ARI Research Note 90-57

Meanings of Nonnumerical Probability Phrases: Final Report

Thomas S. Wallsten
University of North Carolina

for

Contracting Officer's Representative
Michael Drillings

Basic Research
Michael Kaplan, Director

July 1990

FIFTY YEARS OF SERVICE
1940 - 1990

United States Army
Research Institute for the Behavioral and Social Sciences

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

JON W. BLADES
COL, IN
Commanding

Research accomplished under contract for the Department of the Army

University of North Carolina

Technical review by

Nehama Babin

Accession For	
NTIS GRA&I	X
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

NOTICES

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS --
2a. SECURITY CLASSIFICATION AUTHORITY -- 2b. DECLASSIFICATION / DOWNGRADING SCHEDULE --	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Research Memorandum No. 63	5. MONITORING ORGANIZATION REPORT NUMBER(S) ARI Research Note 90-57
6a. NAME OF PERFORMING ORGANIZATION L.L. Thurstone Psychometric Laboratory 6b. OFFICE SYMBOL (If applicable) - -	7a. NAME OF MONITORING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences
6c. ADDRESS (City, State, and ZIP Code) University of North Carolina Chapel Hill, NC 27514	7b. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600
8a. NAME OF FUNDING / SPONSORING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences 8b. OFFICE SYMBOL (If applicable) PERI-BR	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER MDA903-83-K-0347
8c. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 61101B; PROJECT NO. 74F; TASK NO. n/a; WORK UNIT ACCESSION NO. n/a

11. TITLE (Include Security Classification)

Meanings of Nonnumerical Probability Phrases: Final Report

12. PERSONAL AUTHOR(S)

Wallsten, Thomas S.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 83/08 TO 86/08	1990, July	106

16. SUPPLEMENTARY NOTATION

Contracting Officer's Representative, Michael Drillings

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Subjective probability; Linguistic probabilities; Fuzzy set theory; Nonnumerical probability phrases; Behavioral decision theory; Vagueness (Continued)

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This report summarizes three years of research on the meanings of nonnumerical probability phrases. The work is relevant to military needs because often the uncertainty of decisions is not well represented by the probability theory, but rather is imprecise, vague, or based on linguistic input. Techniques were developed and validated for representing the vague meanings of linguistic probabilities to individuals in specific contexts as membership functions over the (0, 1) interval. There are large, consistent individual differences in the meanings of probability phrases within a single context. Additional research investigated context factors that affect the meanings of such phrases, such as the available vocabulary, direction of communication, desirability of the forecasted events, and the base rates of the forecasted events. The researchers also summarized experiments that compare decision making in response to numerical and linguistic probabilities. Finally, a theory that handles virtually all the empirical results was outlined. This theory suggests how (Continued)

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL Michael Kaplan	22b. TELEPHONE (Include Area Code) (703) 274-8722	22c. OFFICE SYMBOL PERI-BR

DD Form 1473, JUN 86 Previous editions are obsolete. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED

ARI Research Note 90-57

18. SUBJECT TERMS (Continued)

Judgment
Choice

19. ABSTRACT (Continued)

the vague meanings of probability phrases are altered by context and integrated into single values to make judgments and choices.

CONTENTS

Meanings of Nonnumerical Probability Phrases

0. INTRODUCTION

Most real-world decisions are made in the face of uncertainty. Occasionally the data base is such that the uncertainty can be represented numerically by a forecaster, expert, or decision maker. More frequently, however, the information is sparse, incomplete, or vague, causing individuals to prefer expressing the uncertainty linguistically, with such terms as doubtful, probable, highly unlikely, and so forth. The research supported by this contract focused on how people understand and use linguistic expressions of uncertainty, with the ultimate aim of enhancing communication between experts, forecasters, and decision makers, and thereby improving the decision making process.

There were three specific goals to this project. The first was to develop and validate techniques for quantifying meanings of probability phrases in a given context, both in terms of the probabilities the phrases imply, and the vagueness with which they imply them. The second was to determine qualitatively the effects of certain context and individual characteristics on the meanings of these phrases. The third was to use the quantitative techniques developed as a first goal to understand the effects documented in service of the second goal. All three objectives were achieved. In addition the research has led to a tentative theory of judgment and choice on the basis of linguistic information, as well as to the means for investigating a new, interesting, and important hypothesis, namely, that human decision making may be more optimal in the presence of linguistic than numerical information. Finally, the techniques have been extended from linguistic probabilities to other linguistic variables, as well.

Because we have regularly reported our work in technical papers, and in order to facilitate communication of the research to other people within the military, this report is organized as follows. We first briefly discuss the relevance of the work to military needs. This is followed by a discussion of the scientific and theoretical issues addressed by the research. The subsequent sections provide a summary of the progress achieved during the term of the contract, with reference to the technical papers. An appendix gives a complete listing of all the papers, publications, and presentations stemming from the supported research.

1. RELEVANCE TO MILITARY NEEDS

For at least two reasons, many analyses or forecasts and much important information are available only in qualitative linguistic form. For example, one might hear that if conditions X, Y, and Z hold then it is likely that the Syrians will do A, but it is more probable that they will do B. Or, a battlefield commander might hear a report that the line to the south is relatively weak, but it is doubtful that it will fall within the next 12 hours. All of the terms that carry information in the preceding examples, likely, more probable, relatively weak, and doubtful, are imprecise, but, nevertheless, in most communication situations would be preferred over numerical counterparts. One reason for this preference is that frequently the information on which the forecasts or evaluations are made is itself not sufficiently precise that there is any natural way to translate it into probabilistic or numerical statements. Thus linguistic terms are used to reflect the nonnumerical nature of the data. Second, many people feel that even if they could translate this information into numerical form, to do so would be to suggest to the user of the information a level of precision and confidence that is inappropriate. Related to this second point is the feeling of many people that they understand and can respond to the information better when it is expressed in a verbal rather than a numerical form.

To make the point even stronger, it should be emphasized that there are circumstances in which people feel that numerical forecasts are appropriate. For example, probability estimates are commonly given when there are good relative frequency data or when probabilities might be estimated through the use of analytic aids such as fault trees. However, even in the latter case there is sometimes an unease with the resulting numbers, because one never knows whether all possible failure mechanisms have been considered. For example, the most sophisticated fault tree for chances of a melt-down at a nuclear power plant might not include the possibility of an operator spilling coffee on an important button. In short, numerical communication seems to be preferred when numerical information is available or can be estimated in a reasonable way, but verbal information is preferred when the available information is itself verbal, indirect, qualitative, or otherwise imprecise.

A particularly tragic example of how probability expressions are used and misused when historical data are sparse or virtually nonexistent was reported recently in Science (Marshall, 1986). According to the article, NASA had estimated the risk of a space shuttle crash as 1 in 100,000. The estimate was achieved by first having the top engineers at the Marshall Space Flight Center give their best judgment in verbal form of the reliability of all the components involved. Subsequently, the adjectival descriptions were converted to numbers. "For example, (Milton) Silvera (NASA's chief engineer in Washington) says, 'frequent' equals 1 in 100; 'reasonably probable' equals 1 in 1,000; 'occasional' equals 1 in 10,000; and 'remote' equals 1 in 100,000. When all the judgments were summed up and averaged, the risk of a shuttle booster explosion was found to be 1 in 100,000" (Marshall, 1986, p. 1596).

In view of the fact that verbal probability judgments will continue to be utilized in operational risk assessments, in part for the reasons described above, it is absolutely crucial that the use of such expressions be understood. Only with such understanding can recommendations be made and policies implemented for the effective communication and use of nonnumerical probability expressions.

There are many possible outcomes that may flow ultimately from this research. First, difficulties in verbal communication may become so well documented that users will agree that alternative methods must be devised. One possible alternative method would be to use upper and lower probability bounds, rather than words or point probability estimates.

A second possible outcome of this research is that verbal methods of communicating uncertainty will become sufficiently well understood that their use can be systematized. Algorithms or methods would be developed to convert different individuals' uses of words to a common base, which itself would communicate levels of both probability and vagueness. In this last regard, numerous systems analysts and artificial intelligence researchers have suggested in recent years that computer decision-support systems be developed based on fuzzy set theory to handle vague, imprecise, and fuzzy information in a systematic way that represents human processing of this information. The present research is directly relevant to the feasibility of such systems.

2. SCIENTIFIC RELEVANCE

The present research is relevant to a number of related scientific issues. One important issue concerns the mathematics of fuzzy set theory, which has developed rapidly since the pioneering paper by Zadeh (1965). The purpose of that work is to represent formally the vagueness or fuzziness that is inherent in much of human categorization, information, and decision making. Nevertheless, empirical research on the descriptive adequacy of fuzzy set models has been sparse and frequently of poor quality. (See the discussion in Wallsten, Budescu, Rapoport, Zwick, &

Forsyth, 1986.) Consequently, numerous questions remain to be answered, some of which have been addressed in the present research. First, can the construct of vagueness be measured in a reliable and valid fashion such that the resulting measurements ("membership functions" in fuzzy set theory) predict independent behavior? Assuming an affirmative answer, subsequent questions concern the shape of membership functions as well as the relative stability of meaning of various terms, expressed as such functions, and the shape of those functions. These questions are of interest to psychologists as well as to designers of expert systems to aid in decision and risk analyses. Currently, the designers of such systems enter membership functions according to their intuitions and to relatively arbitrary rules (Schmucker, 1984). However, the large, stable individual differences in the meanings of vague quantifiers clearly established in the present research means that the simple assumption frequently employed in expert systems, that the meaning of an expression can be represented in a unique way, is in error.

The fact that such an assumption is wrong would be of little surprise to most psychologists who are interested in language. It is well established, at least qualitatively, that meanings of expressions vary systematically with numerous context factors. By developing quantitative techniques for measuring the meanings of expressions that take their values over numerical bases, we are providing additional tools for this line of research. Furthermore, we are specifically developing information that will lead to a theory of how people understand nonnumerical probability expressions.

Finally, there is a vast body of research concerned with how people make judgments and choices in the face of uncertainty. Virtually all of this work has utilized numerically stated probabilities (and numerically stated outcomes) (e.g., see reviews by Einhorn & Hogarth, 1981; Pitz & Sachs, 1984; Rapoport & Wallsten, 1972; Slovic, Fischhoff, & Lichtenstein, 1977), yet, as argued above, these numerical representations may not reflect the most common situations that humans encounter. It has been suggested (e.g., Zimmer, 1983, 1984) that humans process information differently and more optimally when the information is presented in a verbal manner (which is consistent with the mode in which they normally think) than in a numerical manner (which is inconsistent with the normal mode). The present research provides a framework for investigating these claims, and for developing theories about how people make decisions in the face of ill-defined information.

3. VERBAL VS. NUMERICAL COMMUNICATION

In the introduction to this report and in many of our papers we claim that when information is sparse, vague, or otherwise incomplete, most people prefer verbal to numerical communication of uncertainty. It occurred to us rather late in the project that we should obtain data on this point. Thus we devised a simple questionnaire that we are still administering, but to which 37 people have thus far responded. The respondents were people who had completed participation in one or another of our studies. Twenty are from Chapel Hill, North Carolina, and 17 are native English speakers who live in Haifa, Israel (where we are conducting related experiments supported by the U.S.-Israel Binational Science Foundation).

The basic questions and responses are shown in Table 1. Note that 89% of the sample believe that most people prefer verbal communication while only 11% believe that numerical communication is preferred. In contrast, 76% of the sample prefer themselves receiving uncertainty numerically, while 72% prefer communicating to others in a verbal mode. Table 2 shows the cross tabulation on the latter two questions. Note that of subjects who prefer receiving numerical communication, 63% prefer communicating verbally to others. All of those who prefer receiving verbal assessments also prefer giving them.

While these results provide general support for our claim, their pattern is particularly fascinating. Insight into this pattern can be achieved by considering responses to the additional questions that we asked. Specifically, for the latter two questions, i.e., how do you usually prefer receiving communications and how do you usually prefer issuing them, we also asked why individuals have that preference, whether there are conditions under which they have the opposite preference, and if so what those conditions are. Responses to these questions were open ended.

A pattern clearly emerges when the American participants are categorized according to the cell of Table 2 into which they fall. (The Israeli responses will be treated similarly.) Those people who generally prefer both giving and receiving verbal

Table 1

Questionnaire Results

In your opinion, which mode of expresing uncertainty is usually preferred by most people in everyday life?

Numerical	4	(11%)
Verbal	33	(89%)

When you depend on other people's judgments of uncertainty, how do you usually prefer that they be communicated to you?

Numerically	28	(76%)
Verbally	9	(24%)

Which mode do you usually prefer to use when communicating your opinion to others?

Numerical	10. 5	(28%)
Verbal	26. 5	(72%)

Table 2

Preference for Communication

to you	to others		
	Num	Ver	Tot
Num	10. 5	17. 5	28
Ver	0	9	9
Tot	10. 5	26. 5	

communications also generally prefer the opposite when the data base warrants it. Similarly, but to a lesser extent, those people who generally prefer giving and receiving numerical communications opt for the verbal mode when numbers are totally inappropriate. Each of these two groups of subjects believes that their preferred communication mode is generally more understandable by people. Those subjects who prefer receiving numerical but giving verbal communications tend to invoke the same reasons as do the other two groups, but shift their emphasis according to whether they are the recipients or the issuers of the communication. In other words, there is considerable agreement as to what conditions warrant the use of either verbal or numerical communications, but some disagreement on which are the modal circumstances.

We are currently replicating this questionnaire on randomly selected subjects who have not been in our experiments. Our intuition is that the results will be similar. It should be of considerable interest to the Army to determine whether in operational contexts decision makers prefer receiving numerical estimates while experts and forecasters prefer issuing verbal ones, as our data suggest might be the case.

The second issue that we began to explore during the third year of the contract concerns factors that predispose individuals to issue verbal vs. numerical forecasts. An experiment was conducted in which subjects read scenarios about uncertain events, and were required to communicate the uncertainty as if to a friend. Our primary interest was whether they selected to communicate numerically or verbally. Specifically, there were 30 scenario topics, each of which occurred in 30 forms obtained by varying (but not orthogonally) four elements and two relations among elements that we thought might influence preference of communication mode. Each of 63 subjects responded to the 30 scenarios, each in a different form, in a Latin square design.

To understand the factors that were varied, consider two forms of one scenario: (i) In a certain class there are 70 students of different ages, 60 of them are 17 years old. If a student from this class is selected at random, what are the chances that he or she will be 17 years old? (ii) In a certain class there are many students of different ages, almost all of them are young. If a student from this class is selected at random, what are the chances that he or she will be around 17 years old? Elements 1 through 4 were whether (1) the population size was specified numerically or vaguely ("70" in i vs. "many" in ii above), (2) the event cardinality was specified numerically or vaguely ("60" in i vs. "almost all" in ii), (3) the event in the population was defined precisely or vaguely ("17 years old" in i vs. "young" in ii), and (4) the event in the query was defined precisely or vaguely ("17 years old" in i vs. "around 17 years old" in ii). Relations 1 and 2 were (1) whether (as in i but not ii) both population size and event cardinality were numerical, and (2) whether (as in i but not ii) the events in the population and the query were the same.

The results are generally consistent with, but go beyond, those from the questionnaire. An average of 39% of the communications were numerical, and individual differences were enormous. Over subjects, the proportion of numerical responses ranged from 0.02 to 1.00, with a standard deviation of 0.26. Fifteen of the subjects (24%) responded numerically over half the time, which compares favorably to the 28% of the questionnaire respondents who prefer communicating to others numerically (see Table 1).

The proportion of numerical responses was relatively constant over the 30 scenario topics (as we had hoped would be the case), but varied from 0.20 to 0.79 over the 30 forms. Considering only the four elements, Elements 1 and 2 accounted for the largest share of the variance in response proportion over forms -- 23% was uniquely associated with Element 1 and 21% was uniquely associated with Element 2 -- while Elements 3 and 4 accounted for almost none. When Relation 1 is considered with Elements 3 and 4, then Relation 1 is uniquely associated with 70% of the variance. In other words, as the questionnaire respondents indicated, the primary determinant of whether communication is numerical or verbal is whether of not numerical information is available. Analyses are continuing, and details will be given in the technical report.

However even at this stage of the work, it can be said that the results of the questionnaire and the experiment strongly support our earlier claims that people generally prefer communicating uncertainty to others in a verbal rather than a numerical manner. Nevertheless, when the nature of the information warrants it, they use the numerical mode. Interestingly and unexpectedly, though, there appears to be a general preference for receiving numerical rather than verbal communications about uncertainty.

4. MEASURING VAGUE MEANINGS OF PROBABILITY PHRASES

Most previous studies on the meanings of probabilistic phrases have had subjects give numerical equivalents to linguistic expressions. The universal result has been large intersubject variability in the numerical values assigned to probability terms and great overlap among terms. Within-subject variability is not small, but is considerably less than between-subject variability. The relevant studies are reviewed and references given in Budescu and Wallsten (1985) and in Wallsten et al. (1986). Although these results have generally been interpreted as demonstrating that linguistic probabilities have imprecise meanings that vary over individuals, a critic could reasonably argue that the variability is due to how people use and understand numbers rather than to how they use and understand words. We evaluated this criticism in an experiment (Budescu & Wallsten, 1985).

In that study 32 faculty and graduate students in the Psychology Department of the University of North Carolina at Chapel Hill rank ordered 19 probability phrases on each of three occasions. If the results of previous experiments were due to how people use numbers rather than to how they understand language, then everybody should have rank-ordered the phrases in approximately the same way in our study. The result, however, was quite the opposite. Individuals ranked the phrases consistently over time, but different individuals had very different rankings. An illustration of the results is provided in Table 3 taken from Budescu and Wallsten (1985). This table is based on the method of pair comparison, which is one of the ranking methods used in the experiment, and shows the probability that two randomly selected people will order the indicated pairs of probability words in the same direction. The words in the table are ordered according to their mean ranks. Note that in general the probability of agreement increases as one moves from the main diagonals of the table toward the lower left corners, indicating that the probability of agreement is roughly inversely related to the proximity of the two words in a pair. Of the 60 pairs of words, there is perfect agreement for only 23 (38%), while agreement is not better than chance for 15 pairs (25%). These data illustrate what we have come to call the "illusion of communication." Two people, communicating with a probability phrase, will each be relatively confident of what the phrase means, yet the meaning may be very different for each person.

Note also in Table 3 that agreement probabilities for pairs with the anchor words always, tossup and never are with a single exception 1.00 or 0.88. Thus, the meanings of these words are precise relative to the meanings of the other words. It was this set of observations, that the meanings of linguistic probabilities not only vary over individuals, but also are differentially precise within individuals, that lead to our attempt to represent the meanings of such phrases in terms of membership functions.

Several people (e.g., Watson, Weiss, & Donnell, 1979; Zadeh, 1975; Zimmer, 1983) have suggested that the meaning of a

Table 3

PROBABILITY THAT TWO RANDOMLY SELECTED PEOPLE WILL ORDER THE INDICATED PAIRS OF PROBABILITY WORDS IN THE SAME DIRECTION BASED ON PAIR COMPARISONS

	Group 1				
	Never	Rarely	Uncommon	Uncertain	Unpredictable
Rarely	1.00				
Uncommon	1.00	.88			
Uncertain	1.00	1.00	.78		
Unpredictable	1.00	.88	.88	.57	
Toss-up	1.00	1.00	1.00	.78	.78
	Toss-up	Probable	Likely	Often	Usually
Probable	1.00				
Likely	1.00	.70			
Often	.88	.57	.57		
Usually	1.00	1.00	.78	.57	
Always	1.00	1.00	1.00	1.00	1.00
	Group 2				
	Never	Improbable	Usually not	Seldom	Unlikely
Improbable	1.00				
Usually not	1.00	.50			
Seldom	1.00	.57	.51		
Unlikely	1.00	.61	.50	.51	
Toss-up	.88	.88	.88	.88	.88
	Possible	Toss-up	Predictable	Common	Frequently
Toss-up	.70				
Predictable	.63	.57			
Common	.88	.88	.57		
Frequently	1.00	.88	.63	.57	
Always	.88	1.00	.88	.88	.88

probability term can be represented by a function on the [0,1] probability interval, as illustrated in Figure 1. The function takes its minimum value, generally zero, for probabilities that are not at all in the concept represented by the phrase, its maximum value, generally 1, for probabilities definitely in the concept, and intermediate values for probabilities that have intermediate degrees of membership in the concept represented by the term. There are no constraints on the shapes such functions can have, nor must they be describable by particular equations. Within fuzzy set theory, such a function is called a membership function, but it is not necessary to tie the idea strictly to fuzzy set theory.

