ARI Research Note 90-100

Evaluation of a Method of Verbally Expressing Degree of Belief by Selecting Phrases From a List

Robert M. Hamm

University of Colorado

for

Contracting Officer's Representative
Michael Drillings

Basic Research
Michael Kaplan, Director

August 1990

United States Army
Research Institute for the Behavioral and Social Sciences

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

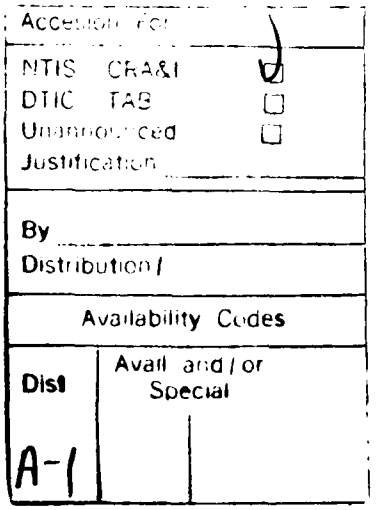
A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

JON W. BLADES
COL, IN
Commanding

Research accomplished under contract
for the Department of the Army

Institute of Cognitive Science, University of Colorado



NOTICES

DISTRIBUTION: This report has been cleared for release to the Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or the National Technical Information Service (NTIS).

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The views, opinions, and findings in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other authorized documents.

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 --	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE --	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) --	5. MONITORING ORGANIZATION REPORT NUMBER(S) ARI Research Note 90-100
6a. NAME OF PERFORMING ORGANIZATION Institute of Cognitive Science 6b. OFFICE SYMBOL (If applicable) --	7a. NAME OF MONITORING ORGANIZATION U.S. Army Research Institute
6c. ADDRESS (City, State, and ZIP Code) University of Colorado, Box 345 Boulder, CO 80309-0345	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 MDA903-86-K-0265
8c. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 61102B; PROJECT NO. 74F; TASK NO. N/A; WORK UNIT ACCESSION NO. N/A

11. TITLE (Include Security Classification)

Evaluation of a Method of Verbally Expressing Degree of Belief by Selecting Phrases From a List

12. PERSONAL AUTHOR(S)

Hamm, Robert M.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Interim	FROM 86/09 TO 88/09	1990, August	49

16. SUPPLEMENTARY NOTATION

Contracting Officer's Representative, Michael Drillings

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)

Probability, Probabilistic inference, Verbal probabilities, Bias

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

This report describes a method for verbal expression of degree of uncertainty. The method requires the subject to select a phrase from a list that spans the full range of probabilities. In a second, optional step, the subject indicates the numerical meaning of each phrase. Alternative list orders were compared to determine the effects of presenting the phrases in ordered sequence or randomly. When the verbal expressions were arranged in random order, ordinal position had a significant effect on the selection of expressions, and the preference for phrases with broader ranges of meaning was stronger in the second half of the list. However, these effects did not occur when the phrases were listed in ascending or descending order. Considerations of accuracy and interpersonal agreement also support the use of ordered phrase lists.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Michael Drillings	22b. TELEPHONE (Include Area Code) (202) 274-8722 22c. OFFICE SYMBOL PERI-BR

DD Form 1473, JUN 86 *Previous editions are obsolete.*

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

EVALUATION OF A METHOD OF VERBALLY EXPRESSING DEGREE OF BELIEF BY SELECTING PHRASES FROM A LIST

CONTENTS

Evaluation of a Method of Verbally Expressing Degree of Belief by Selecting Phrases from a List.

1. Introduction.

The issue whether people think and communicate better using numerical or verbal expressions of probability has received recent attention (Beyth-Marom, 1982; Kong, Barnett, Mosteller, and Youtz, 1986; Wallsten, Budescu, Rapoport, Zwick, and Forsyth, 1986; Zimmer, 1983). In a number of contexts, communication using verbal expressions of probability is preferable, even though it may be less precise than numerical communication (Zwick, 1987). Reasons for this include people's preference for verbal probabilities and the possibility that linguistic terms may facilitate thinking about uncertainty in complex problems (Zimmer, 1983). Accuracy on probabilistic inference word problems is not generally better in either mode (Hamm, 1988).

This paper describes a method designed to avoid several problems that may limit the usefulness of verbal expressions of probability, and reports a study that evaluates possible confounding factors.

Though verbal expression of probability is justified in some situations, it presents problems that must be solved if it is to be broadly useful. First, the meanings of phrases differ between people, although they seem stable for individuals over time (Budescu and Wallsten, 1985) and Kong, Barnett, Mosteller, and Youtz (1986) found no systematic differences between occupational groups. Second, there is an indefinitely large number of words and phrases that could be used to express degree of belief. This makes it difficult to develop a lexicon of the numerical meanings of all verbal expressions of probability. Any new phrase would pose a problem of interpretation, in contrast with a new number that can be easily understood because it can be placed unambigously on the [0,1] number line. The method to be described below solves the problem of individual differences by having the subject assign a numerical value to each phrase (either before or after the phrases have been used in problem solving or communication). It addresses the problem of the indefinitely large lexicon by confining the subject's responses to a limited set of verbal phrases, selected to cover the full range of degrees of belief (though it risks using phrases that subjects may not understand as precisely as they understand their own words).

A third problem with verbal expressions of probability is that the meaning of a phrase may depend on contextual factors. For example, it may depend on the object whose probability is being discussed (Wallsten, Fillenbaum, and Cox, 1986; Mapes, 1979). Thus, "highly likely" may have a different numerical interpretation if applied to the possible failure of a Broadway play than if applied to the possible meltdown of a nuclear reactor. Although this issue is not addressed in this study, the subject's assignment of numbers to phrases could be done on a context specific basis. Next, the meaning of a phrase may depend on the other phrases available in the choice set. For example, the meaning of "probable" may depend on whether "not probable" is present in the list. A term and its negation may mutually influence their meanings to be equidistant from the midpoint of 50%. Two similar terms such as "fairly unlikely" and "somewhat unlikely" may be assigned the same broad meaning if only one of them is in a list, but may be assigned adjacent but non-overlapping meanings if both are present. Finally, a phrase's immediate neighbors in a list may affect its meaning. Thus, "rarely" may mean something different if positioned between "very unlikely" and "absolutely impossible" than if its neighbors are "good chance" and "slightly less than half the time".

A fourth problem is that when subjects read a list of candidate phrases they must do so sequentially. Phrases early in the list may be more likely to be chosen if subjects stop reading after finding one that is good enough. Or phrases late in the list may be favored if subjects read through the whole list and choose from among phrases that are still in short term memory when they finish. Fifth, while the meanings of all verbal expressions of probability may be inherently vague, some phrases may be more vague than others (Wallsten, Budescu, Rapoport, Zwick, and Forsyth, 1986). There are a number of possible mechanisms (detailed below) by which these differences in vagueness might affect the selection of a term from a list.

The problems of context, primacy/recency, and differences in vagueness are addressed experimentally in the present study. Another issue explored is the effect of presenting the phrases in sequential (ascending or descending) or random order. A sequential list would allow a subject to more rapidly find the phrase he or she wants, but it might also constrain the subject's interpretation of the meanings of the phrases. Whether such a constraint is an advantage or disadvantage will be discussed below.

1.1. Description of the method for verbal expression of degree of uncertainty.

In order for subjects to use verbal rather than numerical expressions of probability, while avoiding the requirement of an ever-expanding lexicon, subjects can be asked to select verbal expressions of probability from a pre-defined list. To decrease miscommunication due to individual differences in interpretation of words or phrases, they can be asked in a separate procedure to supply numerical interpretations for the terms in the list.

In the version of the method used in the present study, nineteen verbal expressions covered the range from 0% to 100%, with symmetry about an easily identifiable midpoint ("tossup"). The list was structured so that there was a term for each 5% mark, except that there was only one term in between 25% and 40%, and only one in between 60% and 75%. Other researchers may wish to use shorter (or longer) lists, lists without a sharply defined midpoint, lists that are not balanced around 50% (see Kong, Barnett, Mosteller, and Youtz, 1986) or different phrases. These distinctions, while important, are not pertinent to the present investigation of factors affecting the selection of phrases from the phrase list. The results of this study are applicable to lists comprised of any set of verbal expressions of probability.

The present list was produced by reviewing previous studies that elicited numerical values for verbal expressions of probability (Budescu and Wallsten, 1985; Lichtenstein and Newman, 1967; Simpson, 1944; Shanteau, 1974; Wallsten, Budescu, Rapoport, Zwick, and Forsyth, 1986), in order to identify a set of words and phrases that (a) have interpretations that cover the entire probability range, in about evenly spaced steps, and (b) have relatively narrow interpretations, as indicated by small standard deviations, compared to other candidates with the nearby means (see Table 1). To cover the ends, "absolutely impossible" and "absolutely certain" were chosen. "Almost certain" was used to cover the 95% range; however, it was subsequently learned that Kong, Barnett, Mosteller, and Youtz (1986) had found that people assign this phrase a mean value of .78 (median .90).

****

Insert Table 1 about here.

