Semantics and Pragmatics of Vague Probability Expressions

Bernhard Kipper
Graduate School of Cognitive Science
Dept. of Computer Science
University of Saarbrücken
P.O. Box 151150
66041 Saarbrücken
Germany
kipper@cs.uni-sb.de

Anthony Jameson
SFB–314
Dept. of Computer Science
University of Saarbrücken
P.O. Box 151150
66041 Saarbrücken
Germany
jameson@cs.uni-sb.de

Abstract

Two experiments assessed the membership functions that German speakers assign to 12 adverb phrases and 17 modal verb forms that express probability assessments. These expressions fall largely into three rather homogeneous classes. The membership functions are used as part of the semantic knowledge base of the natural language dialog system PRACMA, one of whose purposes is to model pragmatic and contextual influences on the use of vague expressions. The system's normative model accounts for the role, in the selection and interpretation of vague probability expressions, of the listener's prior expectations, the speaker's dialog motivation, and the expressions that the speaker could have used but did not.

In an old episode of the television series "L. A. Law", adoption agent Ross tells Ann and Stuart that they may be forced to give up their adopted daughter, because her natural mother is suing to reclaim her.

Ann: You never told us she might do that!

Ross: I told you it *can* happen. I said it ***hardly ever*** happens.

Stuart: No, no! What you said was that the *chances were 1 in 100*. Now this is the second damned time: Something that *never* happens is happening to us!

This exchange illustrates several results that have emerged from extensive experimental studies of the use of verbal and numerical expressions of probability judgments. People tend to prefer verbal to numeric formulations when they *express* probability judgments, although this preference is markedly less strong when it comes to *receiving* the judgments (Wallsten, Budescu, Zwick, & Kemp, 1993). And the same probability (here, 1%) can be expressed verbally in quite different ways by different speakers, which makes the interpretation of such expressions a challenging and error-prone task.[1]

The example also illustrates two points that have received little attention in research to date. First, in everyday discourse, probability assessments are often expressed using modal verbs (as in "it can happen"), as well as the adverbs whose meaning has usually been studied. Second, the motivation of the speaker influences his or her choice of a vague probability expression in important and predictable ways; for example, the different paraphrases by Ross and Stuart do not represent random interindividual variation. In experiments, to the extent to which speaker motivation has been identifiable at all, it has generally consisted in the desire to convey as accurate as possible an impression of the probability the speaker has in mind.

The present paper aims to shed new light on both of these last two issues, as well as on some previously documented empirical phenomena. In the next section, we report on a pair of experiments in which the probability implications of adverb phrases and modal verb forms, respectively, are investigated with German-speaking subjects. These experiments were conducted in a neutral context designed to yield data that could be used to explain differing uses and interpretations of the expressions in different contexts. In the subsequent section, we present an idealized computational model of the selection and interpretation of vague probability expressions that takes speaker motivation (and its perception by the listener) into account. In the final discussion, we consider some empirical results reported by other researchers in the light of the model.

Empirical Derivation of Membership Functions

Although the most novel aspect of these new experiments is their inclusion of modal verbs, to ensure comparability the first experiment looks at German-language adverb phrases such as those used in previous (mainly English-language) research. These phrases satisfy the condition that they can express probability estimates when they are used alone as well as when they are used in combination with modal verbs.[2]

The Adverb Experiment

Method. Subjects were 24 graduate and undergraduate students, all native speakers of German. The subjects were asked to imagine that they were spectators at a game of chance. In this game, one of eleven wheels of fortune is spinned (cf. Wallsten, Budescu, Rapoport, Zwick, & Forsyth, 1986). The wheels differ widely in the sizes of their black and white portions. A player wins if the arrow to the right of the wheel points into the black sector when the wheel stops. The subjects were asked to imagine that a player was asked by the emcee about the result of the game before the player could see where the wheel had stopped. Given a particular wheel and a particular adverb phrase, the subjects were to indicate how "realistic" they judged this phrase to be as an answer of the player in the given situation; subjects did this by placing a mark on a graphical rating scale (cf. Rapoport, Wallsten,

[1] Zwick and Wallsten (1989, p. 72) discuss evidence that communication failures of this sort may have contributed to the Bay of Pigs fiasco and the Challenger space shuttle disaster.