We have conducted three experiments to ascertain whether such functions can be established reliably and validly to represent the meanings of nonnumerical probability terms to individuals in specific contexts. Assuming positive results, subsidiary issues addressed by these studies included making some preliminary statements about the meanings of nonnumerical probability expressions, assessing the extent of individual differences in meanings, and developing a scaling technique that is easy to use. Two of the experiments are reported by Wallsten et al.(1986) and the third by Rapoport, Wallsten, & Cox (in press).

We have criticized previously published empirical techniques for establishing membership functions and have proposed a graded pair-procedure instead (Wallsten et al., 1986). The technique can be understood with the aid of Figure 2, which represents the

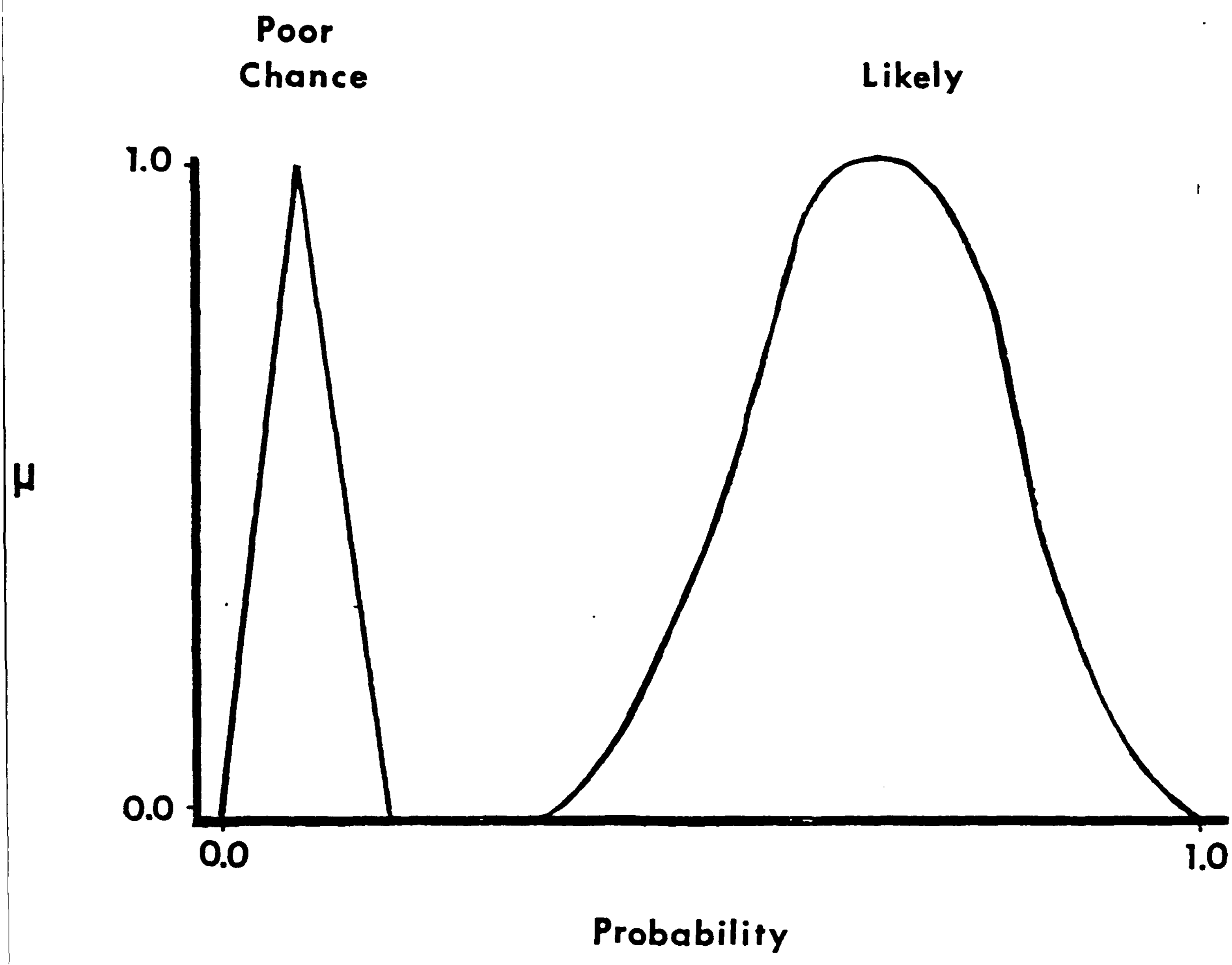


Figure 1. Illustrative membership functions. (From Wallsten, Budescu, Rapoport, Zwick, & Forsyth, 1986)

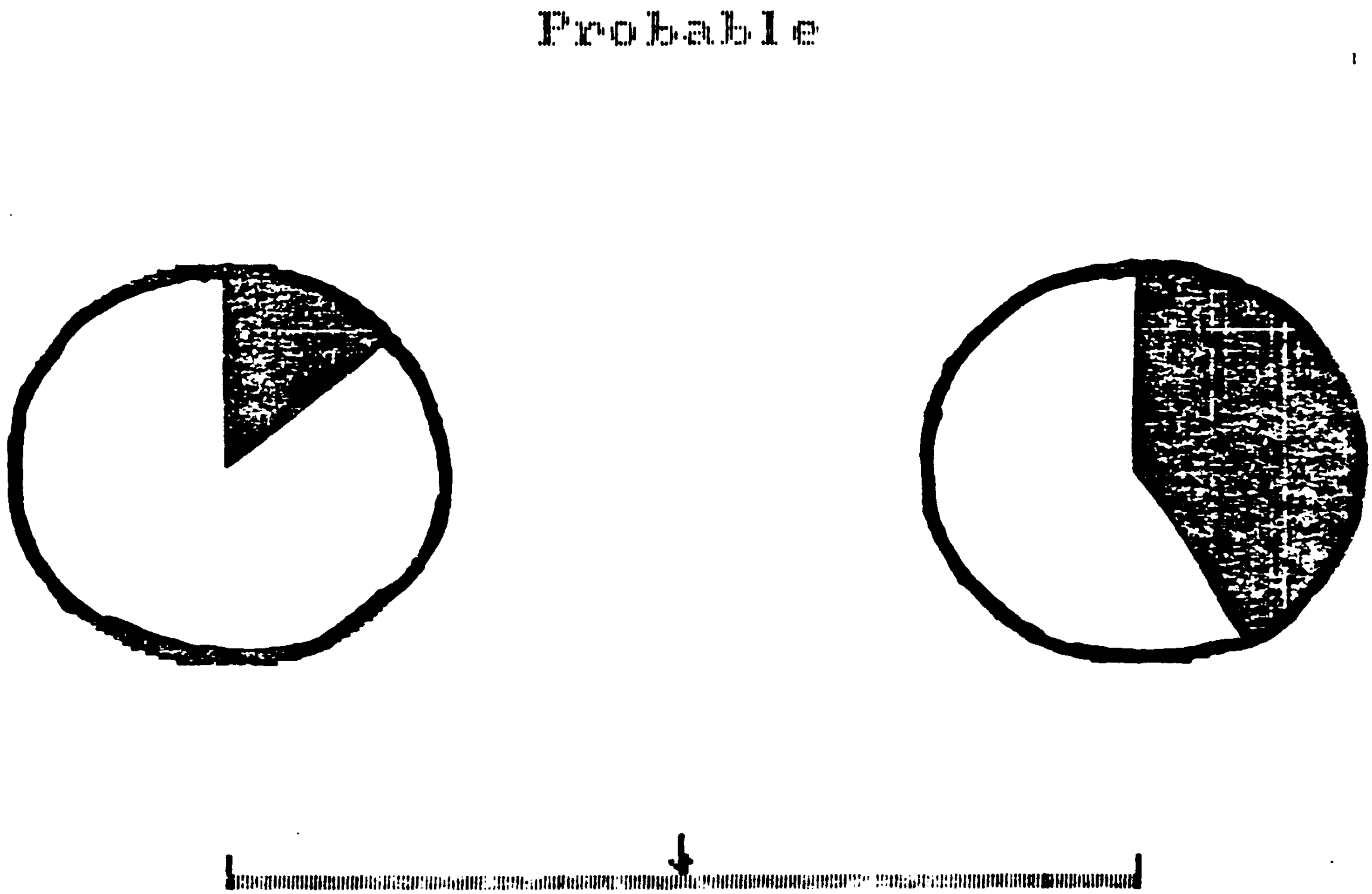
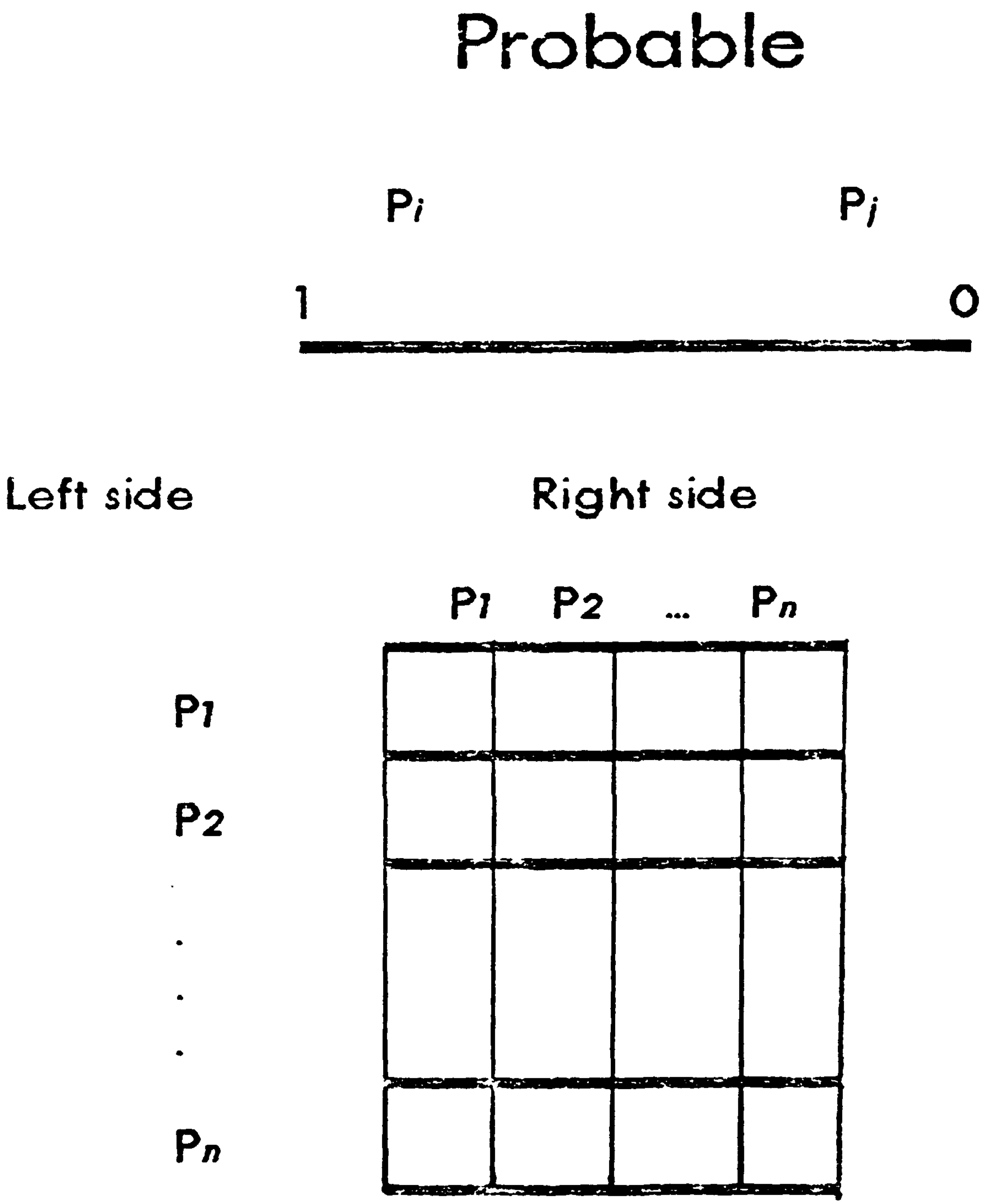


Figure 2. Sample computer display for pair-comparison procedure.

computer display seen by a subject at a beginning of a pair-comparison trial. The subject was instructed to consider the phrase at the top of the screen, probable in this case, as well as the two spinners, each of which represents a different probability of landing on white. The subject was asked "If you had to assign the phrase at the top of the screen to one of the two spinners to describe the probability of landing on white, to which spinner is it more appropriately assigned and how much more appropriate is the assignment of the phrase to that spinner than to the other one?" The subject was told to indicate his or her judgment by moving the arrow on the response line, specifically to "place the arrow so that its relative distance between the two spinners represents (the phrase's) relative appropriateness for the two probabilities."

Over the course of a session, each subject saw a number of phrases, and for each phrase the probabilities on the left and the right sides were manipulated in a factorial fashion. The design for one phrase is illustrated in Figure 3. The two spinners are shown generically by p_i and p_j, while the bottom of the figure illustrates the factorial design. If the response line is imagined to run from zero on the right to one on the left, as illustrated in the figure, then for any particular pair $(p_i p_j)$ the subject's response setting, expressed as a number, can be entered into cell $(p_i p_j)$ of the matrix. The cells of the matrix thus can be rank ordered according to the degree that the left side probability is better described by the phrase than is the right side probability. The basic data consist of such a matrix for each of the phrases under consideration.



$$P_iP_j \precsim P_kP_l \quad \text{iff} \quad R(ij) \lesssim R(kl)$$

Figure 3. Pair-comparison design for one phrase.

Detailed descriptions of how membership functions are derived from these matrices are given by Wallsten et al. (1986) and Rapoport et al. (in press), but the main ideas are illustrated in Figure 4. First, ordinal conjoint-measurement properties necessary for scaling are checked within the matrix. If the properties are satisfied, then metric scaling procedures are used to assign values to each probability such that the differences (or the ratios) of the row and column scale values for each cell are rank- ordered in the same manner as are the data. These values, scaled to [0,1], can be interpreted as membership values representing the degree to which each probability belongs to the vague concept denoted by the phrase. Finally, in the Wallsten et al. (1986) studies, the derived values were used to predict independent judgments, which the subject was shown one spinner and two probability terms (the display was just the converse of that in Figure 2, with the spinner at the top of the screen and the two probability phrases at either side). The subject was to move the arrow on the response line to indicate which phrase better described the spinner and how much better it did so. Thus, there were three validity checks: The judgments had to satisfy the ordinal conjoint-measurement conditions, the scaling procedures had to yield high goodness-of-fit measures, and the resulting scale values had to predict independent judgments.

The Rapoport et al. (in press) study differed from those of Wallsten et al. (1986) in that the former did not use the same types of independent judgments. Rather, we asked in that study whether membership functions derived from the graded pair-

Obtaining Membership Values

- If for a given phrase, the P x P response matrix satisfies ordinal conditions from conjoint measurement, then scale values, υ, can be assigned to the P_i:

  $P_iP_j \lesssim P_kP_l$ iff
  $\upsilon(P_i) - \upsilon(P_j) \leq \upsilon(P_k) - \upsilon(P_l)$.

  I.e., the differences (or ratios) of the row and column scale values for each cell are rank ordered as are the data.

- For each phrase, the $\upsilon(P_i)$ scaled to [0, 1] can be interpreted as membership values representing the degree to which P_i belongs to the vague concept denoted by the phrase.

- To further validate the interpretation, the values should predict independent judgments.

Figure 4. Procedure outline for obtaining membership function values from pair-comparison judgments.

comparison procedure would be similar to those obtained from a much simpler direct estimation procedure, as illustrated in Figure 5. Here, the subject was shown one phrase and one spinner and had to move the arrow on the response line to indicate how well the phrase describes the probability displayed on the spinner. The arrow could be moved from "not at all well" on the left to "perfectly well" on the right. Considering this response line to run from zero on the left to one on the right, membership functions were obtained directly by plotting the subject's judgment as a function of the spinner probability, with a separate curve for each phrase.

We compared the two scaling procedures for two reasons. First, the obtained function should be independent of the method used to derive it. If this result obtains, then that is further evidence bearing on the validity of the methods. Second, as a practical matter, the pair-comparison procedures are long and tedious whereas the direct estimation procedures are relatively simple and quick. If the two provide equivalent results, we are justified in using the latter in subsequent studies. It must be pointed out that we could not have begun with the direct estimation technique, as many other people have done, because there are few independent means for evaluating it. The pair-comparison procedure has the advantage of considerable internal constraint, thereby providing many opportunities for empirical testing. It is only through correlating the results of the two procedures that the direct estimation one itself becomes validated.

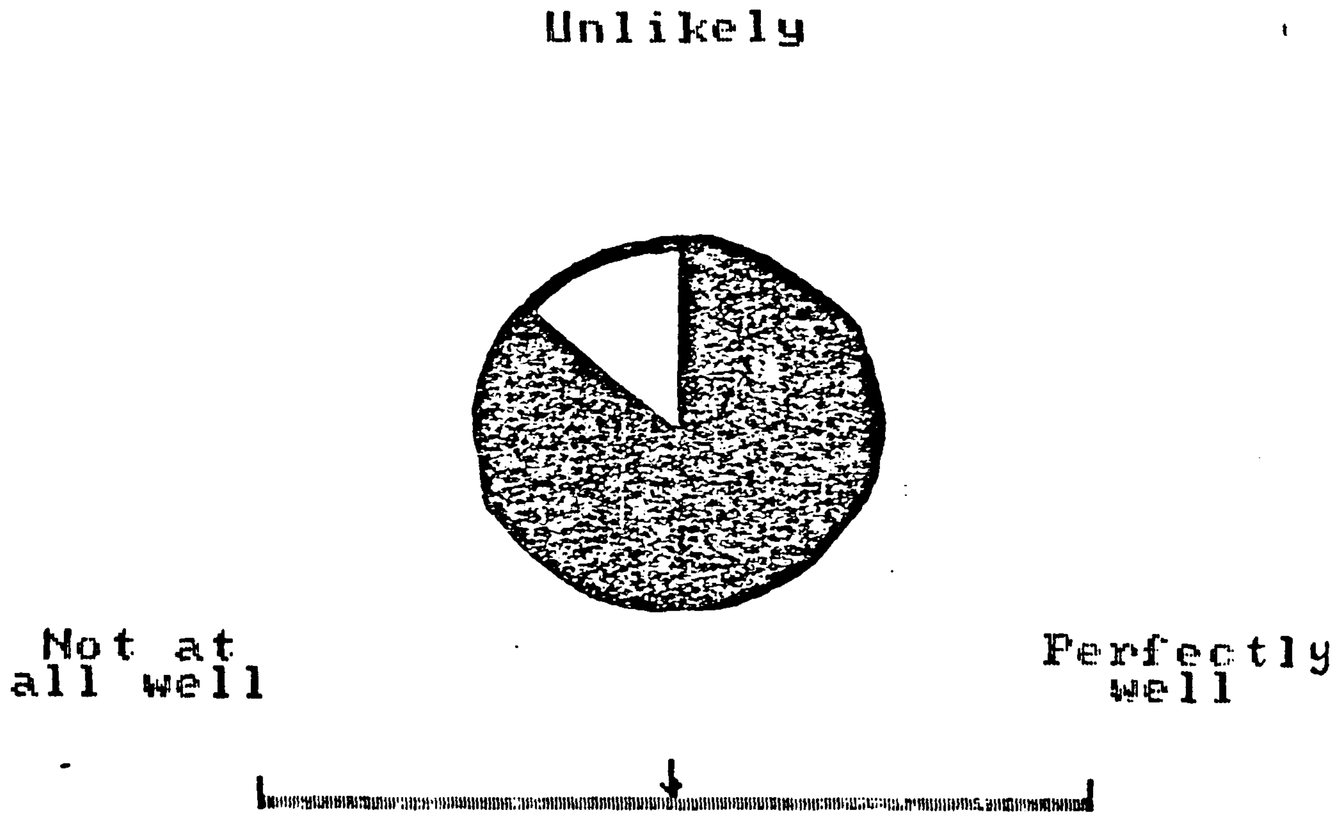


Figure 5. Sample computer display for direct estimation procedure.