****

The list of verbal expressions of probability may optionally be presented in sequential order (ascending (Table 1) or descending) or in random order (e.g., Table 2). The subject's instructions are as follows:

> In this study you will be asked to select verbal phrases that represent your estimates of the probability or likelihood that statements are true or that events have happened. Please look over the following list of phrases.
>
> **[The list of verbal expressions of probability was presented.]**
>
> **[Random order conditions:]** The verbal phrases in this list are arranged in random order.
>
> **[Ascending or descending order conditions:]** The verbal phrases in this list are arranged in order. The top ones in the list express a very [high/low] degree of probability, and the bottom ones express a very [low/high] degree. These meanings were determined in surveys of a large number of people.

Table 1

Means and standard deviations of numerical interpretations of verbal expressions of probability measured in previous studies.

Verbal phrase	Mean (Median)	Standard Deviation	Source	Value adopted for this study
Absolutely impossible	~	~	Author	.00
Rarely	.05	.07[a]	S 1944	.05
	.08	.06	B&W 1985	
Very unlikely	.09(.10)	.07	L&N 1967	.10
Seldom	.16	.09	B&W 1985	.15
	.10	.12[a]	S 1944	
	.16(.15)	.08	L&N 1967	
Not very probable	.20(.20)	.12	L&N 1967	.20
Fairly unlikely	.25(.25)	.11	L&N 1967	.25
Somewhat unlikely	.31(.33)	.12	L&N 1967	.33
	.27		Sh 1974	
Uncertain	.41	.13	B&W 1985	.40
	.40(.50)	.14	L&N 1967	
Slightly less than half the time	.45(.45)	.04	L&N 1967	.45
Toss-up	.50(.50)	.00	L&N 1967	.50
	.47	.11	B&W 1985	
	.54		Sh 1974	
Slightly more than half the time	.55(.55)	.06	L&N 1967	.55
Better than even	.58(.60)	.06	L&N 1967	.60
	.66		Sh 1974	
Rather likely	.69(.70)	.09	L&N 1967	.70
Good chance	.74(.75)	.12	L&N 1967	.75
Quite likely	.79(.80)	.10	L&N 1967	.80
Very probable	.87(.89)	.07	L&N 1967	.85
Highly probable	.89(.90)	.04	L&N 1967	.90
	.84		Sh 1974	
Almost certain	~	~	Author	.95
Absolutely certain	~	~	Author	1.00

[a] Interquartile range.

Note: Sources are: B&W = Budescu and Wallsten, 1985; L&N = Lichtenstein and Newman, 1967; Sh = Shanteau, 1974; S = Simpson, 1944. Author = author's judgment.

Table 2
Random Phrase List Order, "Random Order A".

Uncertain
Rather likely
Somewhat unlikely
Rarely
Slightly less than half the time
Good chance
Fairly unlikely
Absolutely impossible
Toss-up
Quite likely
Not very probable
Absolutely certain
Slightly more than half the time
Very probable
Seldom
Almost certain
Better than even
Highly probable
Very unlikely

Please use one of these phrases to answer every question in the problems that follow. It will help the people who will be reading your answers if you will write neatly and write the whole phrase. Do not leave any answers blank!

Please be careful to consider all the possible phrases and select the best one for each answer. To help you consider the phrases for each word problem, you should separate this page from the questionnaire booklet and set it beside the booklet for easy reference as you work on the problems.

At a different time from when the phrase lists are used in the problem solving or communication task (immediately after, in Hamm, 1988), the subjects are asked to refer to the lists and say "the numerical probability that most closely represents what each of these verbal phrases means."

The study's purpose was to investigate whether the following factors influence subjects' use of verbal expressions of probability that are presented in a list: (a) the context, i.e., a phrase's neighbors in the list; (b) the phrase's position in the first or second half of the list; and (c) differences in the vagueness of the phrases. In addition, the effects of appearing in a sequentially ordered versus random list will be investigated. The influence of these factors on (1) subjects' tendency to select a phrase, (2) subjects' assignment of numbers to phrases, and (3) the accuracy of subjects' reasoning (on word problems) using the phrases, will be determined.

2. Method.

One hundred and forty seven subjects from the Introductory Psychology subject pool participated. Each did 4 probabilistic inference word problems (Appendix 1; see Hamm, 1988, for details). Half responded with verbal expressions of probability (the others used numbers). All subsequently assigned numerical values to the phrases. Total response time on the questionnaire was recorded.

One subject was dropped from the analysis for using the wrong response mode. A number of individual responses were dropped because subjects did not follow directions (e.g., used a phrase that was not on the list, assigned a range of values to a phrase, or assigned the same value to every phrase).

The phrases were presented in one of four possible orders. Two of the lists were ordered sequentially, either ascending (Table 1) or descending. The other two lists were arranged in a random order (Table 2), which was produced by folding the ordered list (splitting the list at a phrase, reversing one half, and interleaving the two halves) repeatedly. In answering the problems, 25 subjects used the ascending phrase list, 14 the descending, 16 the random list in Table 2, and 16 the reversed random list. The numbers of subjects assigning values to phrases in the 4 list orders were 48, 27, 31, and 32, respectively. Those subjects who used the verbal response mode subsequently assigned values to phrases that were presented in the same order.

****

Insert Table 2 about here.

****

3. Results.

The presentation of the results will be organized around three topics: the effect of phrase list order on the selection of phrases as answers to the problems, its effect on the numerical values subjects assign to the phrases, and finally its effect on the accuracy of the subjects' answers to the problems.

3.1. Effect of phrase list order on problem answers.

Subjects selected phrases from a list to express their estimates of the probability of a hypothesis sixteen times: four times (after 0, 1, 2, and 3 pieces of key information had been provided; see Appendix 1) in each of 4 problems (concerning Cabs, Doctors, Insurance, and Twins; see Hamm, 1988).

3.1.1. Preference for phrases in particular ordinal positions.

To reveal preferences for phrases presented in particular positions in the phrase list, consider the answers after all 3 pieces of information were provided (Table 3). The phrases in the first and last positions were rarely used by either the 39 subjects presented with ascending or descending lists (where the extremes were "absolutely impossible" and "absolutely certain"), or the 32 subjects presented with random lists (whose extremes were "uncertain" and "very unlikely"). However, the phrases *next to* the extreme positions were chosen frequently in the Cab and Doctor problems. There is thus evidence that phrases that appear both early and late in the list are used. Further, in the Cab and Insurance problems the phrase in the middle position is used frequently. This may be due to its meaning ("tossup", in the middle of the ordered lists, expresses "I don't know"), rather than its location (10th in a list of 19 phrases).

Insert Table 3 about here.

3.1.2. Preference for phrases in the first or second half of the list.

Because the identity of the phrase occupying a particular position varies across phrase lists, we must consider the lists separately. The sequentially ordered and the random phrase lists were each presented in two orders that are reverses of one another. Comparison of the reversed lists can reveal the overall tendency to pick answers that are early or late, separate from the identities of the phrases. The average ordinal position of the phrases subjects selected from each list is given for all 16 problem answers in Table 4. If there were no effect of ordinal position, the unweighted mean ordinal position of the selected phrase for the ordered lists (or random lists) would be 10. (The unweighted mean is taken, to control for different numbers of subjects using the ascending and descending lists.) Looking over all four answers for all problems, the mean ordinal position of the chosen phrases is 9.74 for the ordered lists, an average of one quarter position (out of 19) in front of the midpoint. For the random lists, the mean ordinal position is almost exactly the middle position, 10.

Insert Table 4 about here.

When there are 0, 1, or 2 pieces of information in the word problems, the answers frequently are strongly constrained (see Hamm, 1987), and so little effect of list reversal would be expected. Looking therefore at only the answers after all three key pieces of information had been presented, there is a slightly larger effect of position in the list. The mean ordinal position of the answers is 1/3 of a position in front of the midpoint for the ordered lists, and 4/5 of a position after the midpoint for the random lists. The small magnitude of this effect suggests that position in the list has little effect on the probability that a phrase will be used.

3.1.3. Comparison of ordinal position effect on phrase selection in random and ordered lists.

Though the overall effect of ordinal position is small, there may be differences between the ordered and random lists in the magnitude of the effect, which would have implications for the design of the optimal method for selecting verbal expressions of probability. In order to measure the effect of ordinal position on the tendency of subjects to select individual phrases, so that the ordered and random lists may be compared, an index was computed for each phrase, measuring its tendency to be used more when it appears in the first half than the second half of the list. First, a measure $D_{i,a,b,p}$ is computed separately for each phrase in each problem, separately for the

Table 3

Number of subjects who chose the phrase occupying each ordinal position in the list for their final answer on each problem.