[2] At the time of this writing, we are analysing the results of a further experiment which looked at combinations of adverbs and modal verbs, to see how the interpretation of such an expression can be predicted on the basis of the meanings of its parts.

& Cox, 1987). Twelve hypothetical answers of the player were presented to the subjects. Each such statement used a different adverb phrase and had the same form as "Ich habe *vermutlich* gewonnen" ("I *presumably* won"). Along with each such statement, 11 wheels were presented, representing the probabilities 5%, 15%, ... , 45%, 50%, 55%, ... , and 95%. The order of the wheels was varied, and a cardboard mask was used to prevent subjects from seeing their ratings for more than one wheel at a given time.

Results. For each subject, a curve was obtained for each adverb phrase showing for each probability (corresponding to a wheel) the value assigned by the subject. These curves will be called *membership functions* (as in fuzzy set theory); they are comparable to the curves resulting from the experiments of Wallsten, Budescu, Rapoport, Zwick, and Forsyth (1986) and Zwick, Budescu, and Wallsten (1988) on English adverbs. The upper part of Figure 1 summarizes the data by giving, for each adverb phrase, a curve showing the *mean membership value* assigned by the subjects to each probability.[3]

The curves show that about half of the 12 adverb phrases investigated have monotonically increasing mean membership functions, and that the differences between these functions are rather small. The generally lower membership values for the first three phrases in the figure suggest that they are seen as being truly appropriate only for very high probabilities (above 95%).

Three of the phrases in a second group—*möglicherweise* (*possibly*), *vielleicht* (*maybe*), and *eventuell* (*perhaps*)—are mainly judged realistic when associated with medium probabilities. The membership values given for these adverbs are less often close to the extreme values of 0 and 1; and since intermediate membership values are less well-defined, these adverbs show a greater amount of disagreement among subjects (as reflected, e.g., in the average standard deviation of the membership values for a given probability). The membership function for the one low-probability expression included—*auf keinen Fall* (*no way*)—is approximately the mirror image of that for the opposite expression, *auf jeden Fall* (*in any case*).

The Modal Verb Experiment

Method. The same experimental arrangement as in the Adverb Experiment was used, but modal verb forms were used instead of adverb phrases. The statements presented had the same form as "Es *dürfte* der Fall sein, daß ich gewonnen habe" ("It *should* be the case that I won"). Of the 17 modal verb forms presented, 9 were negated. As before, 24 graduate and undergraduate students, native speakers of German, served as subjects.

Results. Here again, for each modal verb a mean membership function was derived (cf. the lower part of Figure 1). For the 8 nonnegated modal verb forms, the results are very similar to those of the Adverb Experiment: Of these forms, 5 show monotonically increasing membership functions, which differ mainly in that two forms—*muß* (*must*) and *wird* (*will*)—are judged to be generally less realistic than the others for prob-

[3] A mean membership function could conceivably have a shape that was atypical of the shapes of the membership functions for individual subjects; but inspection of the individual functions showed that this was not the case here, in spite of considerable differences among the individual functions.