The R popoport et al. (in press) experiment included a third measurement technique as well, as indicated in Figure 6. The subject was shown six spinners with each phrase and was to indicate which spinner was best described by the phrase. The selected spinner was then removed from the screen and the subject had to pick the best of the remaining spinners. In this fashion, the subject rank-ordered the six spinners with regard to how well they were described by the particular phrase. This procedure was included on the assumption that it was the easiest of all. Thus yet another validation measure was the degree to which the pair-comparison and direct estimation techniques predicted the observed rank orderings.

We turn now to a summary of the three experiments. The subjects in all cases (20 in Experiment 1 of Wallsten et al., 1986; a selected 8 of that group in Experiment 2; and 20 in Rapoport et al., in press) were social science and business graduate students, who were paid well for their time over three to five sessions. Judgments were highly reliable. The mean realiability correlation in Experiment 2 of Wallsten et al. (1986) was 0.90. In Rapoport et al. (in press) reliability correlations were not computed, but individual membership functions estimated from responses obtained on two separate occasions were very similar. (Certain features of Experiment 1 in Wallsten et al., 1986, precluded general reliability calculations in that study. See the paper for details.)

Overall the pair-comparison scaling method worked very well; the conjoint measurement axioms generally were well satisfied, the scaling model fit the data very well, and independent

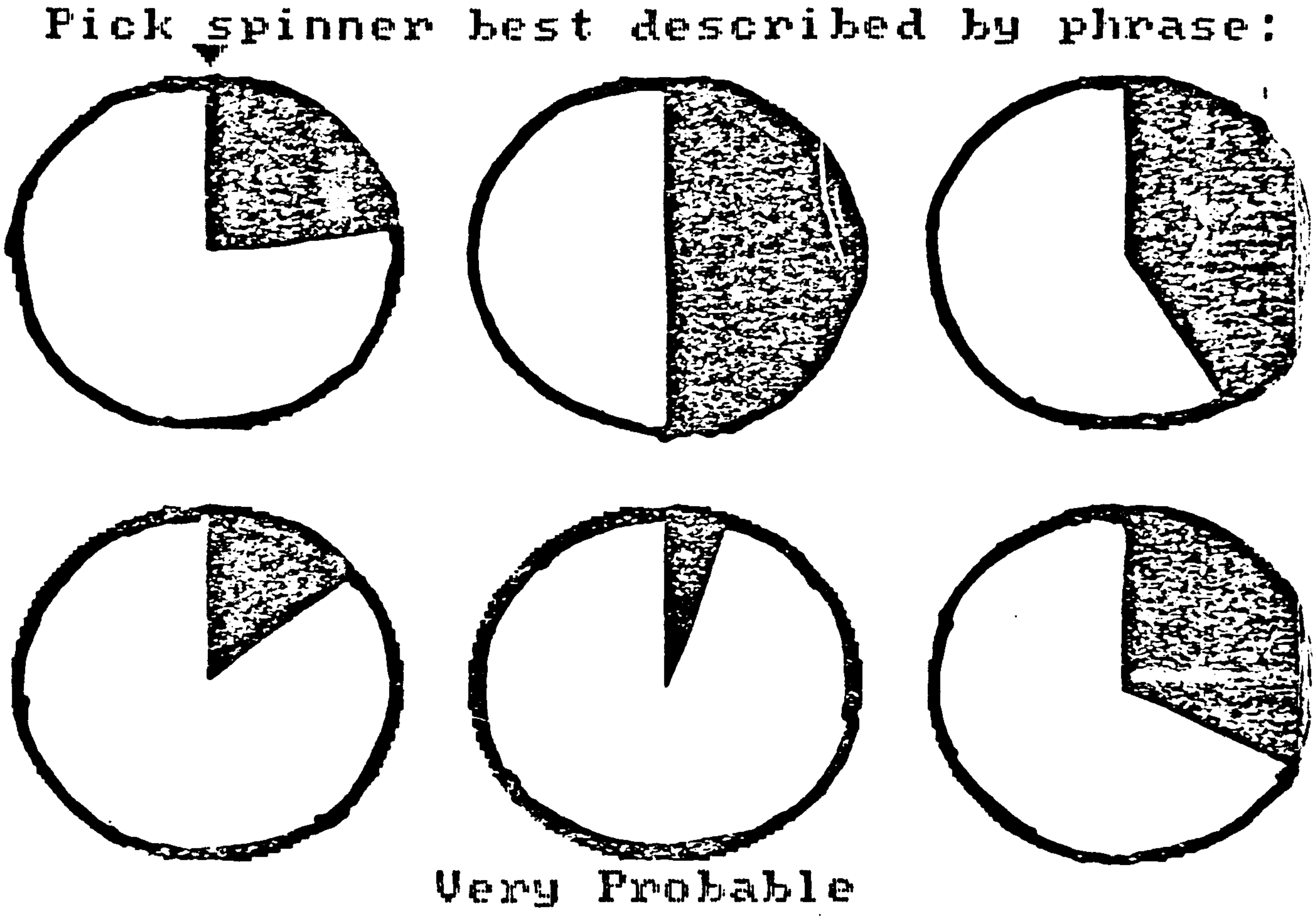


Figure 6. Sample computer display for rank-order procedure.

judgments were well predicted. For example, without estimating any free parameters, the mean correlations between scaling model values and judgments were 0.79 in experiment 1 of Wallsten et al., 0.95 in experiment 2 of Wallsten et al., and 0.84 in the second session data of Rapoport et al. (It should be emphasized that the subjects in experiment 2 of Wallsten et al. had been recruited from experiment 1 and therefore were very highly practiced).

In this report we will look only at the nature of the derived membership functions and the relation between functions obtained from the pair-comparison and direct estimation methods. Figure 7 shows in generic form the three kinds of membership functions that were obtained. Phrase 1 illustrates a monotonic decreasing function. Low probabilities were definitely represented by the phrase (i.e., have membership values of 1) and increasing probabilities have membership values decreasing to zero. Phrases 2 and 3 illustrate single-peaked functions, differing only in that one is roughly symmetric and the other is skewed. In these examples the central probabilities have maximal membership in the phrase with probabilities on either side having decreasing membership. Finally, phrase 4 illustrates a monotonic decreasing membership function.

Table 4 lists all the phrases that were used in the three experiments. Certain phrases are shown together because over all subjects there were no differences in the distributions of membership functions for them. (This does not mean that the phrases are necessarily synonymous to individuals.) Grouped

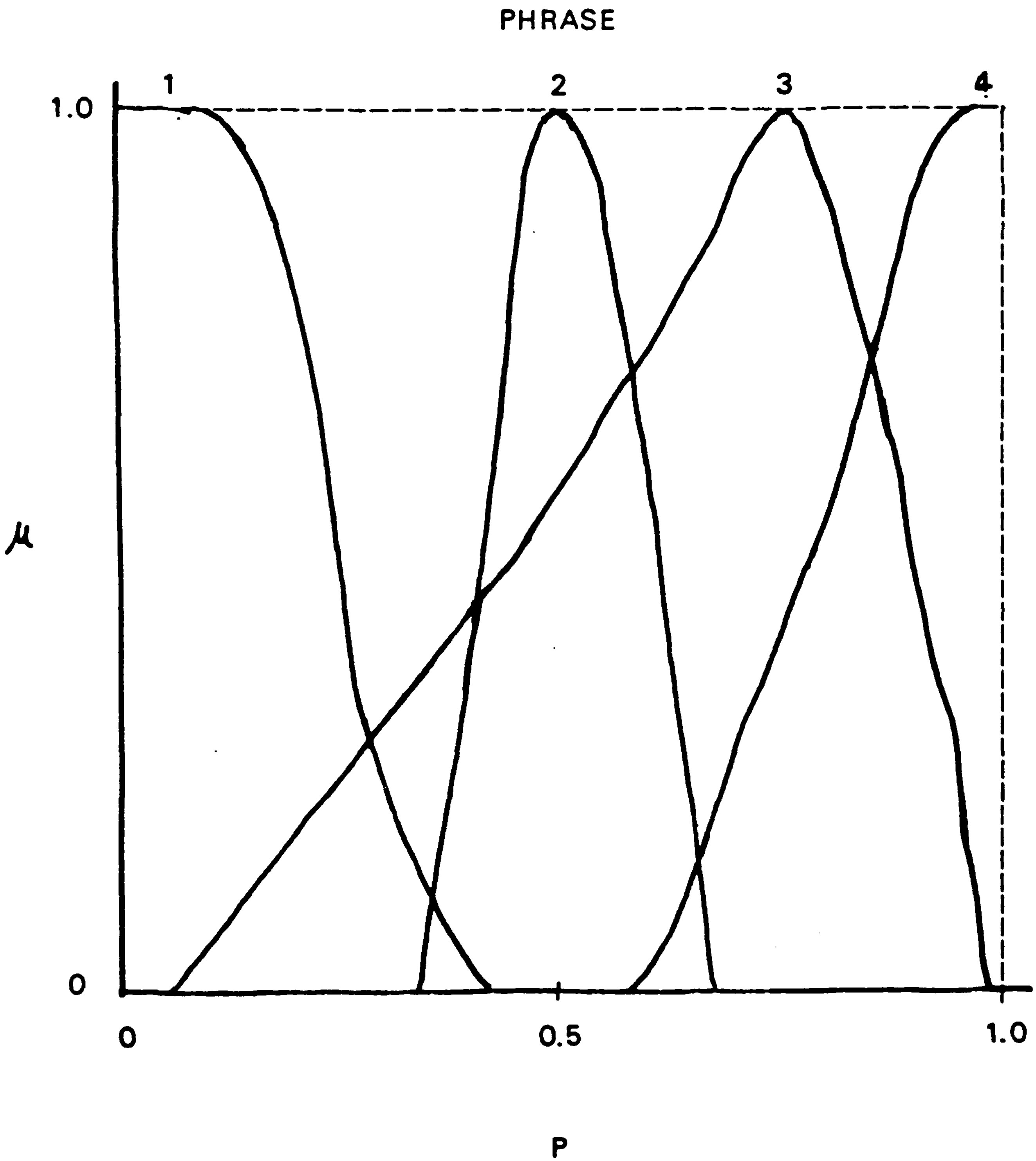


Figure 7. Generic forms of empirically obtained membership functions.

Table 4

Distribution of membership function shapes (in %)

Phrase	M.D.	S.P.	M.I.	Oth	#
Almost certain	0	10	90	0	10
Very (p., l.)	0	5	85	10	20
Probable Likely Good chance	4	48	46	2	90
Rather (p., l.)	0	30	50	20	20
Tossup	0	75	0	25	20
Possible	13	56	14	17	80
Rather (u., i.)	35	60	0	5	20
Unlikely Improbable Doubtful	58	39	0	3	90
Very (u., i.)	95	5	0	0	20
Almost impossible	60	20	0	20	10

phrases in the table include _very probable_, and _very likely_--shown as _very (p,l)_; _probable_, _likely_, and _good chance_; _unlikely_, _improbable_ and _doubtful_; and _very unlikely_ and _very improbable_--shown as _very (u,i)_. The last column in the table shows the number of subjects for whom membership functions were estimated for each grouping of words. The four central columns in the table show the percentage of membership functions classified as monotonic decreasing (M.D.), single peaked (S.P.), monotonic increasing (M.I.), and Other. Note that phrases denoting high probabilities are represented primarily by monotonic increasing functions with a few single peaked ones. The frequency of single peaked functions increases toward the central phrases, while some monotonic decreasing functions are also noted. Monotonic decreasing functions then predominate at the low probability phrases. Thus, although the distribution of meanings, as represented by the membership functions, is systematic and interpretable, it is far from constant over individuals.

Meanings are constant within individuals, however, as shown both in the reliability data already discussed and in the similarities between the functions obtained by the methods of pair-comparison and direct estimation. These latter results are illustrated for the phrases _likely_ and _unlikely_ in Figure 8. The solid lines are derived from the direct estimation procedure, the dashed lines from the pair-comparison procedure, and each panel represents a different subject. Note that functions within a panel are generally the same shape. In fact, in 90 out of 100 comparisons (five phrases for each of 20 subjects) the two methods yielded the same shaped function. Note also in Figure 8

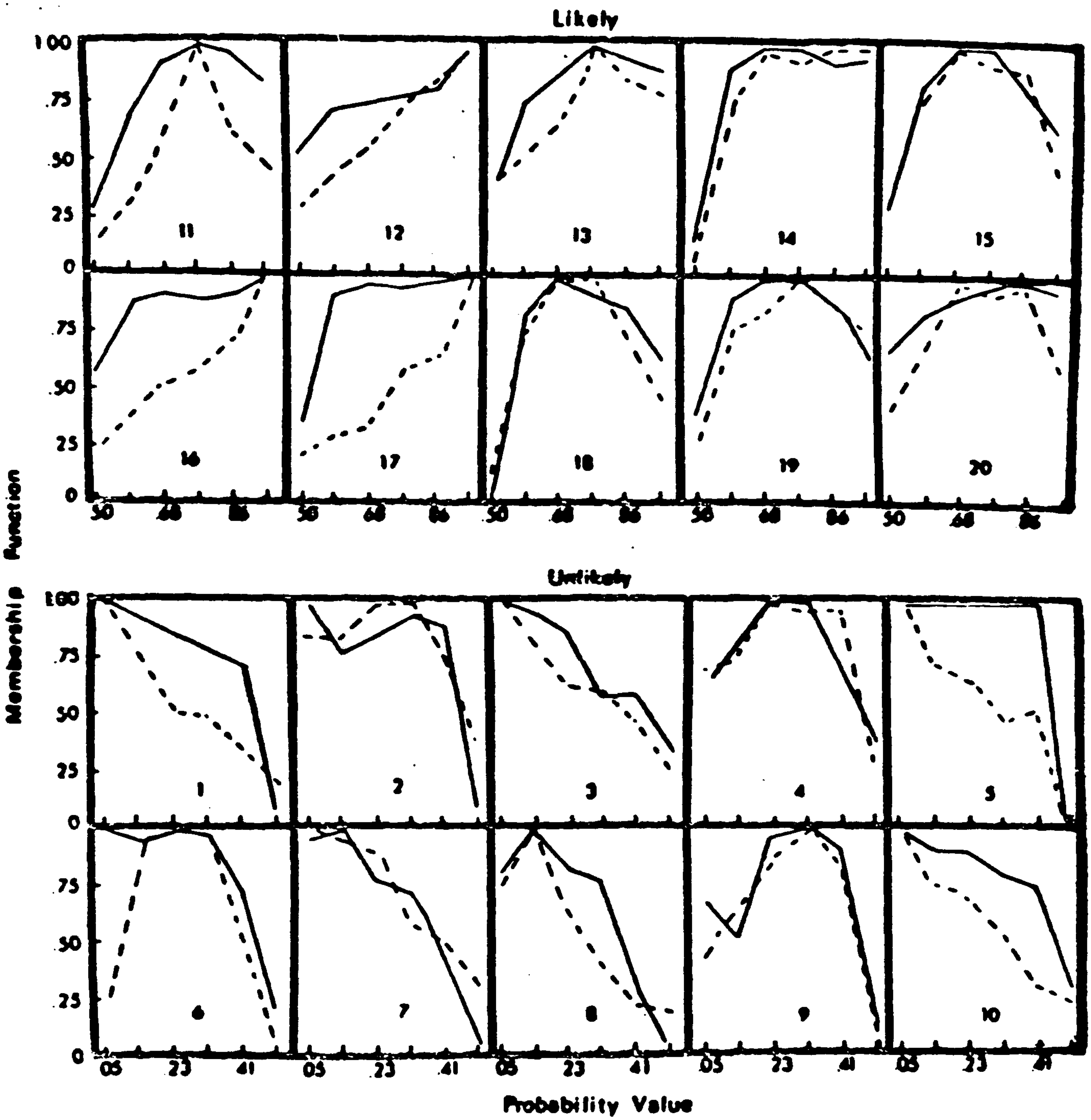


Figure 8. Pair-comparison and direct estimation membership functions for two phrases. (From Rapoport, Wallsten, & Cox, in press)

that the direct estimation functions generally lie above the pair-comparison functions. Overall, this relation occurred 52 times, the reverse occurred 11 times and neither occurred 37 times. Thus, generally, the direct estimation functions resulted in higher membership values than did the pair-comparison functions. In other words, the direct estimation procedure implied that more probabilities were better members of the concept represented by the phrase. Finally, both the direct estimation and the pair-comparison functions correlated well with the outcomes of the rank ordering procedure, although the pair-comparison was slightly superior in this regard.

Figure 9 summarizes the main conclusions from the three experiments, some of which have been highlighted above: (1) Subjects can make the judgments required to obtain interpretable membership functions representing the vague meanings of probability phrases. (2) There are large individual differences in the phrase meanings. (3) The two scaling procedures yield sufficiently similar results that we can use the simpler direct estimation technique in subsequent research studying factors that affect meanings as well as how judgments and choices are made from linguistic uncertainties.

5. FACTORS AFFECTING MEMBERSHIP FUNCTIONS.

The research described in the previous section demonstrated considerable variation over individuals in the vague meanings of phrases within a specific context. It seems highly likely as well that within an individual, phrase meanings vary over contexts. For example, if one says "it is likely to rain

Conclusions

- Subjects can compare degrees of membership such that consistent, reasonable, and interpretable scaling of vague meaning is possible
- Despite the use of precise probabilities, there were large individual differences in the vague meanings of the phrases
- Approximately half the membership functions are monotonic, with the rest generally single-peaked.
- The more extreme phrases are more frequently represented by monotonic functions, and the more central phrases by single-peaked functions
- Judgments appear to be based on differences rather than ratios
- DE yields functions similar to those of PC, but generally somewhat more vague
- The methods may be useful in studying effects of context and other factors on meaning

Figure 9. The main conclusions from Rapoport, et al. (in press) and Wallsten, et al. (1986).

tomorrow afternoon," or "California is likely to have a major earthquake in the next few years," the meaning of likely may be very different in the two situations. There are two possibilities as to how these different meanings might relate to membership functions. At one extreme, the meanings of phrases to an individual might be represented by entirely different functions in each context. If this were the case, one would have the theoretical task of relating changes in membership functions to changes in contexts. At the other extreme, individual membership functions would remain fixed over contexts, even while the uses of the phrases changed. Here the theoretical task would be to explain how the function is evaluated for the purpose of understanding or using the phrase in a particular situation.

A recently completed study (Fillenbaum, Wallsten, Cohen, & Cox, in preparation) looked at two factors that may affect individual membership functions of specific phrases. One factor was the nature of the communication task, namely, whether one receives the phrase in communication from another person or selects the phrase in order to communicate to someone else. Assuming one has better knowledge of one's own vocabulary than of other people's, it might be expected that phrases are treated as more precise when selected than when received. This effect would translate into sharper membership functions covering a smaller interval of probabilities. The second factor that may affect membership functions is the available vocabulary. Specifically, we thought it possible that the availability of extreme phrases (such as almost certain) or modified phrases (such as very unlikely) might affect the meaning of core phrases (such as unlikely or probable).