	Cab	Doctor	Insurance	Twins	Total
Position 1	0	0	0	4	4
Position 2	6	10	2	2	20
Position 3	3	5	4	4	16
Position 4	4	5	3	1	13
Position 5	2	4	4	6	16
Position 6	3	3	5	3	14
Position 7	2	1	4	4	11
Position 8	1	1	5	3	10
Position 9	1	1	3	3	8
Position 10	6	2	10	4	22
Position 11	1	2	3	6	12
Position 12	0	0	4	2	6
Position 13	6	4	0	9	19
Position 14	9	4	6	6	25
Position 15	6	5	3	3	17
Position 16	5	5	4	0	14
Position 17	6	4	8	6	24
Position 18	10	13	3	1	27
Position 19	0	1	0	3	4
Total	71	70	71	70	282

Table 4

Mean ordinal position of phrase selected (from list of 19) for each phrase list.

		Ordered Lists			Random Lists		
		Asc	Desc	Unwtd Mean	Ran A	Ran B	Unwtd Mean
	N	25	14		16	16	
Problem	Amt of info						
cab	0	9.38	9.64	9.51	7.94	11.94	9.94
cab	1	10.92	9.71	10.32	10.25	10.50	10.38
cab	2	14.64	5.07	9.86	10.81	6.44	8.63
cab	3	15.04	5.79	10.41	11.81	11.06	11.44
doc	0	9.20	9.57	9.39	7.63	12.19	9.91
doc	1	5.36	15.29	10.32	12.88	8.27	10.57
doc	2	17.44	2.21	9.83	13.87	5.38	9.62
doc	3	15.00	4.00	9.50	9.60	10.50	10.05
ins	0	7.80	12.93	10.36	6.63	11.81	9.22
ins	1	8.16	10.93	9.54	7.75	11.38	9.56
ins	2	9.60	9.36	9.48	8.75	11.38	10.06
ins	3	10.68	8.29	9.48	12.13	10.00	11.06
twn	0	9.32	10.00	9.66	7.31	11.06	9.19
twn	1	9.16	9.64	9.40	6.38	13.67	10.02
twn	2	11.92	7.21	9.57	9.06	9.75	9.41
twn	3	9.88	8.64	9.26	12.19	9.40	10.79
Mean of all amounts of information:				9.74			9.99
Mean of 3-inf problems:				9.66			10.84

ordered and random lists:

$$D_{i,a,b,p} = 100 \times (H_a \times \frac{C_{i,a,p}}{N_a} + H_b \times \frac{C_{i,b,p}}{N_b}),$$

where *i* indexes the 19 phrases; *a* is a particular phrase order (ascending or random A) and *b* is its reverse (descending or random B); *p* signifies the particular problem; H_j is 1 if the phrase appeared in the first half of list *j*, 0 if in the middle, and -1 if in the second half of the list; $C_{i,j,p}$ is the count of subjects using list *j* on problem *p* who chose phrase *i* as their answer; and N_j is the total number of subjects using list *j*. The $D_{i,a,b,p}$ indices for each phrase, each problem, are presented in Table 5. In the sequentially ordered lists, there are approximately the same number of phrases that have negative and positive indices. However, in the random lists, there are more phrases with negative indices. This suggests that in random lists subjects tend to select phrases that are in the second half of the list.

Insert Table 5 about here.

The mean of the index, across the 18 phrases that are not in the middle of the list (the $D_{i,a,b,p}$ for the middle phrase is 0), is given by:

$$D_{a,b,p} = \frac{\sum_{i=1}^{19} D_{i,a,b,p}}{18}.$$

This mean is produced separately for the ordered and random lists, for each problem. In addition, an overall index is produced for the ordered and the random lists, by averaging over the 4 problems:

$$D_{a,b} = \frac{\sum_{p=1}^{4} D_{a,b,p}}{4}.$$

The two elements of the $D_{i,a,b,p}$ index are the percents of subjects who chose the phrase when it appeared in two lists that have reversed orders. Every subject chose one phrase on each problem, and there were 19 phrases, so the percent of subjects expected to choose each phrase is 5.263%. The index subtracts the percent choosing the phrase when it is in the second half of the list from the percent choosing the phrase when it is in the first half of the list. The expected difference is 0% if ordinal position has no effect. A positive index would signify that subjects chose the phrase more often when it appeared in the first half of the list. Table 6 shows the mean $D_{a,b,p}$ and $D_{a,b}$ indices for the Ordered (*a* = ascending, *b* = descending) and Random (*a* = random order A, *b* = random order B) phrase list orders, for the subjects' final answers on each problem. These represent the average difference in the percent of subjects choosing a phrase when it is in the first compared to the second half of the list. Dividing the mean answer by 5.263% expresses the ordinal-position effect as a proportion of the percent of subjects expected to use the average phrase (Columns 2 and 6 of Table 6). Table 6 also shows the standard deviation of the index across the 18 phrases, and the t-test for whether the mean is different from 0%.[1]

Insert Table 6 about here.

The $D_{a,b,p}$ index is positive (indicating a tendency to choose phrases early in the list) for 3 of the 4 problems when the lists were ordered, but negative (indicating preference for phrases in the

second half of the list) for all problems when the lists were random. In the ordered lists the $D_{a,b}$ index (averaged over problems) indicates a .064% preference for phrases in the first half of the list, which is a .012 proportion of the expected 5.26%. Thus, a phrase from the first half of the list would be used 5.327% of the time on average, compared with 5.199% for a phrase from the second half of the list. In the random lists the $D_{a,b}$ index is -.844% (a .16 proportion of the expected 5.26%), reflecting a preference for the phrases in the second half of the list. A phrase from the second half of a random list would be used 5.685% of the time, while a phrase from the first half would be used 4.841% of the time. These small effects are not significant for the individual problems, although when averaged across problems the tendency of subjects faced with randomly ordered lists to select phrases in the second half of the list is statistically significant ($t = -2.861$, $p < .02$).

The $D_{a,b,p}$ index allows a statistical test of whether there is a difference between the ordered and random lists in the direction and extent of the ordinal position effect. With the random lists, the phrase chosen for the final answers in all problems tended to come from the second half of the list, but with the ordered lists there was a slight preference for the first half. The difference in ordinal position effect between the ordered and random lists is shown in Table 7. The effect is very small -- the mean difference is .908% (a .173 proportion of the 5.263% of the subjects expected to select a given phrase) -- although the difference is statistically significant for the overall indices (and marginally so for the Insurance problem).

Insert Table 7 about here.

3.1.4. Effect of list reversal on the selection of phrases with broad and narrow membership functions.

Verbal expressions of probability differ in the range of numerical probabilities to which they refer. Some phrases, such as "absolutely certain" and "tossup," would be expected to refer to narrow ranges of probabilities (see also Kong, Barnett, Mosteller, and Youtz, 1986), while other phrases, particularly those with meanings near 25% or 75%, would refer to broader ranges. The tendency to use a phrase with a broad "membership function" (Wallsten, Budescu, Rapoport, Zwick, and Forsyth, 1986) may be more strongly affected by its ordinal position in a list than the tendency to use a phrase with a narrow range. Broad phrases may be strongly affected even though when all phrases are considered, as in the above analysis, the ordinal position effects are very small. In order to measure the breadth of the membership functions of the 19 verbal expressions of probability used in this study, an auxilary study was carried out.

Method. Sixty-five subjects, primarily from the Introductory Psychology subject pool, filled out a questionnaire (Appendix 2) which asked them to state the lower and upper bounds of the numerical probabilities that each phrase refers to. Half of the subjects named the lower bound for each phrase before the upper bound, and half did the reverse. Crossed with this factor, half of the subjects named the phrases in Random Order A (Table 2), and half in its reverse, Random Order B.

Results. The mean lower and upper limits, across all conditions, are presented in Columns 1 and 2 of Table 8. The midpoint between these bounds is an estimate of the meaning the individual assigns to the verbal expression of probability. The mean and median midpoints of these ranges and their standard deviation (Columns 3, 4, and 5) can be compared with the values in Table 1. The 6th column shows the standard deviations of the differences between the upper and lower bounds, which reveal that there is an exceptionally high variation across subjects (s.d. = .336) in the range of meaning attributed to "uncertain".

Insert Table 8 about here.

The difference between a phrase's upper and lower bounds is a measure of the range of meaning the individual assigns to the phrase, and can be used as an estimate of the breadth of the

Table 5

The ordinal position effect indices for each phrase, $D_{i,a,b,p}$ for each problem and $D_{i,a,b}$ for all problems.