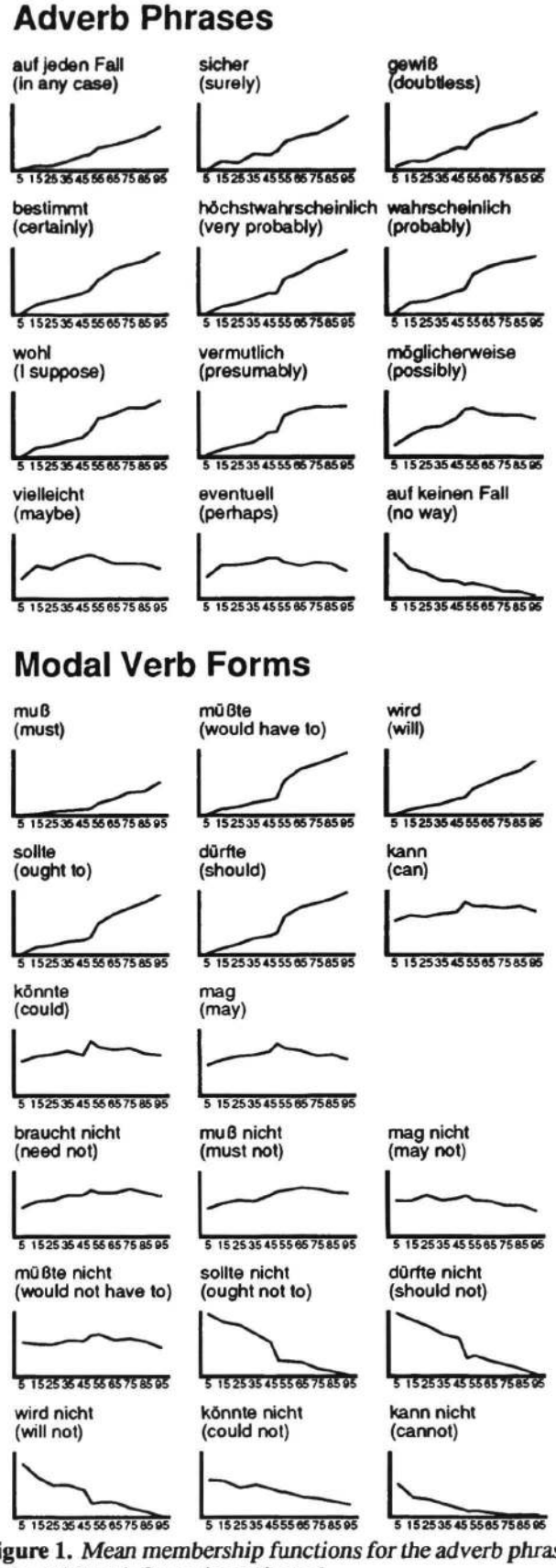


Figure 1. *Mean membership functions for the adverb phrases and modal verb forms investigated.*

abilities up to 95%. The other three nonnegated forms are single-peaked, but almost flat, being viewed as more or less realistic over a large part of the [0, 1] probability interval.

The negated modal verb forms show some different characteristics: The first 4 negated forms shown in the last three rows of Figure 1 were judged rather inconsistently (i.e., the membership function for a given subject in a number of cases exhibited several different peaks) and show virtually flat membership functions. The next 3 negated forms show membership functions that are the mirror images of those of the corresponding nonnegated forms. By contrast, *könnte nicht* (*could not*) and *kann nicht* (*cannot*) are monotonic even though the corresponding nonnegated forms are single-peaked (though almost flat).

Discussion of Experiments

The results of these two experiments suggest that, at least for speakers of German, the set of modal verbs does not constitute a differentiated set of expressions from which they can choose in order to express their probability assessments, even when it is supplemented with a sample of 12 adverb phrases. Although a total of 29 expressions were investigated, essentially three main membership functions emerged: two monotonic ones expressing low and high probabilities, respectively, and a much flatter single-peaked curve covering a large part of the interval [0, 1]. In addition, 4 of the negated modal verb forms appear not to be familiar to German speakers at all as expressions of probability, or to have unclear probability implications.

The apparent inefficiency of the set of modal verb forms investigated for expressing probability judgments becomes more understandable when one considers that this is not their sole linguistic function. As in English, each of the German modal verbs has at least one sense that does not directly involve probability, but rather concerns, e.g., permission or physical capability. Note also that even adverbs with similar membership curves—e.g., *vermutlich* (*presumably*) and *wahrscheinlich* (*probably*)—can have different semantic implications (cf. Teigen, 1988a).

Evidently, in order to be able to express many different assessments of probability, the speaker must make use of linguistic hedges (as in "It is very likely that ...") or combinations of adverbs and modal verbs (as in "It may perhaps be the case that ...").

To the extent to which these experiments can be compared to previous experiments on English-language expressions, the results are generally similar. For example, the noncomplementarity of the negated and nonnegated forms of some modal verbs was found in English for some adverbial forms (Lichtenstein & Newman, 1967). Some further relationships with previous results will be noted in the final discussion.