In this experiment, each of 23 subjects served in each cell of a 3 communication task by 2 vocabulary condition design. Following a day of practice, each of the six combinations was run in a separate session. The two vocabulary conditions consisted of a core (likely, probable, possible, unlikely, and improbable) and a core plus context condition. For 11 subjects there was an anchor context consisting of almost certain, tossup, and almost impossible. The remaining 12 subjects had a modified context consisting of quite probable, very likely, very improbable, and quite unlikely. The three communication tasks were selection, comprehension and evaluation.

The selection task was intended to model the situation in which an individual must choose a phrase to communicate to someone else, and is illustrated in Figures 10 and 11. Figure 10 shows the initial computer display on a selection trial. The subject was shown a particular spinner and a list of phrases from which he or she was instructed to select the phrase that would best communicate to a friend who was going to bet on the spinner the probability of its landing on white. The trial was actually iterated so that the subject first selected the best description, then the next best, and so on, until he or she felt that no remaining phrase was sufficiently descriptive to warrant selection. In Figure 10, improbable had already been selected on the previous iteration; it remained in view but was boxed off and not available. The subject in the illustration, therefore, is about to select the phrase unlikely. After the selection had

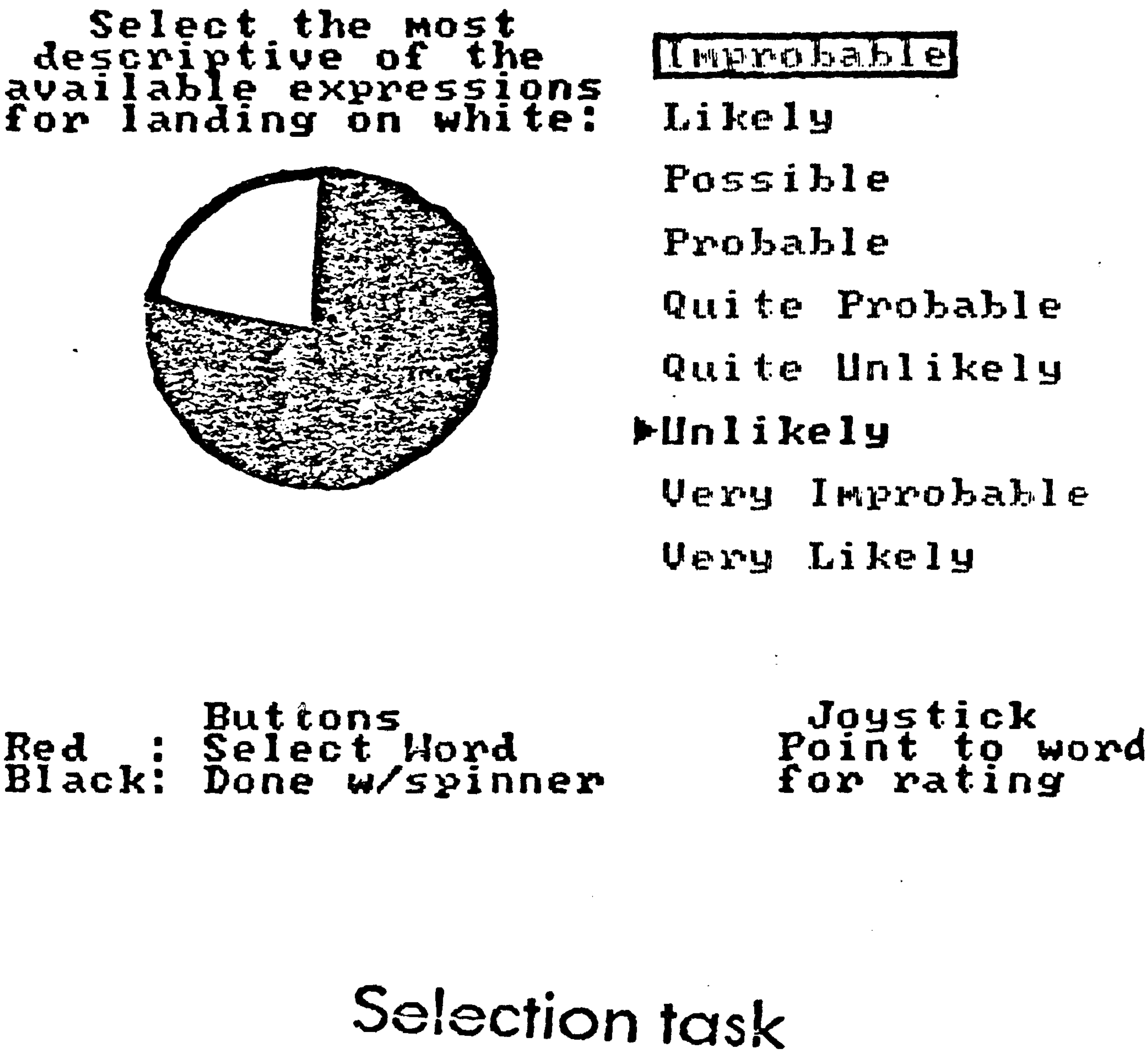


Figure 10. Example of the first computer display in the selection task.

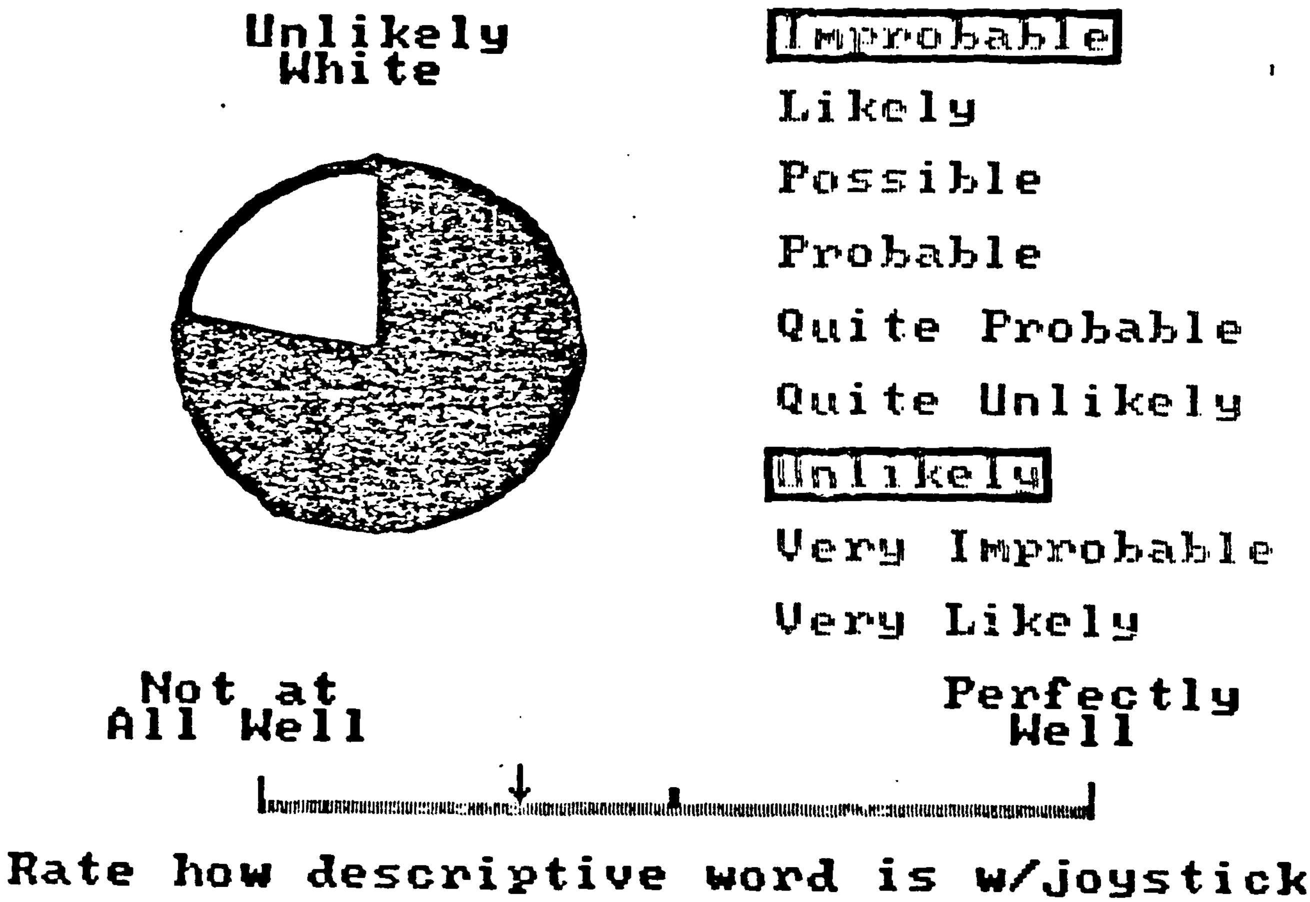


Figure 11. Example of the second computer display in the selection task.

been made, the screen changed to that shown in Figure 11, on which the subject was required to move the arrow on the response line to indicate how well the phrase describes the probability of the spinner landing on white. Following that judgment, the screen reverted back to that shown in Figure 10 but with the previous choice also boxed off. The subject could then select another phrase or indicate that he or she was done with the spinner. The rating technique (illustrated in Figure 11) applied to many probabilities for the same word, yielded the membership function for that word.

Not being certain of the best way to model the reception situation, we employed two distinct tasks which we term evaluation and comprehension. In both cases, the subjects were instructed to imagine that they were going to be required to bet on a spinner landing on white. However, the spinner would be invisible to them. A friend of theirs who could view the spinner would use a probabilistic phrase to communicate to them the chances of its landing on white. In the evaluation task, then, the subject was shown a particular spinner and a list of phrases available to the friend, as illustrated in Figure 12. At the beginning of a trial the computer randomly selected a phrase. The subject then rated how descriptive the phrase would have been if it had been used to describe the particular spinner on the screen. That phrase was then boxed off, another was selected, and so on through the list. Thus, operationally, the selection and evaluation tasks differed only in that in the former the subject selected the phrase, choosing whatever subset of phrases he or she thought appropriate for the spinner in question,

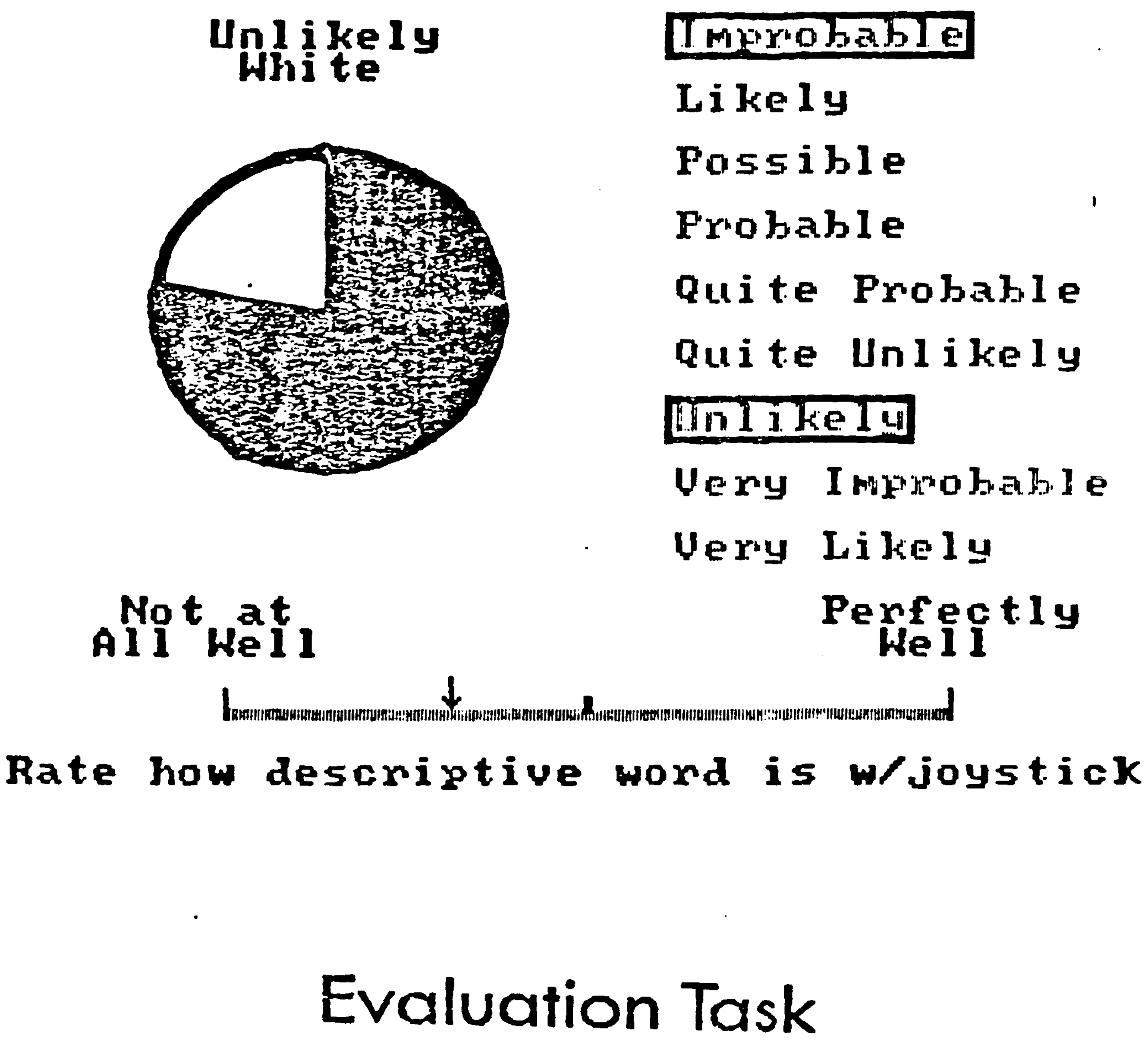


Figure 12. Sample computer display for the evaluation task.

whereas in the latter case the computer selected all the phrases in random order. Both methods yielded membership functions that can be compared.

The comprehension task, illustrated in Figure 13, provides an alternative modeling of the reception situation. The same cover story was used regarding a friend's description of a spinner that cannot be seen by the subject. However, here the list of probability phrases available to the friend was shown on the screen along with a spinner evenly split between the white and shaded regions. The computer selected the phrases in random order. For each, the subject was required to adjust the spinner to show the highest probability the friend may have been viewing. When this was done, the screen changed to request that the subject set the spinner to show the lowest probabiity at which the friend may have been looking. Finally the subject was requested to select a value between these two that represented the probability the friend most likely was describing. This task did not yield membership functions, but rather a lowest, best, and highest probability for each phrase.

The dependent variables for the selection and evaluation tasks included the membership function characteristics of shape, location along the probability interval and width. Location was indexed by Yager's (1981) W (similar to a weighted mean), defined as

$$W = [\Sigma p_i \mu(p_i)]/\Sigma\mu(p_i).$$

Width was indexed by a measure V (similar to a standard deviation) defined as

$$V = [\Sigma(p_i - W)^2 \mu(p_i)]/\Sigma\mu(p_i).$$

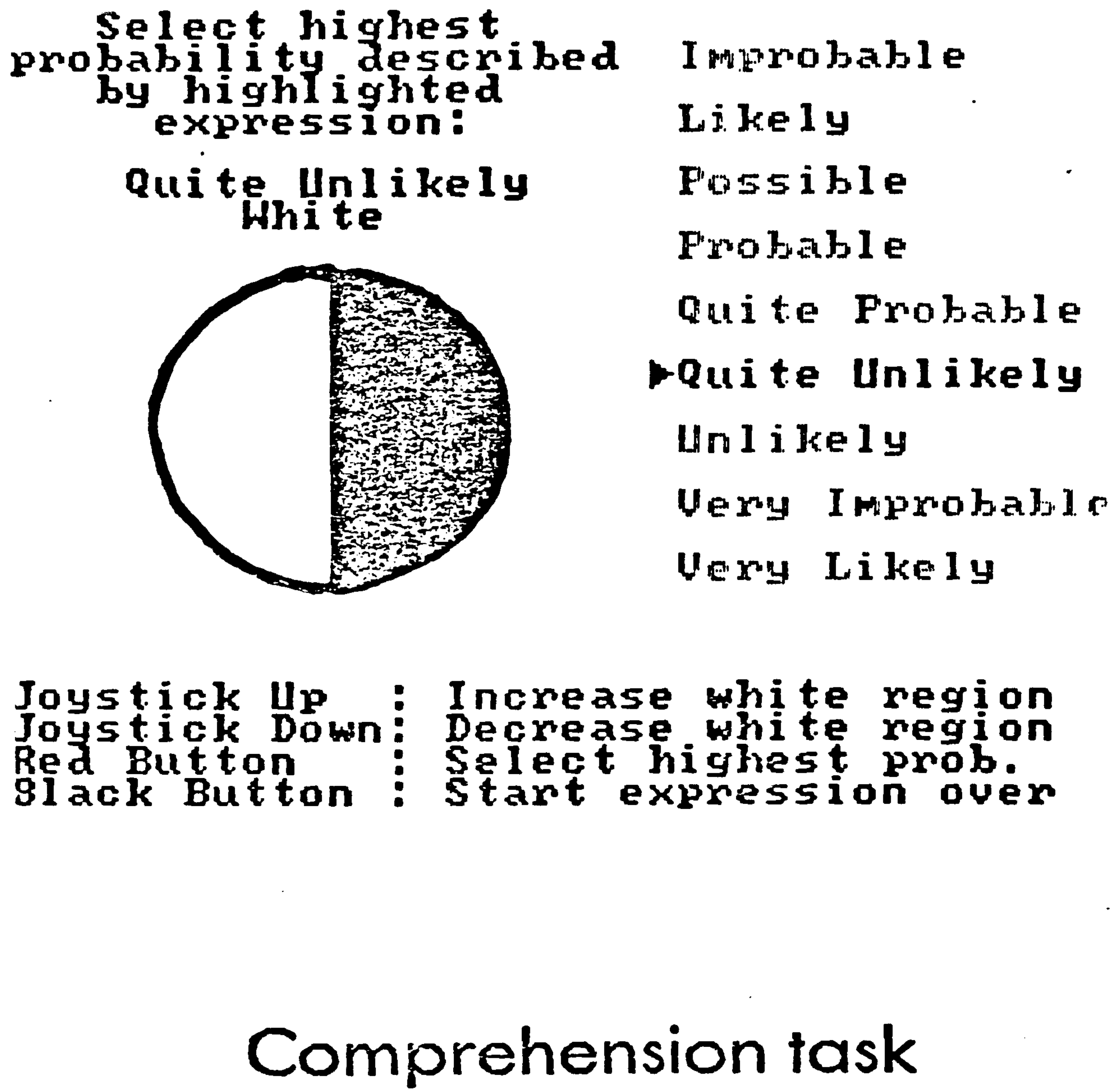


Figure 13. Sample computer display for the comprehension task.

Dependent variables for the comprehension task included the lowest, best, and highest probabilities for each phrase.

The results will just be summarized here; for details see Fillenbaum et al. (in preparation). Within communication task available vocabulary had little effect on the meanings of the core phrases. The membership function shapes were unaffected; there were no differences in the ranges of meanings, either as measured by the index V or by the difference between the highest and lowest probabilities in the comprehension task; nor with one exception, were the locations of the core phrases altered, as measured either by the index W or by the best probability in the comprehension task. The single exception is a small effect of the modified context on the locations of <u>improbable</u>, <u>unlikely</u>, <u>probable</u>, and <u>likely</u>. Generally, those words have slightly more extreme locations (i.e., away from 0.50) when they are presented alone than when they are presented in the context of the modified more polarized phrases.

The effect of communication task is more profound, however, and is shown in the next three tables. Table 5 shows the effect of communication task on the shapes of membership functions for the core phrases. Note that monotonic (increasing or decreasing) functions predominate in the evaluation case but are a minority in the selection case. Conversely, there are 55% more single peaked functions in the selection than in the evaluation tasks.