	$D_{i,a,b,p}$								$D_{i,a,b}$	
	Cab		Doctor		Insurance		Twins		All	
	Ord	Ran	Ord	Ran	Ord	Ran	Ord	Ran	Ord	Ran
Absolutely imp.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Rarely	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Very unlikely	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.3	0.1	0.1
Seldom	-3.6	0.0	0.0	0.0	0.0	0.0	2.1	0.0	-0.4	0.0
Not very prob.	2.1	0.0	0.0	0.0	0.0	0.0	8.5	-3.2	2.7	-0.8
Fairly unlikely	0.0	0.1	0.0	0.0	0.0	0.0	-1.6	-3.0	-0.4	-0.7
Somewhat unlik.	-3.7	-6.3	0.0	6.3	4.2	-6.3	-10.6	-3.1	-2.5	-2.3
Uncertain	0.0	0.0	0.0	0.0	0.5	0.0	2.1	3.3	0.6	0.8
Slightly less than 1/2 time	0.0	0.0	0.0	-3.1	0.0	0.0	-1.6	-6.3	-0.4	-2.3
Toss-up	~[a]	0.0	~	3.1	~	0.0	~	0.0	~	0.8
Slightly more than 1/2 time	1.6	-3.1	-2.1	0.0	0.0	0.0	-4.2	0.0	-1.2	-0.8
Better than even	3.7	-3.0	0.0	3.1	-2.1	-3.0	3.2	-5.9	1.2	-2.2
Rather likely	-2.6	-0.2	3.7	-6.6	3.7	3.1	-2.6	3.1	0.6	-0.1
Good chance	-6.3	-6.2	-2.1	0.0	5.3	-3.1	-0.5	-3.1	-0.9	-3.1
Quite likely	-8.8	~	3.2	~	-4.2	~	5.3	~	-1.1	~
Very probable	0.9	-3.1	5.1	-6.3	1.5	3.1	0.0	-3.1	1.9	-2.3
Highly prob.	4.5	-6.0	-0.6	6.3	-2.1	-2.9	3.6	0.0	1.3	-0.7
Almost certain	1.1	3.2	-2.4	-2.9	0.0	-6.1	1.8	0.0	0.1	-1.4
Absolutely cert.	0.0	0.0	-2.1	0.0	0.0	0.0	0.0	0.0	-0.5	0.0
# phrases with positive index	6	2	3	4	6	2	7	3	8	3
# phrases with zero index	7	9	10	10	9	11	5	8	2	4
# phrases with negative index	5	7	5	4	3	5	6	7	8	11

[a] ~ indicates that the value was not calculated for a phrase because it appeared in the central position in the list.

Table 6

Mean ordinal position effect indices, for the 18 phrases that are not in the middle position.

	Ordered Lists (a=Ascending, b=Descending)				Random Lists (a=Random A, b=Random B)			
	Mean	Proportion	St Dev	t	Mean	Proportion	St Dev	t
Index $D_{a,b,p}$								
Cab	-.608	-.116	3.304	-.737	-1.363	-.259	2.637	-1.860
Doctor	.153	.029	2.008	.305	-.006	-.001	3.413	-0.007
Insurance	.404	.076	2.238	.723	-.842	-.160	2.545	-1.326
Twin	.306	.058	4.035	.304	-1.167	-.222	2.644	-1.769
Index $D_{a,b}$	.064	.012	1.213	.211	-.844	-.160	1.182	-2.861*

* p = .02, 2-tailed

Table 7

Difference between the ordinal position effect index scores of the ordered and random lists.

	Difference	Proportion	t value	2-tail prob
$D_{a,b,p}$ difference.				
Cab	.755	.143	1.55	.141
Doctor	.159	.030	0.14	.889
Insurance	1.246	.237	1.92	.073
Twin	1.473	.280	1.08	.296
$D_{a,b}$ difference.				
All problems	.908	.173	2.86	.011*

Note: Because a different phrase was dropped (for being in the middle location) from the ordered list than from the random list, N = 19 - 2 = 17 and df = 16.

Table 8
The mean lower limit, upper limit, and midpoint between the limits, for each phrase.

Phrase	Lower Limit	Upper Limit	Midpoint (average)			S.D. of Range
			Mean	Median	St Dev	
Absolutely imposs.	.007	.041	.024	.000	.080	.133
Rarely	.064	.183	.117	.100	.085	.079
Very unlikely	.046	.145	.096	.075	.071	.061
Seldom	.117	.243	.180	.150	.126	.090
Not very probable	.113	.235	.174	.150	.116	.078
Fairly unlikely	.176	.287	.231	.225	.111	.061
Somewhat unlikely	.217	.349	.283	.250	.129	.065
Uncertain	.294	.534	.414	.500	.153	.336
Slightly less than half the time	.390	.470	.430	.440	.050	.048
Toss-up	.484	.526	.505	.500	.045	.087
Slightly more than half the time	.524	.604	.564	.555	.065	.053
Better than even	.540	.703	.621	.600	.088	.123
Rather likely	.585	.735	.660	.700	.222	.087
Good chance	.652	.799	.726	.750	.137	.077
Quite likely	.686	.824	.755	.800	.146	.087
Very probable	.733	.872	.803	.850	.122	.093
Highly probable	.757	.899	.828	.850	.127	.084
Almost certain	.840	.950	.895	.925	.088	.090
Absolutely certain	.928	.980	.954	.100	.105	.142

phrase's membership function. The mean and median range for each phrase are presented in Table 9. "Absolutely impossible" (.034), "tossup" (.042), and "absolutely certain" (.052) have the narrowest ranges, and "uncertain" (.240) and "better than even" (.163) have the widest ranges. The median range measure does not discriminate well among the phrases, for its value for many of the phrases was .10. For comparison with previous work, Column 3 shows the difference between the median upper bound and the median lower bound for the three phrases studied by Wallsten, Budescu, Rapoport, Zwick, and Forsyth (1986) that were used in the present study. These ranges are generally larger than the individual ranges measured in our study (Columns 1 and 2), which may reflect different ellicitation procedures. The standard deviation of the values assigned to a verbal expression of probability can be considered an alternative measure of the breadth of the phrase's membership function, although it is confounded with individual differences in the meaning of the phrase. Column 4 of Table 9 shows the mean of the standard deviations of the values given to phrases when presented in the four list orders in the main study. Column 5 shows the standard deviation of the midpoints of the ranges, from the auxilary study.

****

Insert Table 9 about here.

****

The intercorrelations among these five measures of breadth of membership function are all fairly high, ranging from .55 to .86 (Table 10). This indicates that when a direct measure of the breadth of membership function is lacking, the standard deviation might serve as a useful proxy.

****

Insert Table 10 about here.

****

The question whether subjects prefer to use phrases with broader meanings is addressed in Table 11, which shows the correlations between the indices of breadth of membership function and measures of the number of subjects who used each phrase for each problem, separately for the ordered and random lists. The relations are generally positive, especially for the Cab and Insurance problems. While this suggests people prefer to use phrases with broad, even vague, meanings, it may be due to preferences to answer these problems with particular degrees of probability, e.g., answers between .10 and .40 or between .60 and .90. Further study is needed to clarify this issue.

****

Insert Table 11 about here.

****

To test whether the range of a phrase's meaning influences the impact of ordinal position on subjects' tendency to select it from the list when answering a word problem, Table 12 shows correlations between measures of the breadth of membership function (from Table 9) and the $D_{i,a,b,p}$ and $D_{i,a,b}$ measures of the effect of ordinal position on the probability of selecting phrases (from Table 5). The $D_{i,a,b,p}$ measures are positive for a phrase if it is more likely to be used when in the first rather than the second half of a list. Therefore, a positive correlation in Table 12 means that the broader the membership function of the phrase, the more it is likely that the phrase will be used more when it is in the first half of the list, or equivalently, the narrower the phrase's meaning the more likely that it will be used more when in the second half of the list. There were no significant effects for the mean range measure, which is our best measure of breadth of membership function. The median range measure correlated negatively with the $D_{i,a,b}$ index for the random lists, for all problems, suggesting that subjects' tendency to select phrases from the second half of random lists (noted above) is stronger for phrases with broad than with narrow ranges of meaning. (Note that 14 of the phrases are defined as "broad", 12 of them with median ranges of .10.) The SD of range midpoints measure showed a number of significant positive and negative correlations which are hard to interpret because the measure possibly confounds the breadth of the phrases' interpretations with individual differences in their interpretations.

****

Insert Table 12 about here.

In conclusion, even considering variations in the breadth of the phrases' meanings, there is only very weak evidence that there is any effect of ordinal position in the phrase list on the tendency to select particular phrases as the answers to word problems.

3.1.5. Difficulty finding the desired phrase in ordered and random lists.

A possible advantage of an ordered as opposed to a random phrase list is that subjects can find the verbal expression they want more easily. There are a number of possible reasons for this advantage. Subjects who know the phrase they want may be better able to verify its presence in a list that is structured in an ascending or descending order. Subjects who know the probability they want to express (either as a number, a range of numbers, a verbal phrase not in the list, or an idea that is not modally specific) may be able to find an appropriate expression more easily, presumably by evaluating the available phrases, when those phrases are ordered. Another possibility is that people may not know what degree of probability they want to express until they have considered candidate phrases. If so, it may be easier to check whether the meanings of the phrases apply to the situation when using an ordered list, in which phrases' meanings can be quickly understood because they are implied by the meanings of their neighbors in the list.

Any of these ordered list advantages might result in faster response time. The time to complete the whole questionnaire includes time reading and responding to all four problems, as well as time assigning values to all 19 phrases. Analysis of response time as a function of whether the list was ordered or random and of presentation mode shows that the subjects took only 10 seconds longer on the average (out of 18 minutes) on the random lists, which is not significant in a 2 (list order) X 2 (presentation mode) ANOVA. Therefore the admittedly rough measure of total questionnaire response time gives no indication that responding using random phrase lists is more difficult than responding using ordered lists.