A Model of Pragmatic Processing

The experiments just described aimed to assess the basic and—it is hoped—relatively context-independent meaning of some vague probability expressions by creating an artificial task that called for much more primitive judgments than those involved in everyday speaking and comprehension (cf. Fillenbaum, Wallsten, Cohen, & Cox, 1991, p. 57). We will now present a computational model of how the basic meanings of such expressions, together with a number of contextual factors, determine how the expressions are actually used and understood. The model has been realized within the context of the natural language dialog system PRACMA (Ndiaye & Jameson, 1994), which simulates either the seller S or a potential buyer B in a dialog concerning S's used car.[4] To take a specific example, suppose that B asks "What are the chances that the car will pass the next inspection?" and that S has quite a definite idea about this probability, namely that it is only 25%. As the straightforward answer "25%" would presumably make a poor impression on B, S may look for some vague verbalization, such as "The car could pass the inspection". B, in turn, need not take this statement at face value but may wonder what S's assessment of the car's chances really is.

This situation raises two general issues: 1. What criteria can a speaker use to select a vague probability expression to verbalize a precise probability assessment? 2. What reasoning can the listener perform in an attempt to reconstruct the actual probability assessment of the speaker?

The computational model is best viewed as an idealized normative model of reasonable processing by S and B. It takes into account factors that appear to be relevant to the task, but in order not to distract from the basic logic, the model does not try to reproduce the typical style and limitations of everyday human thinking. The model is based on several assumptions which can be evaluated independently of the details of the implemented model and largely independently of each other. Some of the assumptions enjoy better empirical and theoretical support than others, as will become clearer below.

A basic strategy pursued by S, as modeled by PRACMA, is to make *comments* which alter in a way favorable to S the *impressions* that B has of various aspects of the car. We now address the following questions in turn: 1. When B interprets a comment C made by S, how does B form a corresponding *face-value impression*—i.e. an impression based solely on the meaning of C, without considering S's dialog motivation or the alternative comments she[5] might have made? 2. When S is looking for an appropriate verbalization of her probability judgment, how does she evaluate a possible comment on the basis of the face-value impression it is likely to create in B? 3. When B interprets a comment C, how does he reconstruct S's selection process so as to arrive at a pragmatically based interpretation of the comment? When discussing the processing by S or B in the following, we describe the way it is performed when PRACMA takes the corresponding role.

How B Interprets a Comment at Face Value

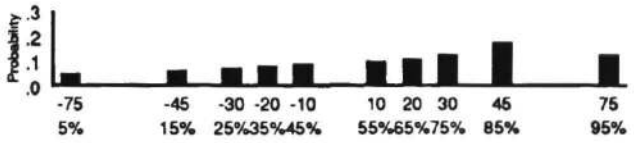
The concept of an impression. For each evaluation-relevant aspect of the car, B is assumed to distinguish some set of *possibilities* that might be realized by the car. For our example aspect "Chances of passing the next inspection", the 10 possibilities are the 10 equal-sized intervals on the scale from

[4] The way PRACMA handles vague probability expressions represents a generalization of some of the techniques used in the system IMP (Jameson, 1989). In IMP, all comments were assumed to be *crisp*, i.e., to have, in any, situation, a truth value of either 0 or 1.

[5] For clarity in exposition, we arbitrarily use feminine pronouns to refer to S and masculine pronouns for B.

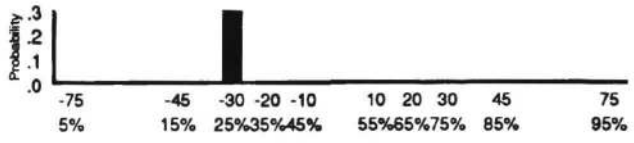
Prior Impression:

1.