Table 6 shows the effects of communication task on the location of the core phrases. <u>Improbable</u> and <u>unlikely</u> are combined, as are <u>probable</u> and <u>likely</u> because there were no substantial differences in the patterns of responses to the

Table 5

Communication Task Effects on Shape of Membership Functions of Core Phrases (in %)

	Selection	Evaluation
SP	48	31
M	23	58
O	29	11

Table 6

Communication Task Effects on Location of Core Phrases

	Improbable Unlikely	Possible	Probable Likely
		Selection	
W	.22	.45	.76
		Evaluation	
W	.29	.48	.70
		Comprehension	
Best	.19	.42	.77

members of each pair. Considering the selection and evaluation tasks first, it is evident that the low and high phrases have more extreme meanings in the selection than in the evaluation contexts. Interestingly, the best probabilities in the comprehension task, which was intended to be similar to the evaluation task, are closer to the locations in the selection situation. We will return to this result after considering the range or spread of meanings, which are shown in Table 7. It is notable that in Table 7 the phrases all have broader or vaguer meanings in the evaluation than in the selection situations. Indeed, in the selection case, there were many probabilities for which some phrases were not selected, implying zero membership value of those probabilities in the particular phrases. However, in the evaluation task, where the computer selected phrases for probabilities, subjects never gave a rating of absolutely not descriptive, although many ratings were very close to that. The range in the comprehension task is not directly comparable to the V index in the other two tasks, but one can note that as in the other tasks, the phrase possible is broader or more vague than are the other core phrases. The comprehension range, however, is considerably less than the range of probabilities with nonzero membership values in the other two tasks.

To understand the location results (Table 6) previously described, it is necessary to consider the nature of the membership functions in the selection and the evaluation tasks. First, the peaks of the functions (i.e., the probabilities with maximum membership) tend to be in the same location for both tasks, although for a variety of technical reasons (described in

Table 7

Communication Task Effects on Spread of Core Phrases

	Improbable Unlikely	Possible	Probable Likely
		Selection	
V	.06	.08	.06
		Evaluation	
V	.10	.12	.10
		Comprehension	
Range	.26	.44	.28

Fillenbaum, et al., in preparation) exact comparison is difficult. The increased spread in the evaluation task (Table 7) combined with the fact that the peaks are off-center (not at 0.50) means that there is greater skew to the implied meanings in the evaluation than in the selection situation. Thus, the shift in location, illustrated in Table 6, is a result of the phrases' covering probabilities further away from the peaks, rather than from movement in the peaks, themselves. Now, the best probability estimates in the comprehension task can be understood as falling between the peaks and the index W for each word in the evaluation task. Recall also that the comprehension range in Table 7 is considerably greater than the range of probabilities with nonzero membership values in the evaluation task. The implication of all this is that when the subject is required to express the meaning of a phrase in terms of three numbers (lowest, best, and highest), he or she considers a subset of the probabilities with sufficiently high membership, and then combines them or averages them in some fashion that yields a best estimate shifted away from the peak towards the more extreme tail, but not so far as W.

On the assumption that the experimental tasks properly model real-world reception and selection situations (see Fillenbaum et al., in preparation, for discussion of this point), the general conclusions to be taken from the study are that probability phrase meanings are relatively unaffected by the available vocabulary, but are affected considerably by the communication task. Specifically, phrases are more precise, more extreme, and more frequently single-peaked when they are selected than received. These results should give pause to designers of expert systems that rely on fuzzy set theory, who must consider whether the systems should include the meanings of phrases as understood by the decision maker or as intended by the forecaster. The latter is probably preferable, but this is a question that requires further research.

With respect to the underlying cognitive psychology, the original hypothesis that individuals understand their own use of language better than other people's use of language was supported. At the very least, theories of inference and judgment based on linguistic probabilities will have to allow for separate functions in the two communication tasks. Such a theory will be described in the last section of the report.

6. OTHER CONTEXT EFFECTS

We have completed three other context studies: one on the effect of event desirability on comparisons of objective probabilities (Cohen, 1986) and two on the effect of base rate on the interpretation of probability and frequency expressions (Wallsten, Fillenbaum, & Cox, 1986). Strong effects were shown in all cases, although membership functions were not derived.

Effect of event desirability. In this study by Cohen (1986), subjects were asked to judge the relative likelihood of two events that were differentially desirable. Figure 14 illustrates a trial in which the subject was confronted with two gambles, each of which depended on a different spinner and both of which were to be played. In one case the spinner was visible so that the subject could judge the chances of its landing on

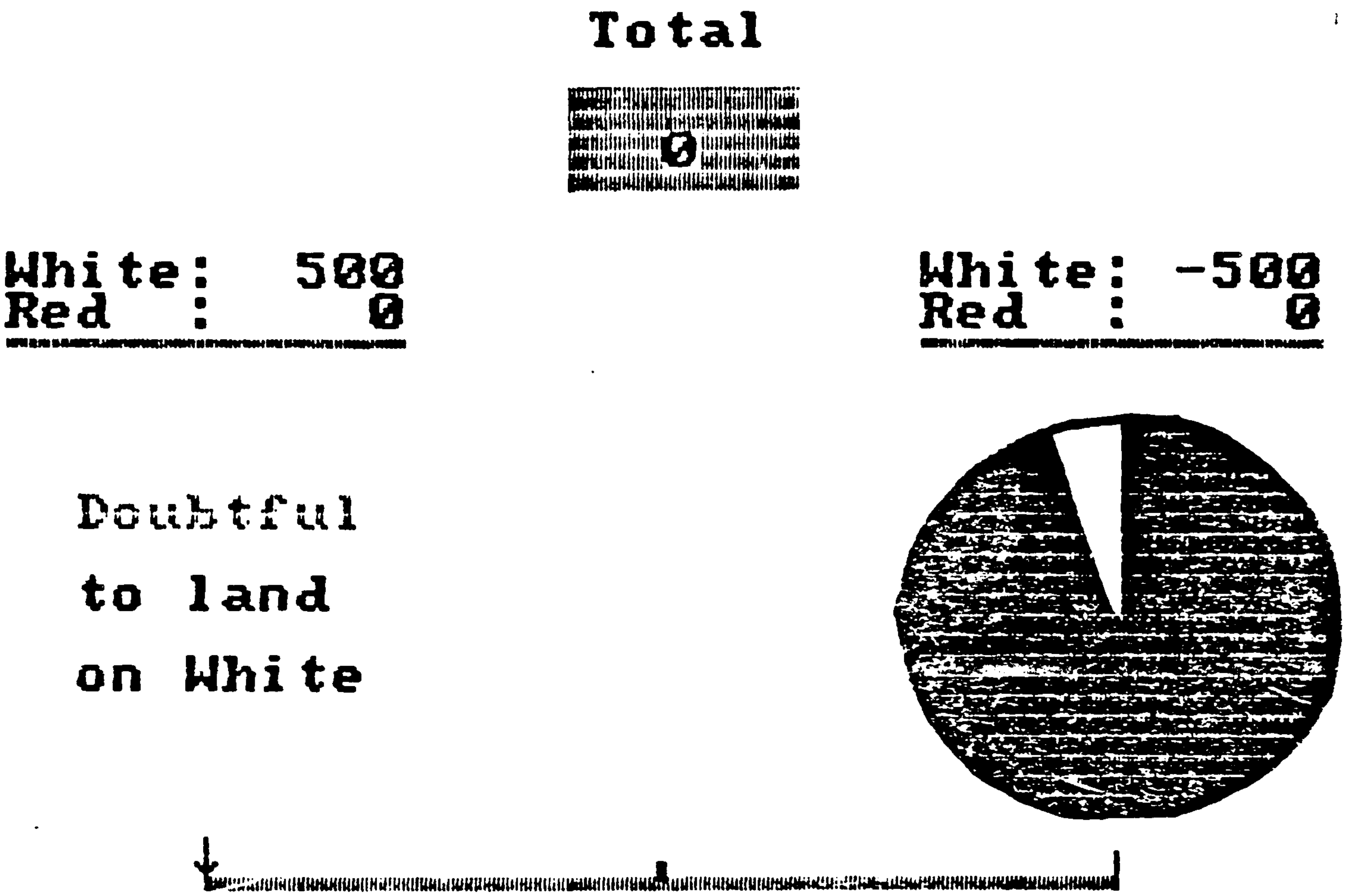


Figure 14. Sample computer display for Cohen's (1986) experiment.

white or red. The other spinner was invisible, but was represented by a probability phrase as shown in Figure 14. The subjects were truthfully told that the phrases had been selected to represent specific probabilities based on the considerable prior scaling we had done. Thus, each phrase in fact corresponded to a specific spinner that was subsequently the basis of that gamble. Gamble outcomes were shown at the top of each side of the screen. The six phrases shown in Figure 15 were utilized, and each was paired with four suitably chosen spinner probabilities. The term "unspecified" was used to convey absolute lack of information and therefore allow investigation of the Ellsberg Paradox in this situation. Thus, there were 6 x 4 = 24 distinct phrase-spinner pairings. Each pairing was combined with three outcome structures designed to manipulate the relative desirability of the events represented by the phrase and by the spinner. For example, in Figure 14 the lefthand (invisible) spinner has positive desirability while the righthand (visible) spinner has negative desirability. The reverse desirability occured when the invisible spinner had outcomes of -500 for white and 0 for red while the visible spinner had +500 for white and 0 for red. On neutral desirability trials, all outcomes were zero. The subject's task on each trial was to move the cursor on the response line to indicate which of the two spinners he or she thought was more likely to land on white. The response actually had no impact on the gambles, so that the outcomes were independent of any judgment made. Following the response, both gambles were played and the point total was incremented or decremented as appropriate. Outcome regarding the specific

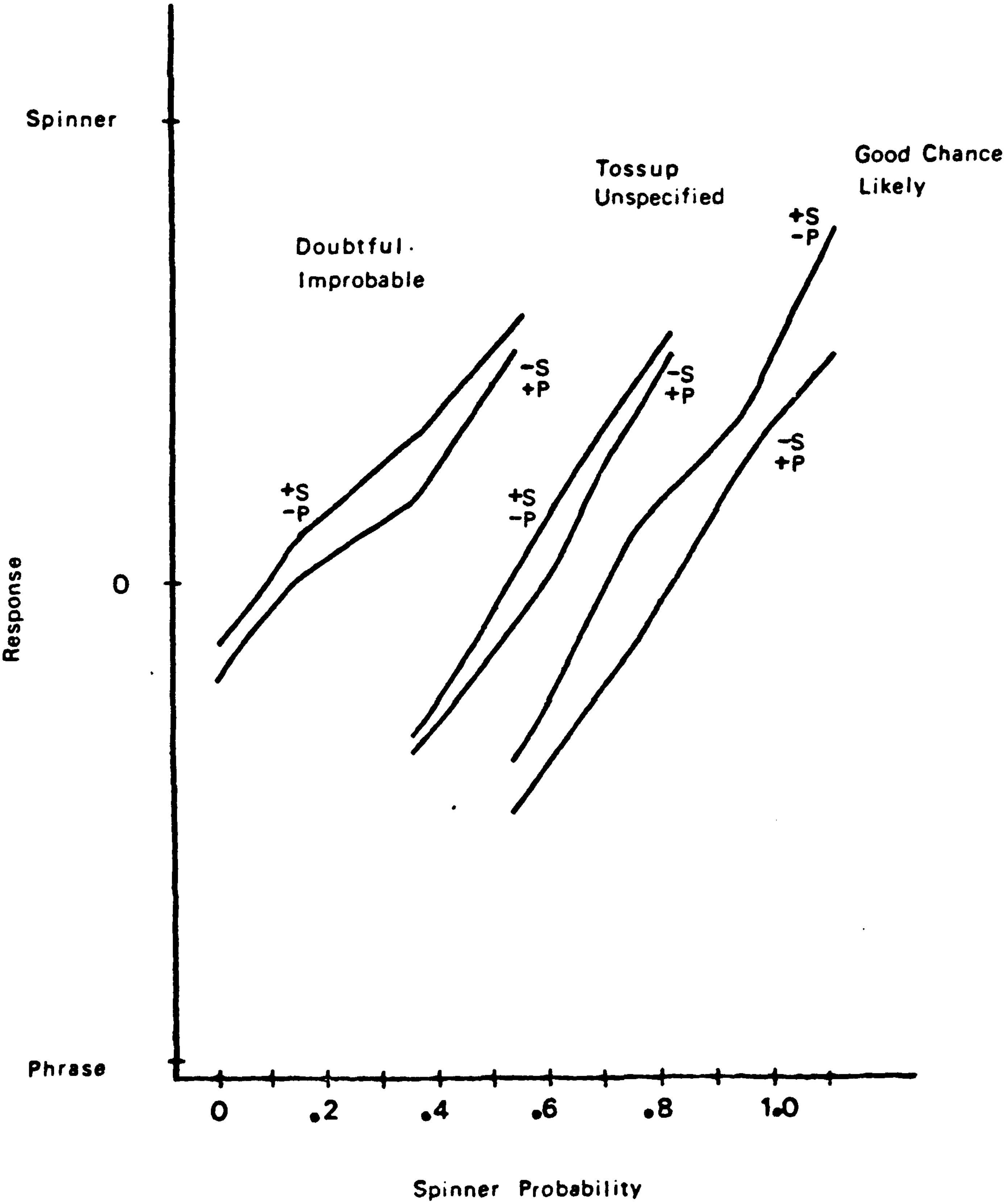


Figure 15. Mean judged likelihood of spinners relative to phrases as a function of desirability.

gambles was not provided.

The results of Cohen's experiment are shown in Figure 15, which plots cursor location as a function of spinner probability separately for the low, neutral, and high phrases. For each of the three sets of terms, there is a separate function for the positive spinner-negative phrase desirability conditions and for the negative spinner positive phrase desirability conditions. The neutral conditions consistently fall between the two and were omitted from the graph for clarity. It is evident from the graphs, and also supported by statistical analyses, that judgment was biased toward the positively desirable and away from the negatively desirable events. Thus, either the interpretation of the probability phrases or the perception of the spinner probabilities, or both, were affected by the levels of desirability. Although one cannot conclude with certainty that the effect is on interpretation of the phrases, that is the most likely possibility because spinner relative areas are so easily and accurately perceived (Wallsten, 1971). In any case, a subsequent study is now underway to check that interpretation as well as to test a theoretical explanation of the results. The theory itself will be described in the last section of this report.

Base rate effects. Two experiments on this issue have been completed and reported by Wallsten, Fillenbaum and Cox (1986). The question addressed by both was whether the meanings of probability and frequency expressions are affected by the perceived base rates of the events to which the expressions refer. Considering the extensive evidence demonstrating that under a variety of conditions people are relatively insensitive to base rates when processing diagnostic information (Bar Hillel, 1983; Kahneman & Tversky, 1973; Tversky & Kahneman, 1982, Wallsten, 1983) one might expect base rates to have no effect on the interpretations of phrases. Other studies (Cohen, Dearnley, & Hansel, 1958; Borges & Sawyers, 1974) have shown that the interpretations of quantifiers of amount such as some, several, or many are affected by the available quantity of the object. In addition, the study by Pepper and Prytulak (1974) and the more general review by Pepper (1981) sugggest that the meanings of quantifiers of frequency, such as frequently or sometimes are influenced by the expected frequencies.

In the first study by Wallsten, Fillenbaum and Cox (1986) meteorologists were asked to interpret medical forecasts. A sample questionnaire is shown in Table 8. Note that the first and third questions concern high probability events, while the second and fourth concern low probability vents. Note also that the four forecasts utilize the phrases likely, possible, slight chance, and chance, respectively. The four scenarios were actually combined with the four probability phrases in two different 2 x 2 designs as shown in the bottom of Table 9. Thus, half the meteorologists received the four forecasts determined by the phrase-context combination corresponding to one of the diagonals in each of the matrices, and half the meteorologists received the combinations indicated by the other diagonals. The four phrases were selected because they in fact are regularly used in National Weather Service (NSW) precipitation forecasts.

Table 8

SAMPLE QUESTIONNAIRE FOR EXPERIMENT I

You normally drink about 10–12 cups of strong coffee a day. The doctor tells you that if you eliminate caffeine it is likely your gastric disturbances will stop.

What is the probability that your gastric disturbances will stop? ______________

You have a wart removed from you hand. The doctor tells you it is possible it will grow back again within 3 months.

What is the probability it will grow back again within 3 months? ______________

You severely twist your ankle in a game of soccer. The doctor tells you there is a slight chance it is badly sprained rather than broken, but that the treatment and prognosis is the same in either case.

What is the probability it is sprained? ______________

You are considering a flu shot to protect against Type A influenza. The doctor tells you there is a chance of severe, life-threatening side effects.

What is the probability of severe, life-threatening side effects? ______________

Table 9

Study Using Meteorologists

NWS Probability to Phrase Conversion

Probability of Precipitation	Phrase
.20	slight chance
.30 or .40	chance
.60 or .70	likely

Mean Probability Judgments in High and Low Base Rate Medical Contexts

	Context	
Phrase	High BR	Low BR
	Coffee	Wart
Likely	.75	.67
Possible	.48	.38
	Ankle	Flu
Chance	.39	.18
Slight chance	.23	.10

In fact, as shown at the top of Table 9, three of the phrases have been assigned to specific probability values by the NSW. Thus if a meteorologist determines that there is a 20% chance of rain, he or she may optionally say there is a slight chance of rain. Similarly, chance can be assigned to 30% or 40% and likely can be assigned to 60% or 70% chance of rain. The phrase possible is never used to express a precipitation probability, but may be used in an ancillary fashion (e.g., "A chance of rain today, possibly heavy at times").

Questionnaires were sent to 60 meteorologists, including forecasters, television forecasters, and research meteorologists, of which 46 (77%) were returned. The main results are displayed in the bottom of Table 9 which shows the mean probability judgments in the high and low base rate contexts for each of the four phrases. It is evident in the table, and confirmed by appropriate statistical analyses, that on the average a given expression was interpreted as reflecting a higher probability when it was used to predict the high base rate than the low base rate event. The variability of the estimates in each of the eight cells of Table 9 is also remarkable, and shown in detail in Wallsten, Fillenbaum and Cox (1986). Although the response distributions cover the NWS assigned values for slight chance, chance, and likely in all cases, in only three of the six instances are the assigned values at the modes of the distributions.

Two results are clear. First, the meteorologists were just as variable in converting probability terms to numbers as have been subjects employed in other studies (as discussed in the introduction to this report), despite the numerical conversion mandated by the NWS for precipitation forecasts. Second, and of more direct interest to the present issue, the meteorologists' interpretations of probability expressions in this medical context varied as as positive function of event base rate. Despite the fact that nothing in the instructions nor in the questionnaire mentioned base rate or suggested that the predicted events actually occur with differing relative frequencies, this variable had a profound effect. Clearly, the influence of base rate is robust.

The second experiment utilized undergraduate subjects to investigate under more controlled circumstances the relation between perceived base rates and the interpretations of probability and frequency expressions. A pilot study was first run to develop sets of scenarios with identical semantic content that differed only in perceived base rate or probability. In the main study the calibrated scenarios were utilized in hypothetical predictions made by experts, in which the expert's level of certainty in each prediction was communicated by means of either a probability or a frequency expression.

Highlights of the complex design are listed in Figure 16, which also gives one of the 36 scenarios employed. Thus, filling every seat in Charmichael Auditorium for a Tar Heel basketball game is a very probable event, while filling every seat in the auditorium for a circus is much less certain. Each of 72 subjects judged 18 pairs of predictions obtained by combining each of 18 scenarios with a different probability or frequency

Holding Semantic Content Fixed while Varying Base Rate

- 36 Scenarios with 2 values each, based on pilot work
  - e.g., Fill every seat in Carmichael Auditorium for a (Tar Heel Basketball game, circus)
- Combined with 18 probability or frequency expressions
  - e.g., "It is likely that every seat in Carmichael Auditorium will be filled for a (Tar Heel Basketball game, circus)."
- Each of 72 subjects judged 18 pairs of predictions, each pair using the high and low version of a scenario with the same probability or frequency expression.