A second measure of whether the ordered phrase lists are easier to use than the random lists is the variability of the meanings of the phrases subjects select as answers for the word problems. If we assume that subjects know the probability they want to express and have more trouble finding a verbal expression that fits it well when they are searching a random list, then we would expect that the numerical values of the phrases selected will be more variable with the random lists. If we assume that at the outset subjects don't know the probability that they want to express, and discover it by looking at phrases and seeing which one "seems right", then we would expect that the context variability in the random lists will cause a wider variation in the subjects' interpretations of the phrases when deciding which one to select. Either way, we expect that the random phrase lists will produce higher variability in the meaning of the answers than the ordered lists.

To measure variability in the meanings of the phrases subjects selected, it is necessary to use the *a priori* values (Column 4 of Table 1). (Use of the subjects' own assigned values would confound list differences in variation in meaning of the selected phrase with list differences in variation of the values subjects subsequently assigned to the phrase.) Table 13 shows the means and standard deviations of the *a priori* values of the selected phrases. While the phrases chosen from the random lists had numerical interpretations with higher average standard deviations (.195) than those chosen from the ordered lists (.181), and this was true for 11 of the 16 subproblems, the difference is not statistically significant ($Chi^2 = 1.25$). Therefore the random lists have only a slight tendency to produce answers with higher variability.

Insert Table 13 about here.

If it is more difficult to make fast, accurate use of a random list of verbal expressions of probability than an ordered list, we have not been able to measure it.

Table 9
Measures of breadth of membership function.

	Measures derived from subjects' estimates of upper and lower bounds			Measures derived from standard deviations of estimates of meanings.	
Phrase	Mean Range[a]	Median Range[a]	U-L diff.[b]	SD of meaning[c]	SD of U-L midpoint[d]
Absolutely impossible	.034	.00		.074	.080
Rarely	.106	.10		.078	.085
Very unlikely	.100	.10		.132	.071
Seldom	.126	.10		.089	.126
Not very probable	.122	.10		.090	.116
Fairly unlikely	.112	.10		.083	.111
Somewhat unlikely	.132	.10		.092	.129
Uncertain	.240	.10		.122	.153
Slightly less than half the time	.080	.07		.046	.050
Toss-up	.042	.00	.13	.021	.045
Slightly more than half the time	.080	.09		.061	.065
Better than even	.163	.10		.074	.088
Rather likely	.150	.15		.121	.222
Good chance	.147	.15	.46	.106	.137
Quite likely	.137	.10		.087	.146
Very probable	.139	.10		.084	.122
Highly probable	.142	.10		.071	.127
Almost certain	.109	.10	.11	.087	.088
Absolutely certain	.052	.00		.008	.105

[a] Difference between upper and lower bounds, auxiliary study, N = 65.

[b] Difference between median upper bound and median lower bound, from Figure 4 of Wallsten, Budescu, Rapoport, Zwick, and Forsyth (1986).

[c] Mean of standard deviations of values assigned to the phrases, from four lists with different phrase orders, main study, N = 138.

[d] Standard deviation of midpoint (average of upper and lower bounds), auxiliary study.

Table 10

Intercorrelations among five measures of breadth of membership function.

	Mean Range	Median Range	U-L Diff.	SD of Value
Median Range	.73**			
U-L Difference	.74	.72		
SD of Value	.67**	.76**	.63	
SD of Midpoint	.65**	.59**	.86	.55**

Note: Indices are defined in notes to Table 9. N = 19 for every correlation except those involving the U-L Difference index, for which N = 3.

**$p < .01$.

Table 11

Correlations between indices of breadth of membership function and measures of the number of subjects who selected each phrase for each problem, separately for ordered and random phrase lists.

Problem List structure	Measures of Breadth of Membership Function Mean Range	 Median Range	 SD of Values	 SD of Midpoints
Cab				
ordered	.26	.40*	.19	.37+
random	.35+	.51*	.27	.13
total	.31+	.47*	.23	.17
Doctor				
ordered	.07	.12	-.06	.18
random	.17	.29	.14	.38+
total	.14	.24	.04	.32+
Insurance				
ordered	.48*	.28	.46*	.26
random	.16	.12	-.05	.15
total	.40*	.25	.27	.26
Twins				
ordered	.19	.26	.05	.22
random	.06	.02	.12	-.28
total	.15	.17	.11	-.06
All problems				
ordered	.35+	.43*	.22	.41*
random	.29	.38+	.22	.29
total	.35+	.43*	.24	.38+

+ $p < .10$; * $p < .05$; ** $p < .01$.

Table 12

Correlations between measures of breadth of membership function and indices of the effect of ordinal position on phrase selection, for ordered and random phrase lists.

	Measures of Breadth of Membership Function			
	Mean Range	Median Range	SD of Values	SD of Midpoints
Index $D_{a,b,p}$				
Cab				
Ordered	-.10	-.25	-.26	-.43*
Random	-.27	-.31	-.07	-.21
Doctor				
Ordered	.29	.25	.32	.49*
Random	-.02	-.22	-.21	-.25
Insurance				
Ordered	.10	.35+	.33	.33+
Random	-.10	-.11	-.01	.17
Twins				
Ordered	.19	-.01	.03	.03
Random	.11	-.05	.25	.38+
Index $D_{a,b}$				
All problems				
Ordered	.25	.09	.14	.09
Random	-.16	-.41*	-.06	.00

Note: For ordered lists, *a* = ascending and *b* = descending; for random lists, *a* = random order A and *b* = random order B.

+$p < .10$; *$p < .05$.

Table 13

Means and standard deviations of the a priori values of the phrases subjects selected to answer the problems.

		Ordered Lists		Random Lists		
		Mean	SD	Mean	SD	List with bigger SD
cab	0	.474	.125	.489	.154	r
	1	.542	.286	.606	.276	o
	2	.781	.106	.801	.144	r
	3	.776	.195	.738	.196	r
doc	0	.482	.112	.516	.161	r
	1	.216	.159	.247	.203	r
	2	.927	.089	.923	.074	o
	3	.813	.162	.737	.201	r
ins	0	.356	.187	.368	.201	r
	1	.413	.251	.437	.263	r
	2	.500	.242	.578	.278	r
	3	.563	.258	.678	.239	o
twn	0	.477	.084	.525	.112	r
	1	.478	.232	.437	.209	o
	2	.632	.175	.566	.174	o
	3	.525	.229	.434	.240	r
Mean SD:			.181		.195	

Note: N = 39 for the ordered lists and 32 for the random lists.

Table 14

Mean numerical values assigned to each phrase, each list order.

	Ordered Lists						Random Lists						Total	
	Ascending		Descending		Both		List A		List B		Both			
N:	48		28		76		32		32		64		140	
	Mn	SD	Mn	SD	Mn	SD	Mn	SD	Mn	SD	Mn	SD	Mn	SD
Absolutely imp.	.021	.130	.000	.002	.011	.066	.002	.005	.035	.160	.019	.083	.015	.074
Rarely	.104	.096	.084	.042	.094	.069	.1[illegible]	.077	.161	.098	.144	.088	.119	.078
Very unlikely	.128	.053	.135	.041	.132	.047	.162	.175	.209	.260	.186	.218	.159	.132
Seldom	.180	.057	.171	.049	.176	.053	.202	.127	.229	.121	.216	.124	.196	.089
Not very probable	.235	.094	.225	.048	.230	.071	.197	.095	.199	.124	.198	.110	.214	.090
Fairly unlikely	.283	.073	.281	.039	.282	.056	.255	.105	.266	.113	.261	.109	.271	.083
Somewhat unlikely	.334	.075	.332	.052	.333	.064	.326	.147	.278	.092	.302	.120	.318	.092
Uncertain	.404	.073	.407	.055	.406	.064	.405	.193	.399	.167	.402	.180	.404	.122
Slightly less than half the time	.448	.045	.445	.036	.447	.041	.435	.030	.428	.072	.432	.051	.439	.046
Toss-up	.496	.040	.493	.038	.495	.039	.500	.000	.501	.004	.501	.002	.498	.021
Slightly more than half the time	.546	.053	.561	.053	.554	.053	.580	.046	.552	.090	.566	.068	.560	.061
Better than even	.598	.069	.608	.063	.603	.066	.609	.069	.627	.095	.618	.082	.611	.074
Rather likely	.671	.080	.680	.070	.676	.075	.738	.121	.619	.211	.679	.166	.677	.121
Good chance	.719	.081	.735	.071	.727	.076	.775	.112	.663	.158	.719	.135	.723	.106
Quite likely	.776	.083	.783	.063	.780	.073	.796	.084	.720	.117	.758	.101	.769	.087
Very probable	.827	.084	.844	.062	.836	.073	.865	.075	.803	.115	.834	.095	.835	.084
Highly probable	.873	.081	.890	.059	.882	.070	.888	.063	.867	.082	.878	.073	.880	.071
Almost certain	.930	.074	.931	.058	.931	.066	.922	.067	.870	.149	.896	.108	.913	.087
Absolutely cert.	.999	.007	1.000	.000	1.000	.004	.998	.010	.996	.013	.997	.012	.998	.008
Mean	.504	.071	.506	.047	.505	.059	.515	.084	.496	.118	.505	.101	.505	.080
Standard dev.	.301	.026	.309	.019	.305	.018	.310	.054	.281	.060	.295	.052	.299	.031

3.2. Effect of list order on values assigned to the verbal expressions of probability.

The second procedure of the proposed method, which is the subject's individual assignment of numerical values to verbal expressions of probability, is important because it potentially increases the accuracy of the proposed method by allowing adjustments for (a) individual differences in the interpretation of phrases, and (b) phrase interpretation differences due to context. In order to evaluate the reliability of the values elicited in this procedure, it is necessary to determine whether the order in which phrases are presented affects the numerical values that subjects assign to the phrases. Table 14 shows the mean value subjects assigned to the phrases when they were presented in each of the 4 orders. It also shows aggregate means for ordered lists, random lists, and all lists. These values may be compared with those in Table 1 and Table 8.

