models of bounded rationality*. Cambridge, MA.: MIT Press.

Wallsten, T. S. & Budescu, D. V. (1995). A review of human linguistic probability processing: General principles and empirical evidence. *Knowledge Engineering Review, 10*, 43-62.

Yule, P., Cooper, R. & Fox, J. (1998). Normative and Information Processing Accounts of Decision Making. In Gernsbacher, M. a. & Derry, S. J. (Eds.), *Proceedings of the 20th Annual Conference of the Cognitive Science Society*. Madison, WI. Cognitive Science Society Incorporated.