Figure 16. Summary of experimental design for Experiment 2 of Wallsten, Fillenbaum, & Cox (1986).

expression. The subjects saw, in separate halves of the session, both the high and low versions of a scenario in combination with the same probability or frequency expression. Subjects were required to indicate the probability they thought the forecaster most likely had in mind, as well as the lowest or highest probabilities the forecaster may have been intending. See Wallsten, Fillenbaum and Cox (1986) for further details of the experimental design.

The main results are summarized in Table 10 and Figure 17. Table 10 shows the mean difference between "most likely" probability judgments to high and low versions averaged over scenarios. Similar results were obtained with the lowest and highest judgments, as well. Although the table shows the mean difference only within the six categories of high, neutral and low probability and frequency expressions, analyses were performed for each expression separately. There were no substantial differences for the expressions shown within each group of the table. Note that base rate had a very large (and statistically significant) impact on the meanings of the neutral and positive terms. The average effect of base rate on the low terms was small and generally nonsignificant.

Figure 17 shows scatter plots of the mean probability judgments for each of the 18 expressions. The probability expressions are shown in the top half of the figure and the frequency expressions in the bottom half. In each case the panels, reading from the upper left to the lower right, are in the same order as are the expressions in Table 10. Each point represents the high or low version of a scenario, and plots the

Table 10

Mean Difference between Probability Judgments to High and Low Versions, Averaged over Scenarios

Probability expressions	Effect	Frequency expressions	Effect
Sure Likely Probable Good chance	.16	Common Usually Frequently Often	.14
Possible	.12	Sometimes	.16
Poor chance Unlikely Improbable Doubtful	.06	Unusual Seldom Rarely Uncommon	.03

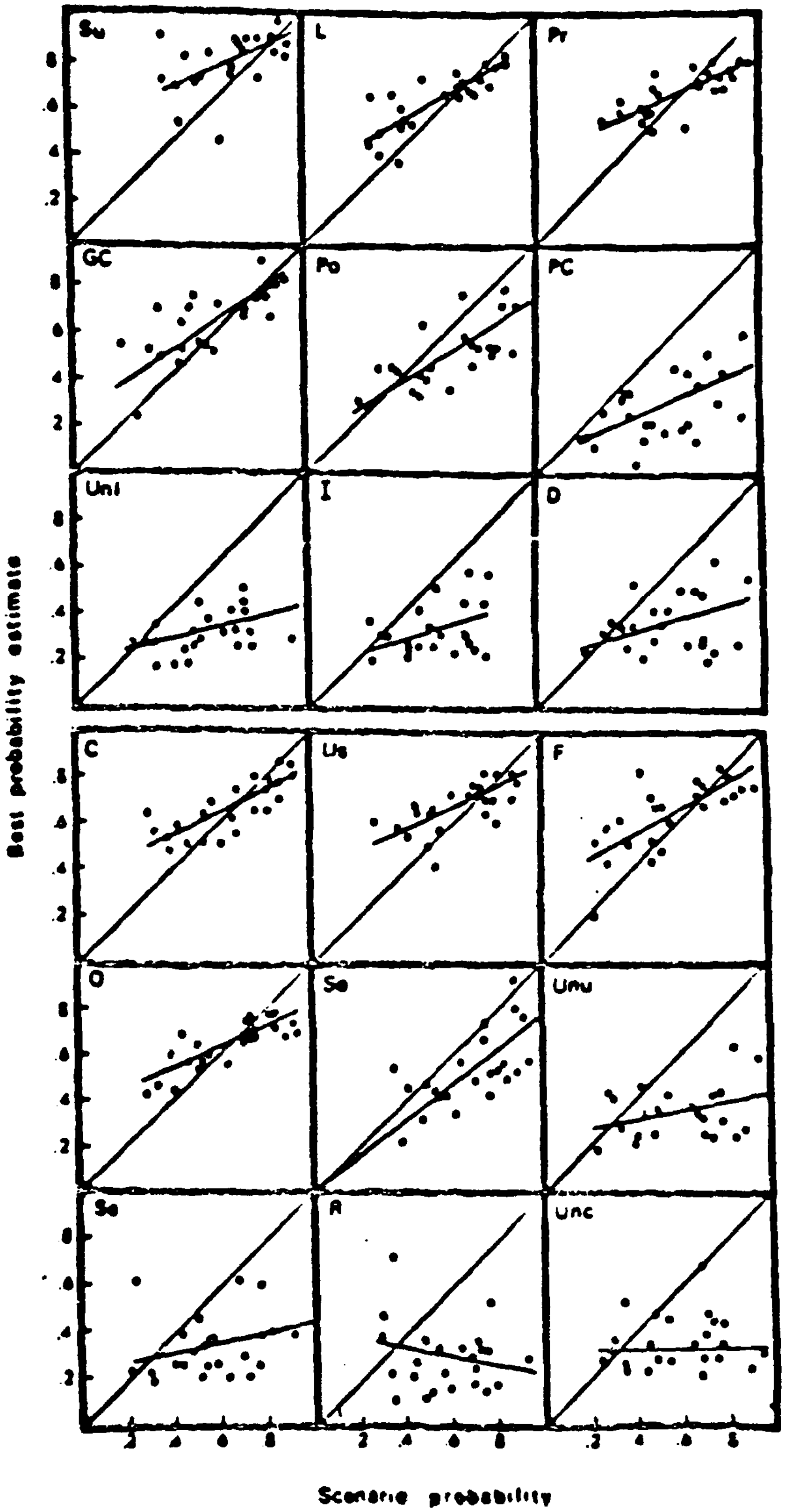


Figure 17. Scatterplots of mean probability judgments for each of the 18 expressions, as a function of scaled scenario base rates. (From Wallsten, Fillenbaum, & Cox, 1986)

mean "most likely" probability estimate as a function of the scenario probability as scaled in the pilot study. The lines represent the fits of linear structural equations (Isaac, 1970), which simultaneously minimize squared deviations in both the x and y dimensions. As would be expected from the results in Table 10, the slopes associated with the low probability or frequency words are generally close to zero, indicating that the probability judgments of forecasts were relatively uninfluenced by the prior or base rate probabilities associated with the scenarios. The remaining scatter plots show that the fitted functions generally cross the diagonal. In other words, a given neutral or positive phrase decreases high scenario probabilities and increases low scenario probabilities. It is as if the subjects' interpretations of the experts' predictions represent some kind of an average between the prior probability or base rate of the event and the meaning of the probabilistic modifier. The point at which the function crosses the diagonal represents the scenario probability that is unchanged by the verbal expression.

Four general and important conclusions emerge from these two studies. First, base rates affect the meanings of probability phrases even for people who regularly use such expressions in their professional work. Second, the meanings of high and neutral probability and frequency expressions are positively related to perceived base rate. Third, the meanings of low expressions depend less, if at all, on base rate. However, it is of interest to note that the base rate effect on <u>slight chance</u> in the meteorologist experiment was just as large as that on the other three phrases, suggesting that at least under some circumstances low expressions are also subject to manipulation by base rate. Finally, the effect of base rate can be represented as that of taking a weighted average of the phrase meaning and the prior probability.

7. SUMMARY OF RESEARCH ON MEANINGS OF PROBABILITY PHRASES

The previous findings can be summarized in four main points:

(1) The vague meanings of probability expressions to individuals in specific contexts can be represented reliably and validly by membership functions.

(2) Individual differences in understanding phrases are substantial.

(3) The particular membership function appropriate for a phrase depends on the direction of communication.

(4) The interpretation of a phrase depends on base rate and event desirability. It will be demonstrated in the final section of the report that the base rate and desirability effects (as well as other context effects that have not yet been demonstrated) can be understood in terms of how the membership function is integrated into a single value for purposes of making a judgment, rather than in terms of changes in the functions themselves.

8. DECISIONS BASED ON LINGUISTIC PROBABILITIES

The ARI contract for which this paper is a final report supported research on the meanings of nonnumerical probabilities. The research proposal did not include issues of how people actually make decisions when confronted with linguistic expressions of uncertainty, because that was considered a subsequent problem. However, while the ARI work was in progress, we (Budescu and Wallsten) received a grant from the U.S.-Israel Binational Science Foundation (BSF No. 82-03394) to conduct work on the related decision issues at the University of Haifa. Because the BSF supported research grew directly out of the work supported by ARI, and because the question of how people actually make decisions in the face of linguistic uncertainties is so important to the Army, we are including a brief summary of some of that work in this report. Two studies are of special interest here: one focusing on individuals and the other on dyads in which one person serves as a forecaster and the other as a decision maker.

Individual decisions based on linguistic probabilities. This study, reported by Budescu, Weinberg and Wallsten (1986), contrasts decisions based on numerically and verbally expressed uncertainties. Specifically, two sets of opposing predictions were tested. One set combines the fact that phrases have vague meanings with the suggestion that individuals tend to avoid decisions under ambiguity (e.g., Ellsberg, 1961) to predict that most people will tend to prefer gambles based on numerical rather than on linguistic probabilities at the sacrifice of expected gain. Furthermore, it was predicted that decision times would be greater when the uncertainties were expressed verbally than when they were expressed numerically. The other set of predictions were derived from Zimmer's (1983, 1984) work, which suggests that the verbal mode of communication is more natural to people than is the numerical. On this basis we predicted that people would generally prefer gambles based on verbal rather than numerical uncertainties, that they would perform more optimally with such gambles, and that decisions about them would be faster.

The experiment was conducted in two stages. In stage 1, each subject selected "best" numerical and verbal descriptors for each of 11 spinners. This was accomplished through an elaborate procedure in which, on separate trials, subjects assigned numerical estimates or verbal descriptors to each of the 11 spinners. As a result of replications, numerical estimates were assigned three times and verbal descriptors six times (three freely selected and three from a list) to each spinner. Each subject was then shown the (up to) three distinct estimates for each spinner and the (up to) six assigned phrases, and was asked to select which of the six phrases best describes the spinner and the numerical values. Similarly, the subject was asked to select which of the numerical values best described the spinner and the verbal expressions. In this manner ultimately 11 "equivalent" triples, consisting of a spinner, a number, and a phrase, were determined for each subject to be used in stage 2.

In stage 2 subjects provided bids for gambles involving wins or losses of $.80, $1.05, or $1.25, with uncertainty described in each of the three modes, graphic (the spinner), numeric, or verbal. Twenty subjects were run and decision times were also recorded. The summary below is based only on the bids.

The stage 1 results are summarized in Table 11 and in Figure 18. The vocabulary of the subjects was impressive. As

Table 11

Stage 1 Results

- Overall, 20 subjects used 114 phrases and 73 numbers to describe 11 displays

Number of Responses per Subject

	Free Phrases	Fixed Phrases	Numbers
Minimum	7	10	12
Mean	13.5	13.3	18.0
Maximum	19	16	29
S.D.	1.7	1.5	4.7

- Ultimately they selected 63 phrases and 51 numbers for use in Stage 2 indicated in the table, the 20 subjects used 114 phrases and 73 numbers to describe 11 displays. The body of the table shows that on average an individual subject used about 13 phrases in the free and the fixed list conditions and 18 distinct numbers to describe the 11 displays. The table also shows the range and standard deviation in numbers of responses per condition over the subjects. Ultimately, the subjects selected 63 phrases and 51 numbers for use in stage 2. Figure 18 focuses on the numerical responses and the freely selected phrases that were used by at least 10 subjects. The figure provides dot charts that shows the range of displays to which the probability numbers or phrases were applied over the multiple replications by the 20 subjects. (The dot chart for the fixed phrase condition shows the same pattern as that for the free phrase condition displayed here). Note that a given probability number was utilized over a relatively small range of probability displays. In contrast, a given verbal phrase was utilized over a very broad range of probability displays. The stage 1 results are thus consistent with our previous measurement work, as well as with other studies in the literature, in showing that a given phrase is applied to a very wide range of probabilities.

Figure 19 shows the mean stage 2 bid adjusted by the gamble probability, as a function of expected value and of display mode. Note that low probabilities are overweighted and high ones are underweighted in all three display conditions. Further, the graphic presentation yields the most nearly linear results, while the verbal presentation is the least linear of all three. Table 12 shows the mean absolute adjusted bid as a function of domain

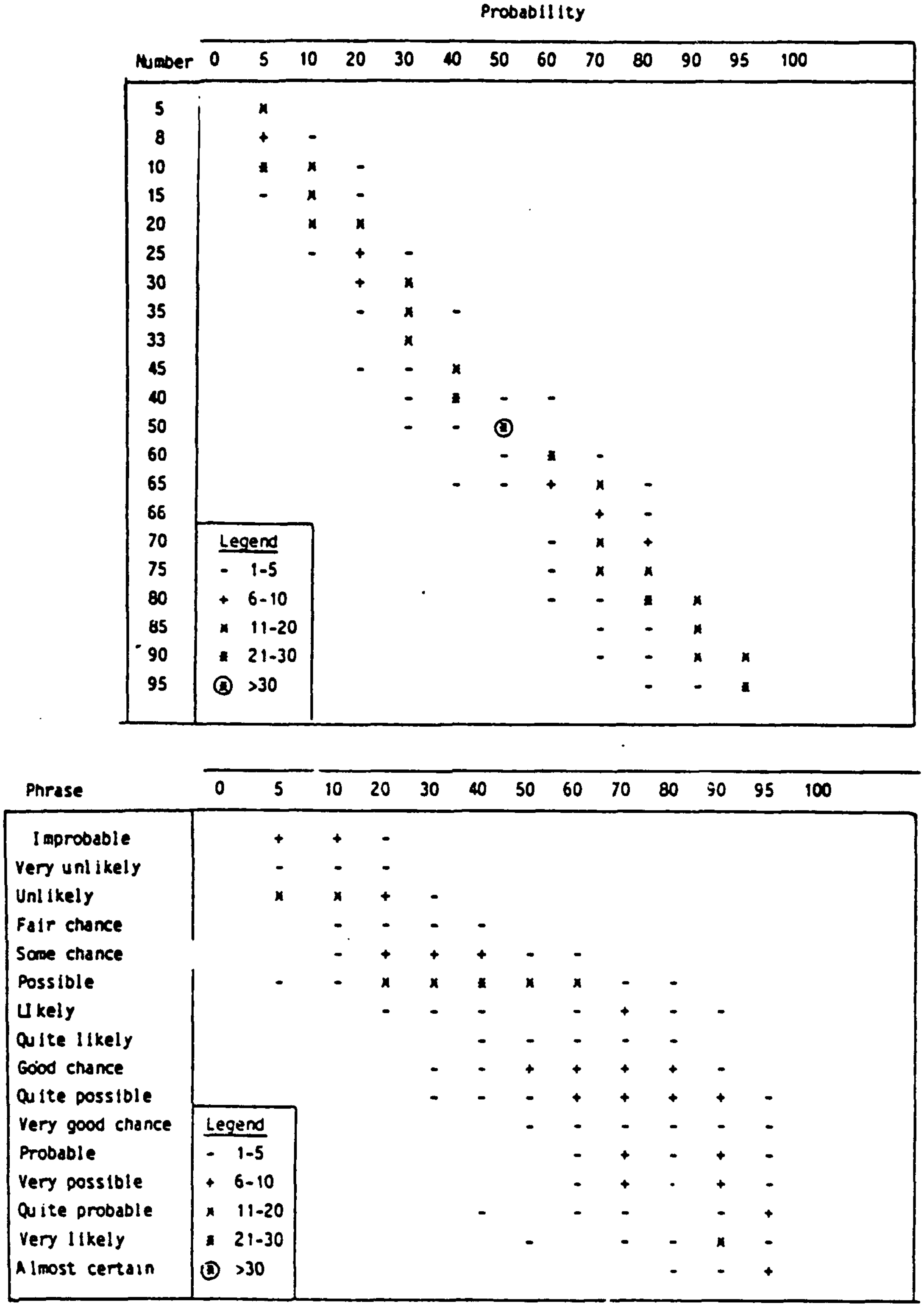


Figure 18. Dot charts for the most commonly used numbers and phrases (in the freely selected condition) in Stage 1, showing the range of displays to which each was applied. (From Budescu, Weinberg, & Wallsten, 1986)

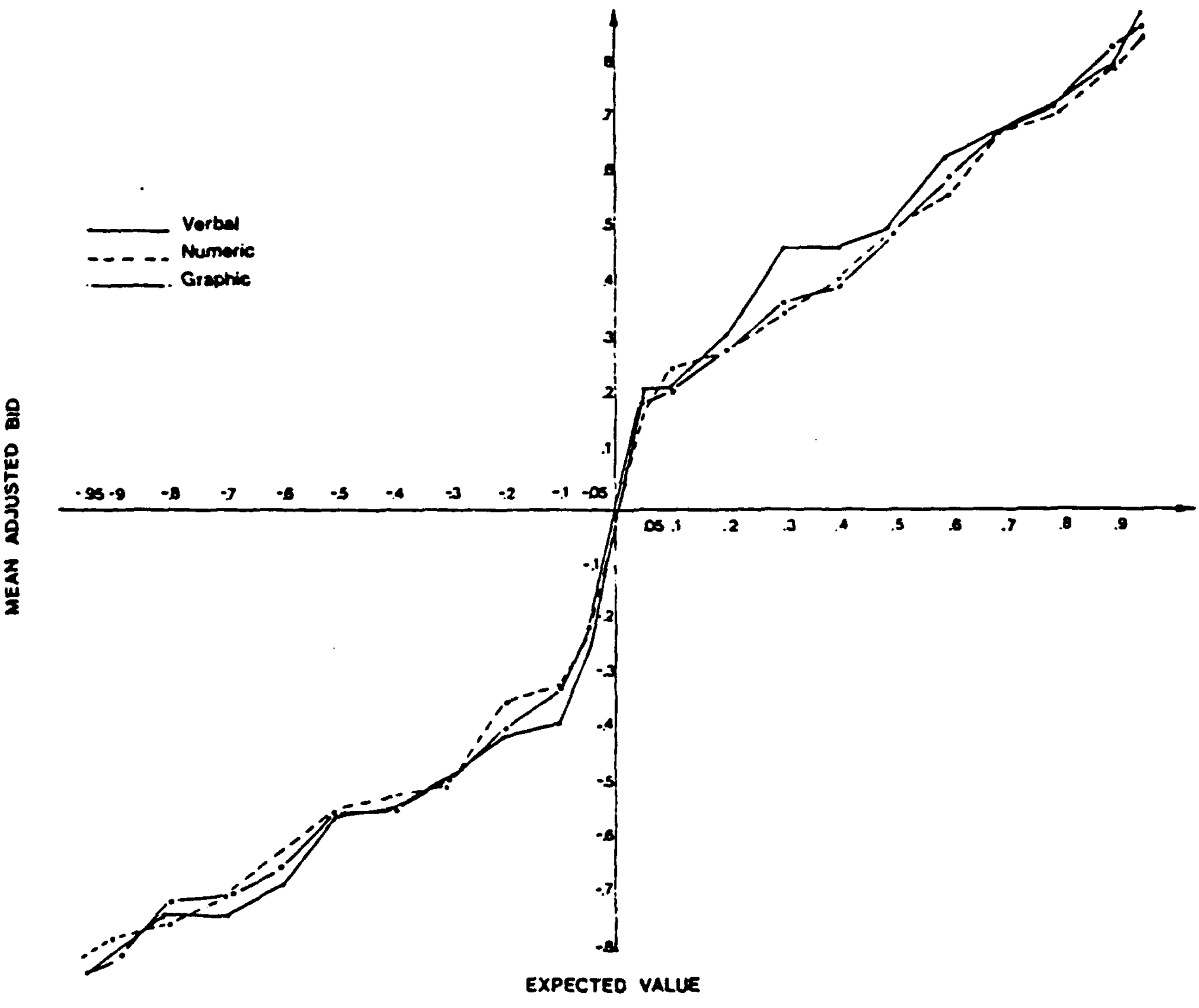


Figure 19. Mean Stage 2 bid adjusted by the gamble probability, as a function of expected value and display mode. (From Budescu, Weinberg, & Wallsten, 1986)

Table 12

Mean Absolute Adjusted Bid as a Function of Domain and Presentation Mode

Domain	Numeric & Graphic	Verbal	Mean
Gains	.51	.53	.52
Losses	.56	.58	.57
Mean	.54	.56	.54

and presentation mode. The numeric and graphic results are combined, because they were not different. If subjects were always bidding the expected value, then all table entries would be 0.50. Thus, subjects demanded more than expected value for gambles involving gains, while simultaneously, they were willing to pay more than expected value to avoid gambles involving losses. Statistical analyses support the conclusion derived from the table that these effects are stronger in the verbal than in the numerical or graphical modes. In other words, subjects' preferences for positive verbal gambles were stronger than their preferences for positive numerical gambles, and similarly, their aversion to negative verbal gambles was stronger than their aversion to negative numerical or graphic gambles. Thus, there was risk seeking in the positive domain, risk aversion in the negative domain, and these effects were stronger for the verbal than the other gambles.