****

Insert Table 14 about here.

****

3.2.1. Effect of list structure on accuracy and variability of assigned values.

Accuracy of assigned values may be measured by subtracting the values the researcher assigned to the phrases *a priori* (based on previous studies; see Table 1) from the values the subjects assigned to them. Table 15 shows the mean accuracy (deviation) scores for the ordered and random phrase lists, and their variability (standard deviations). In both lists the deviations tend to be positive in the first half of the list, and negative in the second half. That is, subjects' numbers were too high when the *a priori* value was low, and too low when the *a priori* value was high. Thus these subjects have shifted toward .5 in 1987, in comparison with the interpretations of these phrases found in previous studies (primarily Lichtenstein and Newman, 1967).

****

Insert Table 15 about here.

****

The hypothesis that an ordered presentation of the verbal expressions of probability allows someone to more readily recognize their meanings predicts that subjects will assign more accurate numerical values (closer to the *a priori* values) when the lists are presented in ascending or descending order than random order. For 13 of the 19 phrases, the absolute value of the mean deviation was larger when the lists were presented randomly. Four of these comparisons (for "rarely", "very unlikely", "seldom", and "almost certain") were statistically significant (in one-way ANOVAs), and two more ("somewhat unlikely" and "slightly less than half the time") were at $p < .10$. Only one of the phrases ("not very probable") had a significantly greater absolute deviation in the ordered list than in the random list.

The hypothesis also predicts that subjects will be less variable in assigning numerical meanings to phrases in the ordered lists than in the random lists. In contrast with the previous analysis, this prediction does not depend on *a priori* assumptions about the "true" meanings of the phrases. Column 8 of Table 15 shows that for 17 of the 19 phrases (all save "tossup" and "high probability"), there was higher variability in the numerical values assigned to the phrases when they were presented in the randomly ordered lists ($Chi^2 = 7.5$, $df = 1$, $p < .005$, one-tailed). The mean standard deviations of the values assigned to the phrases in the 4 lists are shown at the bottom of Table 14. The values assigned to phrases in the list with Random Order B had the highest standard deviation (.118), more than that for Random Order A (.084), Ascending order (.071), or Descending order (.047). The mean random list standard deviation ($M = .101$) was significantly higher than the mean ordered list standard deviation ($M = .054$; $t = 3.92$, $df = 18$, $p < .01$). T-tests between the standard deviations of the values assigned in the individual lists are shown in Table 16. All 4 comparisons between random and ordered lists have the predicted order (3 of them significant). Overall there was, as predicted, less between-subject variation in the means of the numerical values assigned to the phrases when they were presented in an ordered list. In addition, there were significant differences between the two ordered lists (subjects assigned more varying values in the ascending list) and between the two random lists (subjects assigned more varying values in list B).

****

Insert Table 16 about here.

If subjects assign to randomly arranged verbal expressions of probability numerical values that are more variable and farther from the conventional meanings of these terms than they assign ordered lists, and if this reflects their understanding of the meanings of the phrases when they are using them to answer the questions, then it is preferable to use the ordered lists.

3.2.2. Effect of breadth of phrase meaning and list structure on variability and accuracy of value assignment.

It can be expected that subjects will assign more variable numerical values to verbal probability expressions that have broader meanings. This would occur particularly when the phrase lists are randomly ordered, for the context supplies fewer constraints on the meaning of each term. The correlations between the four measures of breadth of membership function of the 19 phrases (from Table 9) and the standard deviations of those phrases when presented in each list order (from Table 14) are shown in the top half of Table 17. (It should be noted that two of the measures of breadth of membership function are in fact standard deviations, and so would have high correlations by definition.) There is a strong positive correlation between the indices of breadth of phrase meaning (Mean and Median Range) and the standard deviations of the values assigned to the phrases, for every list order except the ascending list. This relation was expected and is the reason the standard deviation was proposed as a proxy measure for the breadth of membership function. However, there is a difference between ordered and random lists in the strength of this relationship.

Insert Table 17 about here.

Analogous arguments lead us to expect that the breadth of a phrase's membership function may influence the accuracy of the values subjects assign to the phrase, and that this effect may be moderated by whether the list is ordered or random. The correlations between the measures of breadth of meaning and the accuracy scores (absolute deviations, defined above), presented in the bottom half of Table 12, show that there is no significant relation between accuracy of the value assignment and Mean Range, our best measure of breadth of membership function, although Median Range and standard deviation are significantly positively correlated with accuracy. These latter relations may be attributed to the fact that these two measures distinguish the phrases identified with 0, .5, and 1 from the others, and people know the value of these probability phrases. The structure of the list (ordered versus random) has no effect on the size of these relations.

3.2.3. Effect of nearness of anchor on variability and accuracy of value assignment.

Three of the verbal expressions of probability used here have quite specific meanings: "absolutely certain" (1.0), "tossup" (.50), and "absolutely impossible" (0). It is possible that subjects use these phrases as anchors when assigning values to other phrases. If so, we may expect less variability in the values assigned to phrases that are near to these anchors, than in the values assigned to more distant phrases. The distance of a phrase from an anchor will be more salient in an ordered list than in a random list. Therefore, we may expect the effect of distance from an anchor phrase on the variability of the values assigned to other phrases to be smaller in random lists. However, people already know the meanings of these phrases, and so even in the random lists a phrase whose meaning is near an anchor may have a narrower range of interpretations.

In the context of our list of 19 phrases, the distance of a phrase from the nearest anchor is simply measured by counting the number of steps in the list to the nearest anchor (see Column 9 of Table 15). The hypothesis predicts that this measure will be positively correlated (over the 16 non-anchor phrases) with the standard deviation of the values the subjects assigned to the phrase, and that this correlation will be larger for the ordered lists than for the random lists. Column 1 of Table 18 shows these correlations for each list and for the combined lists (ordered, random, and total). The correlations of phrase value standard deviations with distances from anchors are significantly

Table 15

Means and standard deviations of accuracy scores (deviations) for values assigned to phrases, Ordered and Random lists.

	Ordered Lists		Random Lists							
	Mean	SD	Mean	SD	List with greater deviation score	Test of Mean dif F	sig	List with greater variability	Distance of phrase from nearest anchor
Absolutely impossible	.013	.103	.018	.114	r	.06	.800	r	0
Rarely	.047	.081	.094	.089	r*	10.54	.002	r	1
Very unlikely	.031	.049	.085	.221	r*	4.32	.040	r	2
Seldom	.026	.054	.065	.124	r*	6.07	.015	r	3
Not very probable	.031	.080	-.002	.110	o*	4.10	.045	r	4
Fairly unlikely	.032	.063	.010	.108	o	2.14	.146	r	4
Somewhat unlikely	.004	.067	-.028	.124	r+	3.64	.058	r	3
Uncertain	.005	.067	.002	.179	o	.02	.887	r	2
Slightly less than half the time	-.003	.042	-.018	.054	r+	3.54	.062	r	1
Toss-up	-.005	.039	.000	.003	o	1.19	.277	o	0
Slightly more than half the time	.002	.053	.016	.072	r	1.85	.176	r	1
Better than even	.002	.066	.018	.082	r	1.55	.215	r	2
Rather likely	-.026	.076	-.020	.180	o	.05	.819	r	3
Good chance	-.025	.077	-.032	.148	r	.12	.734	r	4
Quite likely	-.022	.076	-.043	.108	r	1.79	.184	r	4
Very probable	-.017	.077	-.016	.101	o	.00	.947	r	3
Highly probable	-.021	.074	-.022	.073	r	.02	.886	o	2
Almost certain	-.020	.068	-.053	.117	r*	4.45	.037	r	1
Absolutely certain	-.001	.006	-.003	.011	r	2.68	.104	r	0
Number of phrases for which random list has larger statistic:					13			17	
Number of phrases for which ordered list has larger statistic:					6			2	

Note: N = 76 for Ordered lists, and N = 64 for Random lists.

$^{+}p < .10$; $^{*}p < .05$.

Table 16

T-tests of differences between mean standard deviations of values assigned to phrases, in different phrase list orders.