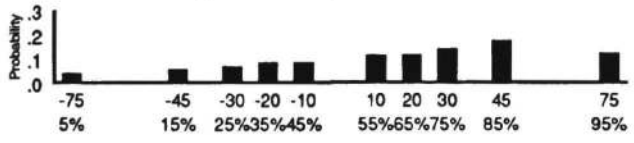


Face-Value Impressions:

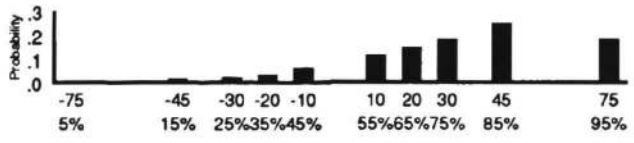
2. After "The chances are 25%"



3. After "It could pass the inspection"



4. After "It will presumably pass the inspection"



Impression Based on Pragmatic Interpretation:

5. After "It could pass the inspection"

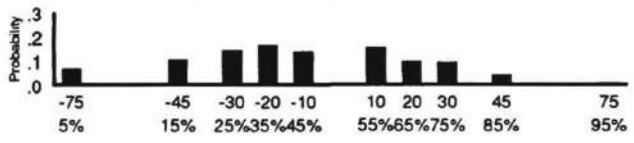


Figure 2. *Impressions involved in the generation and pragmatic interpretation of a comment.*

0% to 100%, each interval being referred to in terms of its midpoint. Thus the intervals correspond to the probabilities 5%, 15%, ... 95% that were examined in the experiments (cf. the lower row of labels in each histogram in Figure 2).[6] PRACMA uses as the *truth values* of a comment the corresponding mean membership values derived in our experiments (Figure 1).

Each possibility has for $\mathcal{B}$ a different *value*. For example, the upper row of labels in each histogram in Figure 2 reflects the assumption that $\mathcal{B}$, when evaluating the car, will add to its evaluation some number of "points" between −75 and +75, depending on its chances of passing the next inspection.

On the basis of general beliefs about cars, $\mathcal{B}$ has even at the beginning of the dialog some idea of how likely it is that the car's chances of passing the inspection lie within a given interval.[7] This *prior impression* of the car's chances is operationalized as a probability distribution. For example, Impression 1 in Figure 2 shows that $\mathcal{B}$ considers it a priori relatively probable that a used car of the type under discussion should have good chances of passing the inspection.

Face-value impressions. The procedure by which $\mathcal{B}$ derives a face-value impression is based on the following assumption: *Independent of pragmatic considerations and alternative comments, the likelihood that $\mathcal{S}$ will make comment $\mathcal{C}$, given a particular true possibility $\mathcal{P}$, is proportional to the truth value of $\mathcal{C}$ given $\mathcal{P}$.* This assumption allows $\mathcal{B}$ to derive his face-value impression by applying Bayes' Rule to his prior impression, using the truth values of the comment as relative likelihoods.[8] Impressions 2 through 4 in Figure 2 represent the face-value impressions that result when this rule is applied for three of the possible comments.

The above assumption can be seen as reflecting the general dialog convention that (a) statements that are false (i.e. which have a membership value of 0) should not be made at all, and (b) all other things being equal, a comment with a higher truth value should be preferred over a comment with a lower one.

How $\mathcal{S}$ Selects a Comment

The importance of the concept of a face-value impression lies in the fact that $\mathcal{S}$ can try to anticipate the face-value impression that a comment $\mathcal{C}$ would create in $\mathcal{B}$ as an initial assessment of the impact that $\mathcal{C}$ would have on $\mathcal{B}$. Of course, $\mathcal{S}$ cannot know exactly what face-value impression a given comment will create, because she doesn't know exactly, for example, what $\mathcal{B}$'s prior impression is. For simplicity of exposition, we now assume that $\mathcal{S}$ makes precise assumptions about these factors, and that these assumptions happen to be correct. How incorrect assumptions by $\mathcal{S}$ are taken into account will be discussed in the next section.

$\mathcal{S}$ determines a *rating* of the desirability of a comment $\mathcal{C}$ on the basis of the following two criteria:[9]

Favorability of the resulting face-value impression. Impression 3 in Figure 2 lies farther to the right than Impression 2, i.e. it is a more optimistic impression of the car's chances of passing the inspection; and Impression 4 is more favorable still. The mean of an impression can be used to quantify its relative favorability. Since $\mathcal{S}$ in PRACMA's dialog situation is *biased* toward creating favorable impressions, her rating procedure favors comments whose face-value impressions have relatively high means. (For an objective speaker, the rating would take into account how close the mean is to the value of the possibility actually realized.)