Table 13 shows the mean expected gain or loss as a function of the domain and presentation mode. It can be seen that decisions in response to numerical or graphical uncertainties led to greater gains and smaller losses than did their decisions in the face of verbal uncertainties. The absolute magnitude of the differences, however, was very small, although it was significant in both cases. Although small in either domain, combined over gains and losses, the inferiority of the verbal presentations is 24% (-5.6 vs. -4.6).

Neither set of prior predictions was completely sustained. Subjects did perform more optimally with numerical or graphical than with verbal gambles, but the magnitude of the effect was

Table 13

Mean Expected Gain/Loss as a Function of Domain and Presentation Mode

Domain	Numeric & Graphic	Verbal	Mean
Gains	15.1	14 9	15.0
Losses	-19.5	-20.5	-19.9
Total	- 4.6	- 5.6	- 4.9

relatively small. Further, verbal gambles were actually preferred in the domain of gains while numerical gambles were preferred in the domain of losses. Perhaps of greatest interest, however, is that despite the greater vagueness of the probability phrases shown in stage 1, there was overall a very similar pattern of stage 2 bids for the three expression modes. This result suggests to us that when an individual must make a decision on the basis of a verbal uncertainty, he or she integrates the range of meaning into a single quantity for the purpose of making that decision. We shall return to this point in the section on theory.

Dyadic decisions based on numerically and verbally expressed uncertainties. This study, to be reported by Budescu and Wallsten (in preparation), was intended to model the common situation in which a decision maker must take action on the basis of information received from a forecaster (e.g., an intelligence agent). Each dyad consisted of a forecaster and a decision maker who were placed in separate cubicles and communicated only by means of the computer. On each trial the forecaster, who was unaware of the gamble outcomes, saw one of 11 spinners and had to communicate the uncertainty to the decision maker by means of either a numerical or a verbal probability descriptor. The decision maker saw the forecaster's judgment, but not the spinner, and on that basis bid for gambles involving gains or losses of one dollar. The forecaster, of course, did not learn what the decision maker had bid. Following verbal trials, both the forecaster and the decision maker provided best numerical judgments of the phrase used. Fifteen dyads participated in the study.

Figure 20 shows the difference between the forecasters' and the decision makers' numerical judgments for phrases, as a function of the spinner probabilities to which the phrases were applied. Note that the decision makers consistently gave numerical judgments that were closer to 0.5 than did the forecaster--a result absolutely consistent with the results obtained by Fillenbaum et al. (in preparation) presented earlier. In other words, decision makers overestimated the meanings of phrases assigned to low probabilities and underestimated the meanings of phrases applied to high probabilities.

Figure 21 shows the mean bid as a function of spinner probability, separately for the numerical and verbal judgments. Consistent with Figure 20, mean bids in both the positive and negative domain were more extreme than expected value for probabilities less than 0.5 and less extreme than expected value for probabilities greater than 0.5. Interestingly, this same pattern occurred for both the numerical and the verbal presentations. A possible explanation for this similarity is that the decision makers treated the forecasters' verbal and numerical judgments as being equally vague. That is to say, the decision maker assumed that numerical judgments were not made precisely, but rather with some variability, and consequently were no more informative than were the verbal judgments. It is also apparent in Figure 21 that the extent of over bidding was greater in the positive than the negative domain. This result is summarized in Table 14 which shows the mean absolute bid as a

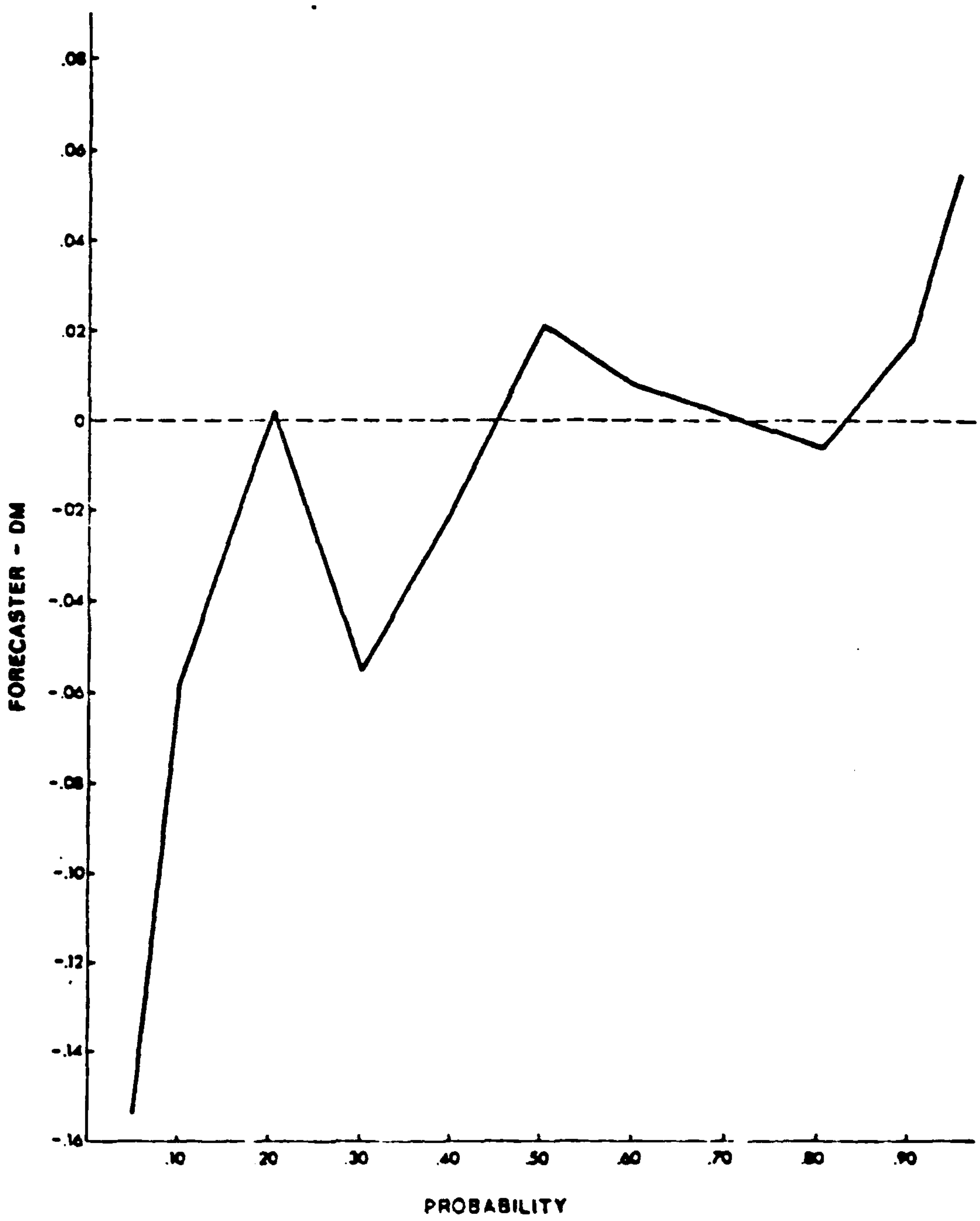


Figure 20. Mean difference between the forecasters' and decision makers' numerical judgments for phrases.

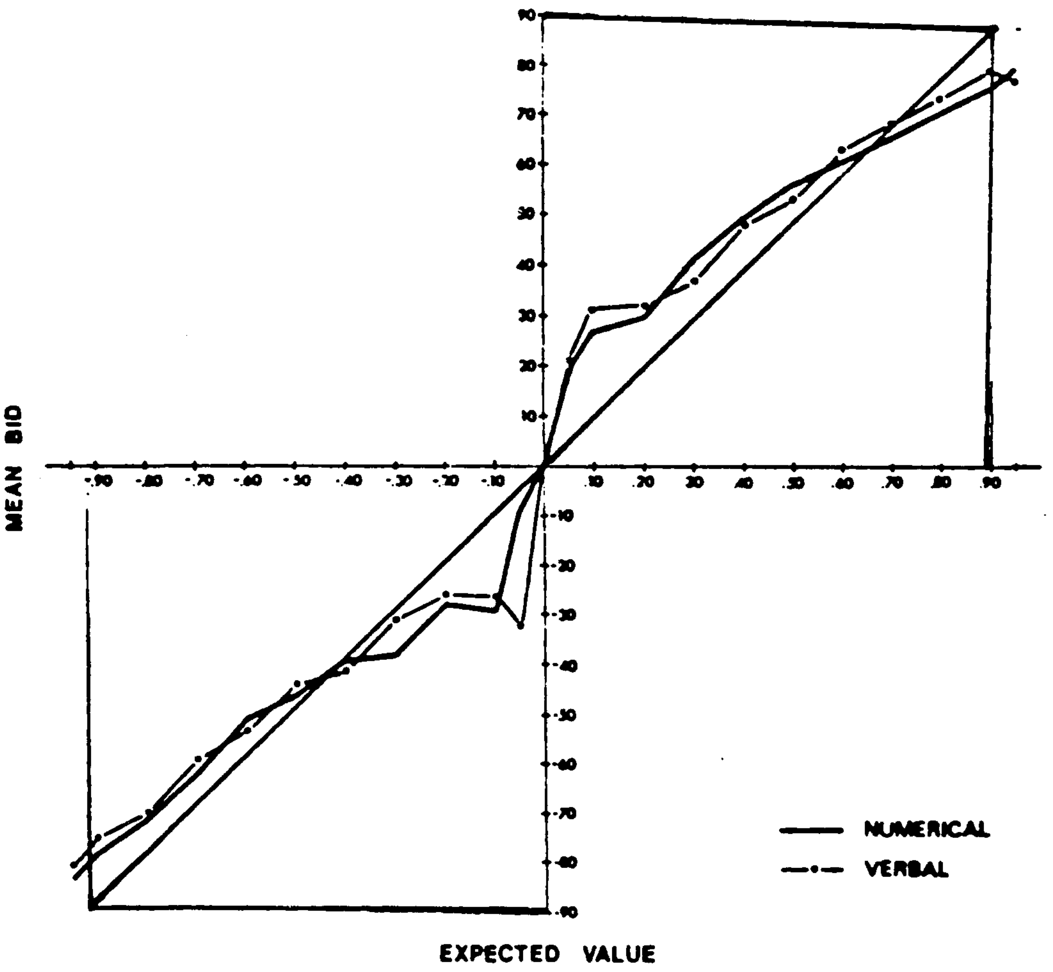


Figure 21. Mean bid in the dyadic experiment as a function of spinner probability and mode of communication.

Table 14

Mean Absolute Bid as a Function of Domain and Expression Mode

Domain	Numeric	Verbal	Mean
Gains	.54	.55	.55
Losses	.49	.49	.49
Mean	.52	.52	.52

function of domain and expression mode. On the average, bids were close to the optimal 0.50 in the domain of losses, but as already seen in Figure 21, this is an artifact of overbidding to low probabilities and underbidding to high ones. In the domain of gains, the average bid is 0.55. The table, therefore, suggests risk seeking in the domain of gains and risk neutrality in the domain of losses, although, as already indicated, the actual explanation in this dyadic situation is more complicated than that. Nevertheless, it must be pointed out that this pattern of results is very different from the usual one seen in individual decision making experiments, such as the previous one or many others in the literature (e.g., Kahneman & Tversky, 1979).

A few main conclusions follow from this study. (1) Decision makers interpret the probability phrases as being less extreme, i.e., closer to 0.50, than the forecasters do. (2) The unusual pattern of bids that on average is close to optimal, actually reflects overestimation of low probabilities and underestimation of high probabilities. The similar pattern of bids in response to numerical and verbal forecasters can be understood by assuming that decision makers treat both kinds of forecasts as imprecise. This latter interpretation must be taken as tenuous until the results are replicated in additional studies. Finally, (3) the general pattern of results can be understood in terms of an overall theory we are designing, and to which we now turn.

9. A THEORY OF JUDGMENT AND CHOICE BASED ON LINGUISTIC PROBABILITIES

In this section we present a tentative theory that ties together the many results described above. The theory is still in an early stage of development and details are subject to change. Nevertheless, it provides a perspective from which the previous work can be understood, as well as a framework for asking additional interesting and useful questions. The main phenomena that we have to explain are the following: Probability phrases have vague meanings to individuals that are systematically affected by context. The context effects thus far demonstrated and that must be handled by the theory include those of event desirability, event base rate, and direction of communication. Despite the fact that linguistic probabilities have vague meanings, they are not responded to in particular choice and judgment situations with much greater variability than are numerical expressions of probabiity. This last result was first evidenced in the Budescu and Wallsten (1985) study, in which individuals consistently rank ordered probability phrases that (it was subsequently learned in other research) are represented within subjects by highly overlapping membership functions. Subsequent demonstrations occurred in the desirability and base rate research in which within subject responses were sufficiently stable to yield large effects of the independent variables. Similarly, in the choice experiments the linguistic gambles were not systematically treated as more vague than the numerical ones and therefore to be avoided, nor was choice variability more extreme in response to the verbal than the numerical gambles.

If one assumes that the vague meaning of a linguistic expression is integrated for the purpose of making a judgment or choice, that the nature of the integration is influenced by context and individual factors, and finally, that the integration is done relatively consistently within a particular situation, then one would expect relatively equivalent response variability to linguistic and numerical expressions of uncertainty, while simultaneously expecting independent variables to have much more profound effects on the interpretation and use of linguistic than on numerical expressions. It is this notion that is at the core of our theory.

We begin by assuming that the meanings of nonnumerical probability expressions for an individual are properly represented by a set of membership functions as illustrated in Figure 22. Because of the Fillenbaum et al. (in preparation) results, we must allow a different set of membership functions, according to whether the individual is selecting the phrases to communicate to another person, or is receiving the phrases from someone else. Indeed, perhaps individuals who work together under pressure or sharing concepts of uncertainty also share the same membership functions. Perhaps also, one attributes differential meanings to the phrases for people from different groups (e.g., politicians vs. weather forecasters vs. physicians). These are intriguing notions that merit investigation, but as of yet we have no data on them. Nevertheless, we expressly do not allow membership functions to vary over context. To do so would be to completely undermine their usefulness as an explanatory construct.

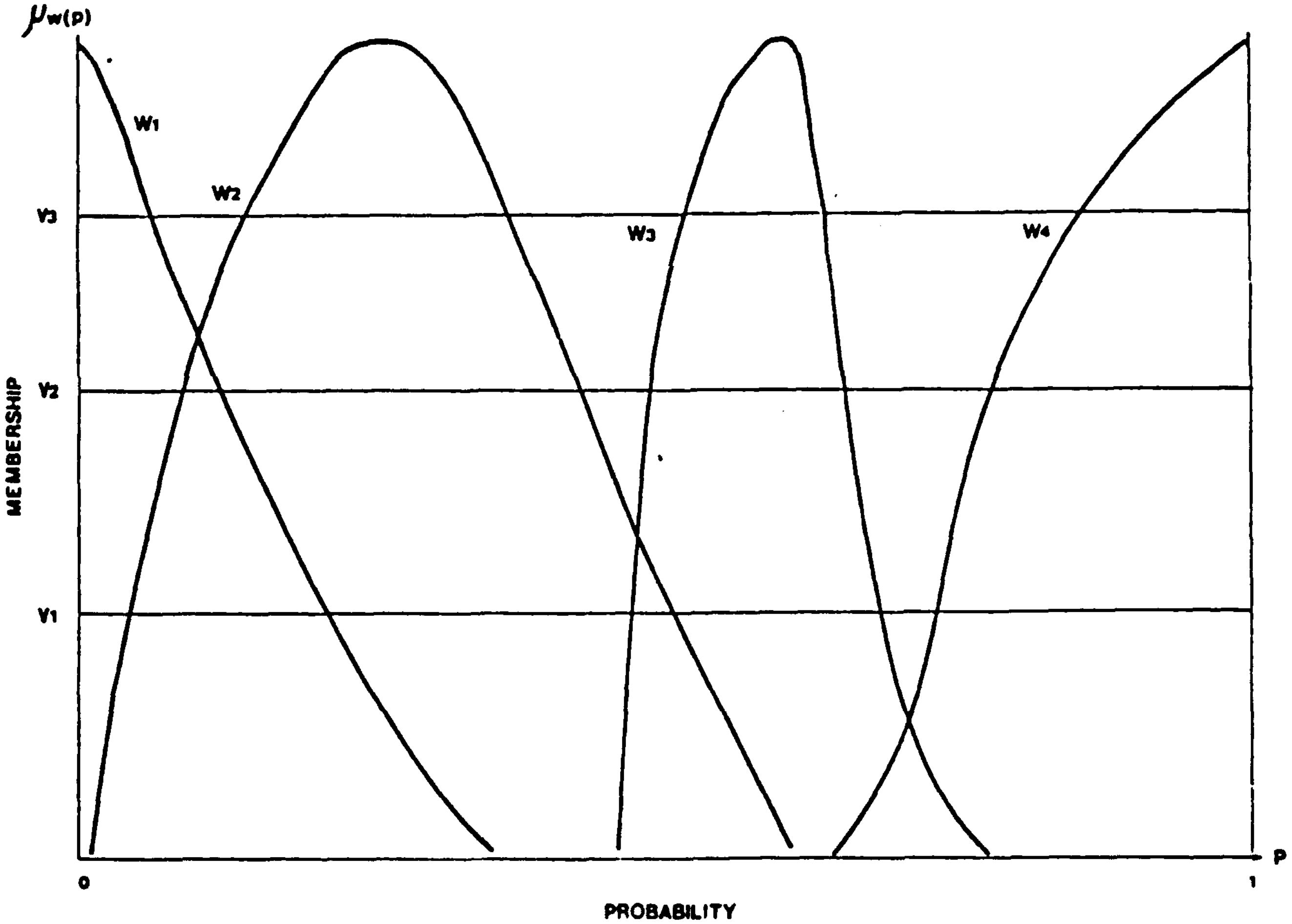


Figure 22. Illustration of membership functions and membership threshold cutoffs.

We assume that when required to make a judgment, choice, or inference on the basis of a linguistic probability, a person first considers the range of probabilities most consistent with the expression. This range can be modeled by assuming that probabilities are only considered if their membership value is greater than or equal to a threshold ν. Three possible thresholds are illustrated in Figure 22. At this point our data do not require the assumption of any threshold membership value, but the assumption seems warranted on other grounds. Specifically, a wide range of literature suggests that under various, but not all, conditions people avoid ambiguity (Curley & Yates, 1985; Einhorn & Hogarth, 1985; Ellsberg, 1961; Becker & Bronson, 1964). Also, on a priori grounds, it would seem that the greater the interval of probabilities under consideration, the more difficult it would be to act. For these two reasons one might postulate that in order to avoid ambiguity and to minimize cognitive effort, the threshold ν is generally kept high.