		SD Asc	SD Des	SD Rana
	Mean	.071	.047	.084
SD Des	.047	-3.54*	-	-
SD Rana	.084	1.02	3.37*	-
SD Ranb	.118	3.80*	5.36*	3.15*

*$p < .01$.

Table 17

Correlations of measures of phrases' breadth of membership function and measures of phrases' standard deviation and accuracy.

	Measures of Phrases' Breadth of Membership Function			
	Mean Range	Median Range	SD of Values	SD of Midpoints
Measures of Phrases' Variability				
SD ascending list	.22	.28	.51*,a	.26
SD descending list	.72**	.86**	.56**	.48*
SD ordered lists	.56**	.68**	.68**	.46*
SD random list A	.75**	.69**	.86**	.53**
SD random list B	.39*	.56**	.91**	.40*
SD random lists	.61**	.68**	.97**	.50*
SD all lists	.67**	.76**	1.0[b]	.55*
Measures of Phrases' Accuracy				
Acc ascending list	.09	.44*	.45*	.27
Acc descending list	.12	.50*	.49*	.08
Acc ordered lists	.11	.49*	.49*	.20
Acc random list A	.02	.40*	.38+	-.02
Acc random list B	.13	.50*	.54**	.24
Acc random lists	.09	.49*	.51*	.15
Acc all lists	.10	.52**	.55**	.18

[a] These row variables are all components of the column index.

[b] The row variable is identical with the column variable.

+ $p < .10$; * $p < .05$; ** $p < .01$.

Table 18

Correlations between standard deviations and accuracies of values assigned to phrases, and distance of phrase from nearest anchor.

Phrase List Order	Correlation of Phrase Value SD with distance from anchor	Correlation of Phrase Value Accuracy (abs. deviation) with distance from anchor
Ascending	.41+	.30
Descending	.30	.13
Ordered (all)	.44*	.24
Random A	.32	-.40+
Random B	.14	.00
Random (all)	.25	-.16
Total	.34+	-.05

Note: N = 16 for every correlation.

+ $p < .10$; * $p < .05$.

positive, as predicted, for the ascending list, and for the ordered lists overall. Though positive, the correlations for the random lists are smaller, as expected, and nonsignificant.

Insert Table 18 about here.

If subjects do indeed use these three phrases as anchors, does this contribute to the accuracy of the values they assign to other variables? Are the phrases near anchors assigned more accurate values, and if this effect occurs is it stronger in ordered phrase lists? Column 2 of Table 18 shows that there is a nonsignificant correlation of .24 between absolute error of assigned value and distance from anchor, which is the predicted direction. The correlation in one of the random lists was -.40, df=15, $p < .10$, in the opposite direction. These results provide weak evidence that when lists are ordered, subjects use an anchoring strategy that both narrows the range of the values they assign to verbal expressions of probability, and makes those value more accurate.

3.2.4. Effect of list structure on amount of duplication in assigned values.

If in the value assignment procedure of the proposed method, subjects assign the same numerical value to more than one phrase, this would degrade the precision of the method. People can be expected to do this more often when the lists are random than when they are ordered. The extent of such duplication can be measured by counting the number of pairs of phrases to which a subject assigns the same value. For example, if "almost certain" and "highly probable" are both assigned the value .90, that is one duplicated pair. If in addition "quite likely" were to be called .90, this would produce 3 pairs. If someone assigned the same value to all 19 phrases, there would be (19*18)/2 = 171 duplicate pairs. Table 19 shows the number of pairs of phrases that were assigned duplicate values for each list. The number of duplications is very small in comparison with the maximum possible count of 171. Significantly more duplicate values were assigned to phrases in the random lists (M = 4.6) than in the ordered lists (M = 1.5), as predicted ($F(3,140) = 6.13$, $p = .0006$).

Insert Table 19 about here.

3.3. Effect of phrase list order on accuracy of problem answers.

A third criterion for evaluating the proposed method of expressing degree of belief by selecting verbal expressions of probability is the accuracy of its use. This accuracy is a joint product of (a) the phrase the subject selects, (b) the meaning assigned to the phrase, and (c) the right answer to the problem. Hamm (1988) has compared the accuracy of the verbal and numerical response modes in this study, and found that verbal responses were more accurate in some probabilistic inference word problems but less accurate in others. Here we ask whether the accuracy of subjects' responses is affected by the order in which the phrases are presented.

3.3.1. Effects of list structure on accuracy of problem answers.

Accuracy of answers using the response mode of selecting answers from a list of verbal expressions of probability can be measured by translating all phrases (those the experimenter included in the word problem, and those the subject selected as response) into numbers, and comparing the response number with the correct answer (produced by applying Bayes' Theorem to the numbers in the word problem; see Hamm, 1988). Translation from phrases to numbers can be done in two ways: using the *a priori* values (Table 1) or the values each individual subject assigned to the phrases. Accuracy using both translations will be studied here, to separate those effects of list order which are due to selection from those due to value assignment. If phrase list order affects accuracy using the *a priori* translations, this can only be due to its effects on selection of a phrase as a response. If list order affects accuracy using the subjects' individual translations but not using the *a priori* translation, this must be due to its effects on subjects' assignment of values to phrases.

Results are presented separately for subjects for whom the word problems were presented with verbal and numerical expressions of probability (Table 20). If the probabilities were presented as phrases, the numerical value of the right answer depends on an assignment of a numerical value to one or more phrases. Because the subject was in the verbal response mode condition, the numerical value of his or her answer also depends on the assignment of a numerical value to a phrase. Answers when no information had been presented are not analyzed here, because there was little variation in response. There was no single correct answer for the Doctor and Insurance problems when only two pieces of information had been presented (see Hamm, 1987; 1988), and so these subproblems too are excluded from the analysis.

Insert Table 20 about here.

Table 20 shows the accuracy scores (absolute errors) for subjects using the verbal response mode, computed using the *a priori* translations from phrases to numbers, for ordered and random phrase list orders, separately for each subproblem and for the numerical and verbal presentation modes. The ordered list produced more accurate answers on 8 of 20 comparisons between the ordered and random lists. Only three of these 20 comparisons were statistically significant. In all three, ordered lists produced more accurate responses. (When the deviation score, rather than the absolute deviation score, was used, the results were similar, which shows that the advantage of ordered lists is not simply their smaller variability.) In conclusion, there is weak evidence that the order in which phrases are presented influences the accuracy of the subjects' performance on probabilistic inference word problems. A parallel analysis, using subjects' individually assigned values to translate the meaning of the phrases and calculate accuracy, had similar results.

3.3.2. Effects of use of subject's own assigned values versus *a priori* values on accuracy of response.

A motivation for the proposed method is to enable subjects to express their degrees of belief in a way that is more natural for them than using numerical probabilities. It might seem that asking the subjects afterwards for their numerical interpretations of the phrases defeats this purpose. However, the virtue of the method is its isolation of the numerical thinking, for it allows subjects to use only the linguistic mode when thinking about the problems. Translating the phrases into numbers is done separately and does not interfere with the all-important problem solving. Nonetheless, the assignment of numbers to phrases places a burden on the subjects, and so it is worth considering whether it is possible to do without this part of the procedure by using a priori numerical interpretations of the phrases. What effect does the use of the subjects' own translations of the phrases have on their accuracy on the word problems?

Table 21 shows the mean accuracy score (absolute error) on each problem using both the *a priori* numerical values and the subjects' own values for the phrases. The comparison includes subjects whose presentation mode and response mode were numerical/verbal, verbal/numerical, or verbal/verbal. The answers using the subjects' own values were more accurate on 8 of the 10 problems (using both absolute deviation scores (Table 21) and simple deviation scores), and significantly so after three pieces of information for the Doctor and Twins problems. However, there is significantly higher accuracy using the *a priori* values after two pieces of information for the Cab problem. Therefore when accuracy is very important, it is probably preferable to use subjects' individual interpretations of the phrases, rather than relying on a universal *a priori* interpretation. However, the evidence is mixed, and the difference in even the significant comparisons is small. In conditions where it is difficult to get subjects to assign values to the phrases, *a priori* interpretations could be used with only a small probable loss of accuracy.

Insert Table 21 about here.

Table 19

Mean number of pairs of phrases to which subjects gave duplicate values, for each phrase list order.

Phrase List Order	Mean # of pairs	SD	N of subjects
Ascending	1.73	5.41	49
Descending	1.03	2.08	30
(Ordered)	1.47	4.44	79
Random A	4.36	4.17	33
Random B	4.84	4.66	32
(Random)	4.60	4.39	65

Table 20

Comparison of word problem accuracy (absolute deviations) between subjects with ordered and random phrase lists, for subjects with numerical and verbal presentation of probabilities.