Truth value. Even a biased $\mathcal{S}$ is constrained by the convention that comments with low truth values are to be avoided. Accordingly, $\mathcal{S}$ includes in its rating of a comment $\mathcal{C}$ a *penalty*

[6] The results for the value 50% are not used in the computational model.

[7] From now on we use the term *chances* to refer to $\mathcal{S}$'s assessment of the probability that the car will pass the next inspection, so as to avoid confusion with the other probabilities that are involved in the processing of the comment.

[8] Note that this assumption does not presuppose that a fuzzy truth value is a sort of probability, only that it may in certain circumstances be reasonable to base a probability assessment on a fuzzy truth value.

[9] $\mathcal{S}$'s rating actually takes into account several other factors as well, such as the relevance of the comment to the current dialog focus, but as these factors are of minor importance for the present examples, they will be passed over here.

whose size is a positive function of the difference between 1.0 and its truth value. This penalty is high for the comment "It will presumably pass the inspection" (cf. Impression 4 in Figure 2), because that comment's truth-value is only .18 here; for the comment "It could pass the inspection" (Impression 3), the penalty is moderate, as the truth value is .63, and for "The chances are 25%" (Impression 2) the penalty is 0.

The comment "It could pass the inspection" emerges with the highest rating among all of the possible comments, because it combines a relatively favorable face-value impression with a reasonably high truth value.

How $\mathcal{B}$ Interprets $\mathcal{S}$'s Comment Pragmatically

When $\mathcal{S}$ has made a comment $\mathcal{C}$, it is not necessary for $\mathcal{B}$ to form only a face-value impression on the basis of $\mathcal{C}$. Since $\mathcal{B}$ is aware of the criteria that speakers use in selecting comments, he can attempt to reconstruct $\mathcal{S}$'s comment selection process and so to derive a more realistic impression of the true state of affairs that $\mathcal{S}$ verbalized using $\mathcal{C}$.

The core of this process of *pragmatic interpretation* by $\mathcal{B}$ is as follows: For each possibility $\mathcal{P}$ (i.e., each of the 10 intervals between 0 and 1), $\mathcal{B}$ adopts the perspective of $\mathcal{S}$ and rates each possible comment under the assumption that possibility $\mathcal{P}$ is realized. Let us denote with $\mathcal{D}$ the comment other than $\mathcal{C}$ which receives the highest rating. If $\mathcal{B}$ could be certain that her reconstruction of $\mathcal{S}$'s ratings were correct, $\mathcal{B}$ could, simply by comparing the ratings of $\mathcal{C}$ and $\mathcal{D}$, identify the likelihood that $\mathcal{S}$ would have created $\mathcal{C}$ as being either 1 (if the rating of $\mathcal{C}$ is greater than that of $\mathcal{D}$) or 0 (in the opposite case). In fact, of course, when reconstructing $\mathcal{S}$'s rating process $\mathcal{B}$ must make assumptions about all of the factors that enter into the process. As these assumptions will in general be more or less inaccurate, $\mathcal{B}$ associates considerable uncertainty with the ratings he ascribes to $\mathcal{S}$. He can take this uncertainty into account in a rough way by being cautious about the conclusions he draws on the basis of his comparison of the reconstructed ratings of $\mathcal{C}$ and $\mathcal{D}$. Specifically, he can assume that the conditional likelihood that $\mathcal{S}$ would have produced $\mathcal{C}$, given a particular possibility $\mathcal{P}$, is a positive monotonic function of the extent to which the reconstructed rating of $\mathcal{C}$ exceeds that of $\mathcal{D}$, approaching 1.0 only when $\mathcal{C}$'s reconstructed rating is much higher, approaching .0 when it is much lower, and taking the value .5 when the two ratings are equal.