On other grounds, however, we also postulate that the more important the problem is, the more important it is to consider a fuller range of probabilities. Therefore, the threshold ν is decreased as problem importance increases (e.g., from ν_3 to ν_1 in Figure 22. Two strains of evidence support this assumption. The first is the work summarized by Slovic, Fischhoff and Lichtenstein (1980) which indicates that an important dimension of perceived risk is the amount of information on which a probability judgment is based. The less information that is available, the more dreaded is the risk. This suggests that the more important the problem, the more strongly an individual wishes to consider the full range of probabilities that are consistent with the data. More directly, Einhorn and Hogarth (1985) suggest within the context of their model that individuals adjust their initial probability estimates over wider ranges as the amount of ambiguity in the data increases. Subsequent research will be aimed at determining whether it is necessary to postulate a membership threshold and, if so, the factors that affect its placement. For the moment we make the assumption to achieve greatest generality.

Once an interval of probabilities is determined for a particular problem, the values within it are integrated to yield a single value for action. A family of models is available to represent the integration process. At one extreme it might be assumed that the probability selected for action is that with the maximum value. This simple model can be ruled out immediately, because it implies that context manipulations have no effect on the interpretation of probability terms, and we know that that is not the case. At the other extreme, it might be postulated that the integration process is a weighted averaging one, in which the weight assigned to each probability above the threshold is proportional to its membership value. This assumption is more consistent with the approach underlying fuzzy set theory, but is not sufficient to explain our results if we want to keep membership functions fixed over contexts and decision problems.

We have developed a family of integration models that utilize the membership functions and a context parameter. One special case that is relatively simple to explain and is consistent with all of our results assumes that the integrated value of a probability phrase, I, is the weighted average of three probabilities: p*, which is the probability with maximum membership value, p_{min} which is the minimum probability with membership value at the threshold ν, and p_{max} which is the maximum probability with membership value at the threshold ν. The averaging equation is

$$I = [(1-\alpha)\nu P_{min} + P* + (1+\alpha)\nu P_{max}/(2\nu+1) ,$$

where α is a context parameter and ν is the membership value threshold. This two parameter model is consistent with all the data presented thus far.

The value of α is influenced by the desirability and base rates of the events. Thus, the more desirable is the event being predicted, the greater is α, and the more heavily is the interpretation of the phrase weighted to the higher probabilities. Similarly, the parameter α is proportional to perceived base rate, and therefore so is the value I representing this subject's probability judgments in the Wallsten, Fillenbaum and Cox (1986) studies.

However, recall that base rate had little effect on the low probability terms in those studies. Similarly, it can be seen in Figure 15 that the effect of event desirability was less in the case of the low than the high probability terms. A possible explanation for these results is that in fact the low probability terms have tighter membership functions, or in other words, less vague meanings. Therefore, the integration varies over a smaller range in the case of the low than the high probability terms. In fact, when we look back on the derived membership functions in the Wallsten et al. (1986), Rapoport et al. (in press), and Fillenbaum et al. (in preparation) experiments, this proves to be the case. Why the meanings of low probability terms should be more precise than those of high probability terms is another question, and it is one to which we do not have an answer.

As already indicated, we are willing at this point to allow different membership functions for the selection and evaluation tasks in the Fillenbaum et al. study. However, from the present perspective, one can now understand the relation between the evaluation and comprehension tasks. Recall that the probability judged best for each phrase in the comprehension task fell between the probabilities with the maximum membership value and those calculated as the weighted means of the membership functions. This is because in each case the threshold ν was set relatively high due to the inconsequential nature of the task to the subjects. As a result of the high threshold, most of the tails of the membership functions were cut off, and therefore the weighted average, given as the best probability estimate in the comprehension task, was moved from the weighted average for the full function toward the location of the peak. Similarly, the highest and lowest probability values given in response to a probability term in the comprehension task have membership values above zero in the evaluation task because subjects give the probabilities that have membership value at the threshold.

Recall that in the decision experiment of Budescu, Weinberg, and Wallsten (1986), subjects assigned phrases to spinner displays in stage 1 with considerable variability. In this stage, they were comparing different phrases' membership values for specific probabilities rather than integrating the overall meaning of a particular phrase. Since various phrases have similar membership values at particular probabilities, response variability was high. In stage 2, however, the meanings of selected phrases had to be integrated for the purpose of generating a bid. Although phrases were selected from stage 1 to be "equivalent" to certain numerical probabilities, they were not responded to as such. Specifically, the existence of a threshold ν resulted in the phrases' somewhat overestimating the corresponding probability values, causing bids to them to be more extreme than to the numerical probabilities. Also, because losses loom larger than gains (Kahneman & Tversky, 1979), the value of α is larger in the face of losses than in the face of gains, yielding greater values of I in the former than the latter case.

One result that is left unexplained by our theory is the fact that on the average decision makers bid larger values in response to gains than to losses in the dyadic experiment (Budescu & Wallsten, in preparation). If this result is replicated, it will surely demand some revision in the theory just outlined.

The theory proposed above provides a parsimonious explanation of a wide variety of results in a manner that is consistent with the literature. It is very general, and in that sense perhaps should be thought of more as a theoretical framework than as a specific model. Nevertheless, within this framework the specific assumptions that we have made are easily testable and subject to falsification. Research now under way within this integration framework will result in the assumptions either being supported, modified, or abandoned.

We end this section with a word about the relative optimality of linguistic information processing. As mentioned in the beginning of this report, Zimmer advanced the intriguing suggestion that because humans are accustomed to thinking in verbal rather than numerical ways, their information processing may in fact be more optimal when the information is linguistic than numerical. Without good measurement techniques, such as those described above, it would be impossible to investigate such an hypothesis. Fuzzy optimal models that make use of membership functions can be derived for specific choice and decision situations. Such models, then, can be put into opposition to the information processing model described above. Experiments designed to compare the two models, as well as to compare the relative optimality of choice and decision making in response to numerical and linguistic information are now underway.

10. OTHER WORK

The previous sections outlined the main body of work accomplished during the contract period, and indicated the theoretical and practical insights it provided. However, additional research was carried out as well both to answer subsidiary questions and to open new directions of inquiry. The additional work will be mentioned here for completeness.

Scaling issues. Two technical issues arose while developing the empirical methods for establishing membership functions (Rapoport, et al., in press; Wallsten, et al., 1986). One of them involved the fact that various ratio-scaling models were available for the purpose of deriving scale values from the pair-comparison judgments. The models were not comparable, because each yielded a different goodness-of-fit measure and none had a natural sampling distribution from which inferential statistics could be calculated. Thus, in order to compare the model results, it was necessary to develop sampling distributions from Monte Carlo runs. This work was done for the eigenvector and geometric mean ratio-scaling procedures, and reported by Budescu, Zwick, and Rapoport (1986).

The second technical issue concerns the nature of the variability in membership function values for specific elements in a fuzzy set. An approach to understanding this variability from a Thurstonian perspective was investigated by Zwick (in press).

Combining two non-numerical probabilities. Most of the research focused on how people understand single probability phrases. However, it is not uncommon in real-world situations for people to receive two or more linguistic forecasts before making a decision. For example, one might obtain opinions from two physicians (one saying it is likely you have problem X and the other saying it is doubtful), from two stock analysists, or from two intelligence analysists before taking action.

We have completed two experiments on how people integrate two linguistic probabilities into a single judgment. The first (Wallsten, Zwick, & Budescu, 1985) tested a number of formal models of the integration process taken from fuzzy logic and fuzzy arithmetic. The most successful of the models was one that treated the two probability phrases as fuzzy numbers and the resulting judgment as their fuzzy mean (Dubois & Prade, 1978). The second experiment further tests this conclusion, and attempts as well to predict the single phrase that an individual would use to summarize his or her integrated judgment. Data analysis of this experiment is still in progress.

Other vague descriptors. All the research discussed thus far has concerned nonnumerical probability phrases. In fact, however, subjective uncertainty may be vague within a particular context because features other than the probabilities are described imprecisely. As Figure 23 shows, either or all of the population characteristics, degrees of uncertainty, or events in question may be defined crisply or vaguely. For example, one may know the probability distribution over people's heights in a particular population, and then be interested in the probability of randomly selecting an individual who is between 65 and 70 inches tall. Alternatively, one may know only that the occurence of very short people is doubtful, that of moderately tall people probable, etc, and be interested in the chances of randomly selecting someone of average size.

The three factors shown in Figure 22 combine to yield eight different situations, each with its own uncertainty characteristics. Further, in each case the uncertainty assessment might be numerical or verbal. When all three factors are crisp and assessment is numerical, then classical probability

Crisp Versus Vague Definitions

- Population characteristics
- Crisp f: A → Re
  e.g. f: People → numerical heights
- Vague f: A → {linguistic phrases}
  e.g. f: people → {very short, ...}
- Uncertainty
- Crisp P: Re → [0, 1]
  e.g. P: normal
- Vague P: Y → {linguistic phrases}
  e.g. P: Y → {doubtful, ...}
- Event
- Crisp x1 < X < x2
  e.g. 65 in. < X < 70 in.
- Vague linguistic phrases
  e.g. average size

Figure 23. Three sources of crisp versus vague definitions.

theory applies. However, different forms of fuzziness emerge in the remaining seven cases, for some of which models have been developed. Each of the models provides a means for combining the different sources of vagueness into an overall judgment. These models are discussed by Zwick (in preparation), who has also empirically evaluated four of them. The purpose of this work is to (a) generalize the techniques and results discussed in previous sections, and (b) pave the way for evaluating optimal models and applying decision analysis to these realistic situations. Initial findings of this project, suggesting that three of the four models are reasonably valid, have been reported by Zwick & Wallsten (1986). The relative success of the models in describing subjects' judgments bodes favorably for the extension of the present work to more complex situations, as well as for the development of realistic optimal and cognitive models for the processing of linguistic information.

11. REFERENCES

Bar-Hillel, M. (1983). The base rate fallacy controversy. In R. W. Scholz (Ed.), Decision making under uncertainty. Amsterdam: North Holland.

Becker, S. W., & Brownson, F. O. (1964). What price ambiguity? Or the role of ambiguity in decision-making. The Journal of Political Economy, 72, 62-73.

Borges, M. A., & Sawyers, B. K. (1974). Common verbal quantifier: Usage and interpretation. Journal of Experimental Psychology, 102, 335-338.

Budescu, D. V., & Wallsten, T. S. (1985). Consistency in interpretation of probabilistic phrases. Organizational Behavior and Human Decision Processes, 36, 391-405.

Budescu, D. V., & Wallsten, T. S. (in press). Consistency in interpretation of probabilistic phrases. Organizational Behavior and Human Decision Processes.

Budescu, D. V., & Wallsten, T. S. (in preparation). Dyadic decisions based on linguistic and numerical probabilities.

Budescu, D. V., Wallsten, T. S., & Zwick, R. (in preparation). Integrating the meanings of two probability terms.

Budescu, D. V., Weinberg, S., & Wallsten, T. S. (1986). Decisions based on numerically and expressed uncertainties. IPDM Report No. 39. Haifa, Israel: Universiy of Haifa.

Budescu, D. V., Zwick, R., & Rapoport, A. (1986). A comparison of the eigenvector method and geometric mean procedure for ratio scaling. Applied Psychological Measurement, 10, 69-

78.

Cohen, B. L. (1986). The effect of outcome desirability on comparisons of linguistic and numerical probabilities. Unpublished MA thesis, University of North Carolina at Chapel Hill.

Cohen, J., Dearnley, E. J., & Hansel, C.E.M. (1958). A quantitative study of meaning. British Journal of Educational Psychology, 28, 141-148.

Curley, S. P., & Yates, J. F. (1985). The center and range of the probability interval as factors affecting ambiguity preferences. Organizational Behavior of Human Decision Processes, 36, 273-287.

Dubois, D., & Prade, H. (1978). Operations on fuzzy numbers. International Journal of Systems Science, 9, 613-626.

Einhorn, H. J., & Hogarth, R. M. (1981). Decision theory: Processes of judgment and choice. Annual Review of Psychology, 32, 1-60.

Einhorn, E. J., & Hogarth, R. M. (1985). Ambiguity and uncertainty in probabilistic inference. Psychological Review, 92, 433-461.

Ellsberg, D. Risk, ambiguity, and the savage axioms. Quarterly Journal of Economics, 1961, 75, 643-669.

Fillenbaum, S., Wallsten, T. S., Cohen, B. L., & Cox, J. A. (in preparation). Effects of available vocabulary and mode of communication on the meanings of probability phrases.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. Psychological Review, 80, 237-251.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica, 47, 263-291.

Marshall, E. (1986). Feynman issues his own shuttle report, attacking NASA's risk estimates. Science, 232, 1596.

Pepper, S. (1981). Problems in the quantification of frequency expressions. In D. Fiske (Ed.), New directions for methodology of social and behavior science (9): Problems with language imprecision. San Francisco: Jossey Bass, pp. 25-41.

Pepper, S., & Prytulak, L. S. (1974). Sometimes frequency means seldom: Context effects in the interpretation of quantitative expressions. Journal of Research in Personality, 8, 95-101.

Pitz, G. F., & Sachs, N. J. (1984). Judgment and decision: Theory and application. Annual Review of Psychology, 35, 139-164.

Rapoport, A., & Wallsten, T. S. (1972). Individual decision behavior. Annual Review of Psychology, 23, 131-176.

Rapoport, A., Wallsten, T. S., & Cox, J. A. (in press). Direct and indirect scaling of membership functions of probability phrases. Mathematical Modeling.

Schmucker, K. J. (1984). Fuzzy sets, natural language computations, and risk analysis. Rockville, MD: Computer Science Press.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1977). Behaioral decision theory. Annual Review of Psychology, 28, 1-39.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1980). Perceived risk. In R. Schwing & W. A. Albers, Jr. (Eds.), Societal risk assessment: How safe is safe enough? New York: Plenum, pp. 181-214.

Tversky, A., & Kahneman, D. (1982). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases. Cambridge, England: Cambridge University Press.

Wallsten, T. S. (1971). Subjectively expected utility theory and subjects' probability estimates: Use of measurement-free techniques. Journal of Experimental Psychology, 88, 31-40.

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., Fillenbaum, S., & Cox, J. A. (1986). Base rate effects on the interpretation of probability and frequency expressions. Journal of Memory and Language, 25, 571-587.

Watson, S. R., Weiss, J. J., & Donnell, M. L. (1979). Fuzzy decision analysis. IEEE Transactions on Systems, Man, & Cybernetics, SMC-9, 1-9.

Zadeh, L. A. (1965). Fuzzy sets. Information and Control, 8, 338-353.

Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning, Parts 1, 2, 3. Information Sciences, 8, 199-249; 8, 301-357; 9, 43-98.

Zimmer, A. C. (1983). Verbal vs. numerical processing of subjective probabilities. In R. W. Scholz (Ed.), Decision making under uncertainty. Amsterdam: North-Holland Publishers.

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

Zwick, R. (in press). A note on random sets and the Thurstonian scaling methods. Fuzzy Sets and Systems.

Zwick, R. (in preparation). The use of linguistic probabilities in a fuzzy environment.

Zwick, R., Carlstein, E., & Budescu, D. V. (submitted). Measures of similarity between fuzzy concepts: A comparative analysis.

Zwick, R., & Wallsten, T. S. (1986). Breking the language barrier: Talking about linguistically expressed probabilities. Paper presented at the 19th Annual Math-Psych Meeting. Boston, MA.

APPENDIX: PAPERS AND PRESENTATIONS

Papers published, in press, or submitted

Budescu, D. V. & Wallsten, T. S. (1985). Consistency in interpretation of probabilistic phrases. Organizational Behavior and Human Decision Processes, 36, 391-405.

Budescu, D. V. & Wallsten, T. S. (in press). Subjective estimation of precise and vague uncertainties. In G. Wright & P. Ayton (Eds.), Judgmental Forecasting. Sussex, England: John Wiley & Sons Ltd.

Budescu, D. V., Zwick, R., & Rapoport, A. (1986). A comparison of the eigenvector method and geometric mean procedure for ratio scaling. Applied Psychological Measurement, 10, 69-78.

Rapoport, A., Wallsten, T. S., & Cox, J. A. (in press). Direct and indirect scaling of membership functions of probability phrases. Mathematical Modeling.

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., Fillenbaum, S., & Cox, J. A. (1986). Base rate effects on the interpretation of probability and frequency expressions. Journal of Memory and Language, 25, 571-587.

Zwick, R. (in press). A note on random sets and the Thurstonian scaling methods. Fuzzy Sets and Systems.

Zwick, R., Carlstein, E., & Budescu, D. V. (submitted). Measures of similarity between fuzzy concepts: A comparative analysis.

Papers in preparation

Budescu, D. V., Wallsten, T. S., & Zwick, R. Integrating the meanings of two probability terms.

Cohen, B. L. & Wallsten, T. S. Effects of independent outcome desirability on the meanings of probability phrases.

Fillenbaum, S., Wallsten, T. S., Cohen, B. L., & Cox, J. A. Effects of available vocabulary and mode of communication on the meanings of probability phrases.

Wallsten, T. S. & Budescu, D. V. Judgment and choice on the basis of linguistic probabilities.

Zwick, R, Wallsten, T. S., Kemp, S., & Budescu, D.V. Factors affecting preference for verbal versus numerical communication of uncertainty: Questionnaire and experimental results.

Zwick, R. & Wallsten, T. S., Models of fuzzy probabilities.

Ph.D. Dissertations and M.A. Theses

Cohen, B. L. (1986). The effect of outcome desirability on comparisons of linguistic and numerical probabilities. M. A. Thesis. Chapel Hill, NC: Department of Psychology, University of North Carolina

Zwick, R. (in preparation). The use of linguistic probabilities in a fuzzy environment. Ph. D. Dissertation. Chapel Hill: Department of Psychology, University of North Carolina.

Presentations at Professional Meetings

Wallsten, T. S. (1984). Effects of base rates on meteorologists' interpretations of probability phrases. Paper presented at the Psychonomic Society Meeting. San Antonia, TX. November 9, 1984.

Wallsten, T. S. (1984). Meanings of non-numerical probability phrases. ARI Basic Researcher Contractors Meeting. Fairfax, VA. November, 1984.

Wallsten, T. S. (1985). Meanings of non-numerical probability phrases. ARI Basic Researcher Contractors Meeting. Atlanta, GA. November, 1985.

Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R., & Forsyth, B. (1984). Measuring the vagueness of probability phrases. Paper presented at the 17th Annual Mathematical Psychology. Chicago, IL. August 22, 1984.

Wallsten, T. S., Budescu, D. V., & Zwick, R. (1986). On the representation and use of linguistic probabilities in judgment and decision making. Paper presented at the Annual Meeting of the Judgment/Decision Making Society. New Orleans, LA. November 14-15, 1986.

Wallsten, T. S., Fillenbaum, S., Cohen, B. L., & Cox, J. A. (1986). Interpreting probabilistic phrases: Effects of available vocabulary and communication direction. Paper presented at the Annual Psychonomic Society Meeting. New Orleans, LA. November November 12-14, 1986.

Wallsten, T. S. & Zwick, R. (1986). Judgment on the basis of linguistic probabilities. Paper presented at the Joint

National Meeting of ORSA/TIMS. Miami, FL. October 27-29, 1986.

Wallsten, T. S., Zwick, R., & Budescu, D. V. (1985). The integration of vague information. Paper presented at the 18th Annual Mathematical Psychology Meeting. La Jolla, CA. August 19-21, 1985.

Zwick, R. (1985). Fuzzy probabilities. Paper presented at the 18th Annual Mathematical Psychology Meeting. La Jolla, CA. August 29, 1985.

Zwick, R. & Wallsten (1986). Breaking the language barrier: Talking about linguistically expressed probabilities. Paper presented at the 19th Annual Mathematical Psychology Meeting. Boston, MA. August 19-21, 1986.