After determining these conditional likelihoods, $\mathcal{B}$ applies Bayes' Rule to update his prior impression (i.e., Impression 1 in Figure 2). Impression 5 is the resulting, pragmatically derived impression. This impression is less optimistic than the corresponding face-value impression (Impression 3). The reason is that it seems very likely to $\mathcal{B}$ that $\mathcal{S}$ would have chosen some more positive-sounding comment if the true chances were higher than about 65%. In sum, by performing pragmatic interpretation, $\mathcal{B}$ is able to compensate in part for $\mathcal{S}$'s attempts to create a maximally positive impression.

Discussion of the Model

Influence of Prior Expectations

Wallsten, Fillenbaum, and Cox (1986) have demonstrated that listeners' interpretations of a vague probability expression can be strongly affected by their prior expectations concerning the probability in question (cf. also Zimmer, 1984). In the context of our scenario, this would mean that a $\mathcal{B}$ who for some reason was initially pessimistic about the car's inspection chances would, after hearing the comment "It could pass the inspection", end up with an impression which was farther to the left than Impression 5 in Figure 2.

Only one assumption of the present model is needed to account for this effect: the assumption that $\mathcal{B}$ derives his final interpretation of $\mathcal{S}$'s utterance according to Bayes' Rule, considering the conditional likelihoods that $\mathcal{S}$ would have produced $\mathcal{C}$, given various specific beliefs about the car's chances. Note that a variety of alternative theories are conceivable as to how $\mathcal{B}$ might assess these conditional likelihoods. Note also that this assumption is compatible with a wide variety of theoretical frameworks. In particular, it does not presuppose that the conditional likelihoods associated with an expression are a fixed aspect of its meaning. On the contrary, as the rest of the model illustrates, $\mathcal{B}$ may base his assessment of each conditional likelihood on his judgment of a number of aspects of the current situation, including $\mathcal{S}$'s dialog motivation.

That the above assumption accounts for the effect of prior expectations follows from the well-known fact that, when Bayes' Rule is applied, the resulting probability distribution depends in general on both the prior distribution and the set of conditional likelihoods. The impact of the prior distribution should be negligible only in special cases, e.g., when the conditional likelihoods are very low except within a limited range. Thus the above assumption fits well with the results of Wallsten, Fillenbaum, and Cox (1986), who noted that the effect of prior expectations was small or nonexistent for the expressions among the ones they investigated which were especially low in vagueness (pp. 584–585).

Differences in Assessments by Speaker and Listener

An experiment by Fillenbaum et al. (1991) asked subjects in two conditions to rate how "descriptive" a probability phrase was of a given probability. One condition (the *evaluation task*) was quite similar to the situation in our two experiments and can be seen as yielding membership functions with a comparable interpretation. In the other condition (the *selection task*), subjects took the role of speakers, selecting from a list of phrases the ones that would be suitable for expressing a given probability. Their ratings of "descriptiveness" can presumably be interpreted as assessments of the extent to which the phrase in question was relatively suitable for the purpose of describing the given probability to another person. As such, the ratings in this condition are comparable to the assessments of the relative desirability of a comment in PRACMA's model—except that the dialog motivation induced by the instructions was a purely cooperative one.

In the evaluation task, the membership functions elicited showed forms similar to those found in our experiments (Figure 1): Most were monotonic, except in the case of the phrase *possibly*, which for most subjects showed a single-peaked curve. By contrast, the curves summarizing the ratings of desirability from the speaker's point of view (elicited in the selection task) were most often single-peaked (except for the expression *improbable*) and they were narrower than the membership curves.

The assumption in our model which accounts for this difference is that the relative desirability of a comment for the speaker is not simply a function of the truth value of the comment itself; rather, it depends on the other possible comments that are available and takes into consideration, in addition to truth value, the location of the face-value impression created by the comment. For example, the phrase *likely* does not have a high relative desirability as a verbalization of a probability of 98%, in spite of its high truth value for that probability, simply because there are other expressions that do a better job of creating an impression of a very high probability.

Further Implications of the Model

Two implications of the PRACMA model can at present mainly be evaluated on the basis of their plausibility, although possible empirical tests readily come to mind. The first implication concerns the importance of the speaker's dialog motivation, both in the selection of comments by the speaker and in their interpretation by the listener. The second implication is that the willingness of a speaker to use an expression with a marginal truth value should depend on the extent to which this drawback of the expression is compensated for by desirable effect of the comment on the listener's impression.

Conclusion

Although our computational model is intended as an idealized normative model, it accounts for several phenomena that have either been documented in empirical research or are familiar from everyday experience. In particular, it explains the contrast between (a) the rather surprising lack of variety found in the basic semantics of the vague probability expressions examined here and (b) the more specific, but highly context-dependent interpretations that listeners and speakers attach to them in concrete situations. Adaptation of the model is required before it can be taken literally as a descriptive model, and such a model of course requires further empirical testing. Also, there are empirically established context effects on the selection of probability expressions about which the model as yet has nothing to say (cf. Gonzalez & Frenck-Mestre, 1993; Teigen, 1988b). Nonetheless, it appears that at the present stage of theorizing about the use of vague probability expressions, a normative model can have considerable heuristic value in the search for an appropriate descriptive theory, by drawing attention to important features of the task.

Acknowledgements

The research of both authors is being supported by the German Science Foundation, through the Graduate School of Cognitive Science at the University of Saarbrücken (the first author) and through the Special Collaborative Research Program on Artificial Intelligence and Knowledge Based Systems SFB 314, Project N1, PRACMA (the second author). Thanks are due to the three anonymous reviewers, who made a number of valuable suggestions for improving this manuscript.

References

Fillenbaum, S., Wallsten, T. S., Cohen, B. L., & Cox, J. A. (1991). Some effects of vocabulary and communication task on the understanding and use of vague probability expressions. *American Journal of Psychology, 104*, 35–60.

Gonzalez, M., & Frenck-Mestre, C. (1993). Determinants of numerical versus verbal probabilities. *Acta Psychologica, 83*, 33–51.

Jameson, A. (1989). But what will the listener think? Belief ascription and image maintenance in dialog. In A. Kobsa & W. Wahlster (Eds.), *User Models in Dialog Systems* (pp. 255–312). Berlin: Springer.

Lichtenstein, S., & Newman, R. (1967). Empirical scaling of common verbal phrases associated with numerical probabilities. *Psychonomic Science, 9*, 563–564.

Ndiaye, A., & Jameson, A. (1994). Supporting flexibility and transmutability: Multi-agent processing and role-switching in a pragmatically oriented dialog system. In P. Jorrand (Ed.), *Proceedings of the 6th International Conference on Artificial Intelligence: Methodology, Systems, Applications*. Singapore: World Scientific Publishing.

Rapoport, A., Wallsten, T. S., & Cox, J. (1987). Direct and indirect scaling of membership functions of probability phrases. *Mathematical Modelling, 9*, 397–417.

Teigen, K. H. (1988a). The language of uncertainty. *Acta Psychologica, 68*, 27–38.

Teigen, K. H. (1988b). When are low-probability events judged to be 'probable'? Effects of outcome-set characteristics on verbal probability estimates. *Acta Psychologica, 68*, 157–174.

Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R., & Forsyth, B. (1986). Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General, 115*, 348–365.

Wallsten, T. S., Budescu, D. V., Zwick, R., & Kemp, S. M. (1993). Preferences and reasons for communicating probabilistic information in verbal or numerical terms. *Bulletin of the Psychonomic Society, 31*, 135–138.

Wallsten, T. S., Fillenbaum, S., & Cox, J. (1986). Base rate effects on the interpretations of probability and frequency expressions. *Journal of Memory and Language, 25*, 571–587.

Zimmer, A. C. (1984). A model for the interpretation of verbal predictions. *International Journal of Man-Machine Studies, 20*, 121–134.

Zwick, R., Budescu, D. V., & Wallsten, T. S. (1988). An empirical study of the integration of linguistic probabilities. In T. Zétényi (Ed.), *Fuzzy sets in psychology* (pp. 91–125). Amsterdam: North-Holland.

Zwick, R., & Wallsten, T. S. (1989). Combining stochastic uncertainty and linguistic inexactness: Theory and experimental evaluation of four fuzzy probability models. *International Journal of Man-Machine Studies, 30*, 69–111.