Problem	Amount of info		Presentation mode: Numerical / Response mode: Verbal — Mean dev Ord	Mean dev Ran	F	sig	Presentation mode: Verbal / Response mode: Verbal — Mean dev Ord	Mean dev Ran	F	sig
Cab	1		.13	.24	2.4	.14	.13	.21	2.1	.17
		N	(9)	(8)			(10)	(8)		
Cab	2		.04	.13	19.0	.00**	.09	.06	1.2	.30
		N	(9)	(8)			(11)	(8)		
Cab	3		.35	.33	.4	.51	.42	.38	.9	.35
		N	(18)	(16)			(21)	(16)		
Doc	1		.06	.20	5.2	.03*	.17	.11	1.3	.26
		N	(18)	(15)			(21)	(16)		
Doc	3		.63	.54	1.3	.27	.62	.47	2.6	.13
		N	(9)	(8)			(10)	(8)		
Ins	1		.23	.25	.1	.74	.15	.16	.0	.87
		N	(18)	(16)			(21)	(16)		
Ins	3		.45	.44	.0	.99	.23	.47	7.7	.01*
		N	(9)	(8)			(11)	(8)		
Twn	1		.08	.13	1.2	.23	.22	.12	1.0	.34
		N	(9)	(7)			(11)	(8)		
Twn	2		.09	.09	.0	.91	.12	.11	.0	.97
		N	(9)	(8)			(10)	(8)		
Twn	3		.25	.22	.2	.64	.30	.26	.5	.49
		N	(18)	(16)			(21)	(15)		

Number of problems (out of 10) where ordered list is more accurate	5	3

$^{*}p < .05$; $^{**}p < .01$.

Table 21

Comparison of Mean accuracies (absolute deviations) for each subproblem, using *a priori* values *versus* subjects' individual values.

Problem	Amount of info	*A priori* values	Own values	t	sig	N
Cab	1	.19	.19	.09	.925	53
Cab	2	.10	.13	-2.67	.011*	53
Cab	3	.38	.36	1.65	.101	105
Doc	1	.17	.17	.18	.857	105
Doc	3	.61	.55	4.06	.000**	53
Ins	1	.20	.19	.69	.492	107
Ins	3	.39	.40	-.92	.361	53
Twin	1	.17	.17	.79	.434	52
Twin	2	.11	.12	.98	.333	52
Twin	3	.28	.24	2.93	.004**	104

* $p < .05$; ** $p < .01$.

4. Discussion.

Expressing uncertainty by selecting verbal probabilities from a list has the advantages detailed by Zwick (1987; e.g., that people prefer to use verbal probabilities) without the disadvantages of unconstrained verbal expression. Because there are only a limited number of phrases in the offered list, it is possible to agree on their meanings, and so communication of uncertainty is feasible with this method. Because the individual gives numerical interpretations for the verbal expressions in the value assignment procedure, it is possible to compensate for individual differences in the meanings of terms.

The present study tested whether the arbitrary features of the method, specifically the sequential order of the list of verbal expressions, and the positions of particular phrases in the list, affect the results. Investigation of the influence of list order on the selection of phrases, the assignment of numbers to represent phrase meanings, and the accuracy of the responses produced using the method showed that sequentially ordered lists are less vulnerable than random lists to ordinal position effects and to the effects of variations in the phrases' breadth of meaning.

The effect of ordinal position on the selection of a phrase was ascertained by comparing the phrases selected from reversed lists. List order reversal made little difference. If there were no ordinal position effect, the mean ordinal position of the selected phrases, averaged across reversed lists, would be the 10th position out of 19. The mean selected position for the final answers on the problems was 9.66 for the ordered lists, and 10.84 for the random lists. For the ordered lists, this is not statistically different from the 10th position. For the random lists, the tendency to pick terms in the second half of the list is significant only when the effect is measured over all four problems.

People seem to prefer phrases with relatively broad meanings, such as "somewhat unlikely" or "good chance". This preference, however, may be due to the particular word problems used in the study. The answers to these problems tended to be in the .60 to .90 range (see Hamm, 1988). The verbal expressions covering this range (as well as the .10 to the .40 range) have broader ranges than the phrases covering other ranges. Therefore subjects are likely to use a phrase with a relatively broad meaning on these problems. An additional effect that is independent of problem content was demonstrated: in random lists, the preference for broad over narrow phrases was greater in the second half of the list. There were two performance measures on which random lists were not significantly different than ordered lists -- the time to complete the questionnaire, and the variance of the *a priori* meaning of the phrases selected as word problem answers.

The method's ability to compensate for individual differences and context effects in the interpretation of the verbal expressions of probability depends on the second step, a separate procedure in which subjects assign numerical values to the phrases. Subjects gave more variable and less accurate values to phrases which were displayed in random order. Similar effects probably occur when people interpret the verbal expressions prior to selecting a phrase to answer a word problem.

The data suggest that subjects produce values for the terms by anchoring on the meanings of known phrases ("absolutely certain" for 1.0, "tossup" for .5, and "absolutely impossible" for 0), and then adjusting. The evidence for this strategy is that the values given to phrases near these anchors were less variable than the values given to phrases farther away. The correlation between phrase value variance and distance from an anchor was statistically significant in the ordered lists, but not significant (though positive) in the random lists. Thus the sequential arrangement of the list seems to facilitate the use of the anchor and adjust strategy in assigning values to phrases. This may be why more accurate values were assigned to phrases in the ordered lists, and is another reason to prefer ordered lists. Additionally, the ordered list promotes more discrimination among the phrases, for fewer duplicate values were assigned to phrases when they were presented in sequence.

The accuracy of the word problem answers depends on the accuracy of the two procedures we have already discussed, selection and value assignment. Although the advantages of ordered phrase lists have been demonstrated for both of these procedures, measurement of their effect on

word problem accuracy gives perspective on the importance of the distinction between ordered and random lists. The overall difference was very small, but the ordered lists produced significantly more accurate responses than the random lists for 3 of the 20 answers tested.

All these comparisons indicate either that the phrase selection method is better using ordered phrase lists than random lists, or that there is no difference. Before recommending ordered phrase lists, however, we must consider a potential criticism. The constraints that an ordered list places on the subject's interpretation of the phrases may distort, rather than clarify, the meanings of the phrases, thus preventing people from using the phrases as they normally would. Consider, for example, the meaning of "almost certain". Kong, Barnett, Mosteller, and Youtz (1986) found subjects assign it a mean value of .78 (median .90), but the author used it to mean .95 in the present study. When subjects assigned values to "almost certain" in the random phrase lists (where there was nothing to indicate that the phrase meant .95), the mean value was .90 (see Table 9). However, in ordered lists (where it appeared in the 18th or 2nd of 19 positions, between "highly probably" and "absolutely certain"), its mean value was .93. This proves that placing a phrase in an ordered list may change its meaning.

Another example is the verbal expression for .40, "uncertain." The range of meaning people assign to this phrase is both very wide (an average of .24 between the lower and upper bounds; see Table 9) and very variable (Table 8 shows an average standard deviation of .34; some subjects gave it a range of 0 and others a range of 100). Although the mean value assigned to "uncertain" was .40 or .41 in both the ordered and random lists (Table 14), agreeing with the *a priori* value, these values were much more variable in the random list (sd = .18) than the ordered list (sd = .06). Thus placing a phrase in an ordered list can change the breadth of its meaning. Because of the exceptional variability of the meaning of "uncertain", an alternative phrase for .40 should be substituted. A candidate is "worse than even", which Shanteau (1974) found to have a mean value of .38 using two different procedures, and which is symmetric with "better than even" whose meaning is .60.

Although it is possible to find replacement phrases for particular inappropriate verbal expressions, still the ordered list will change some phrases' meanings and breadths of meaning, for many individuals. This can be viewed, however, as a necessary cost of adopting a common set of interpretations of verbal expressions of uncertainty. Kong, Barnett, Mosteller, and Youtz (1986) advocate improving the use of verbal probabilities through codifying the meaning of probabilistic expressions. They suggest measuring what people usually mean by phrases, publicizing this, and training people to use the terms with these agreed-upon meanings. Such publicity and training would (a) reduce the differences between people, (b) narrow the individual membership functions for each phrase, and (c) get people to use the phrases to mean the same probability in different contexts. Such a program would require changing people's interpretations of many phrases, in the process of establishing a new convention. The proposed method of selecting verbal probability expressions from a list could be a tool in such a program. The changes that the use of an ordered phrase list in this method would induce in the meaning of its phrases are costs worth incurring in order to improve communication about uncertainty.

Beyth-Marom (1982) proposed an alternative framework for codifying the meaning of verbal probability expressions. It divides the probability scale into ranges .10 or .20 wide, and associates each range with from 2 to 6 verbal expressions. For example, the terms "small chance" and "doubtful" would refer to the .10 to .30 range. Although this reflects the fact that verbal expressions apply to ranges of probability, it has disadvantages. It does not distinguish between probabilities within a range. It requires people to learn a number of sharp boundaries (e.g., at .10 and .30) that are somewhat arbitrary. If establishing a convention requires people to relearn the meanings of phrases, it seems more useful to associate phrases with points and allow for fuzzy boundaries, than to associate phrases with specific ranges.

The method proposed here optionally elicits subjects' own numerical meaning for each phrase. Use of subject supplied values rather than *a priori* values to interpret the phrases used in the word problems, in order to evaluate the accuracy of the subjects' reasoning, resulted in improved

Probability Response Scale

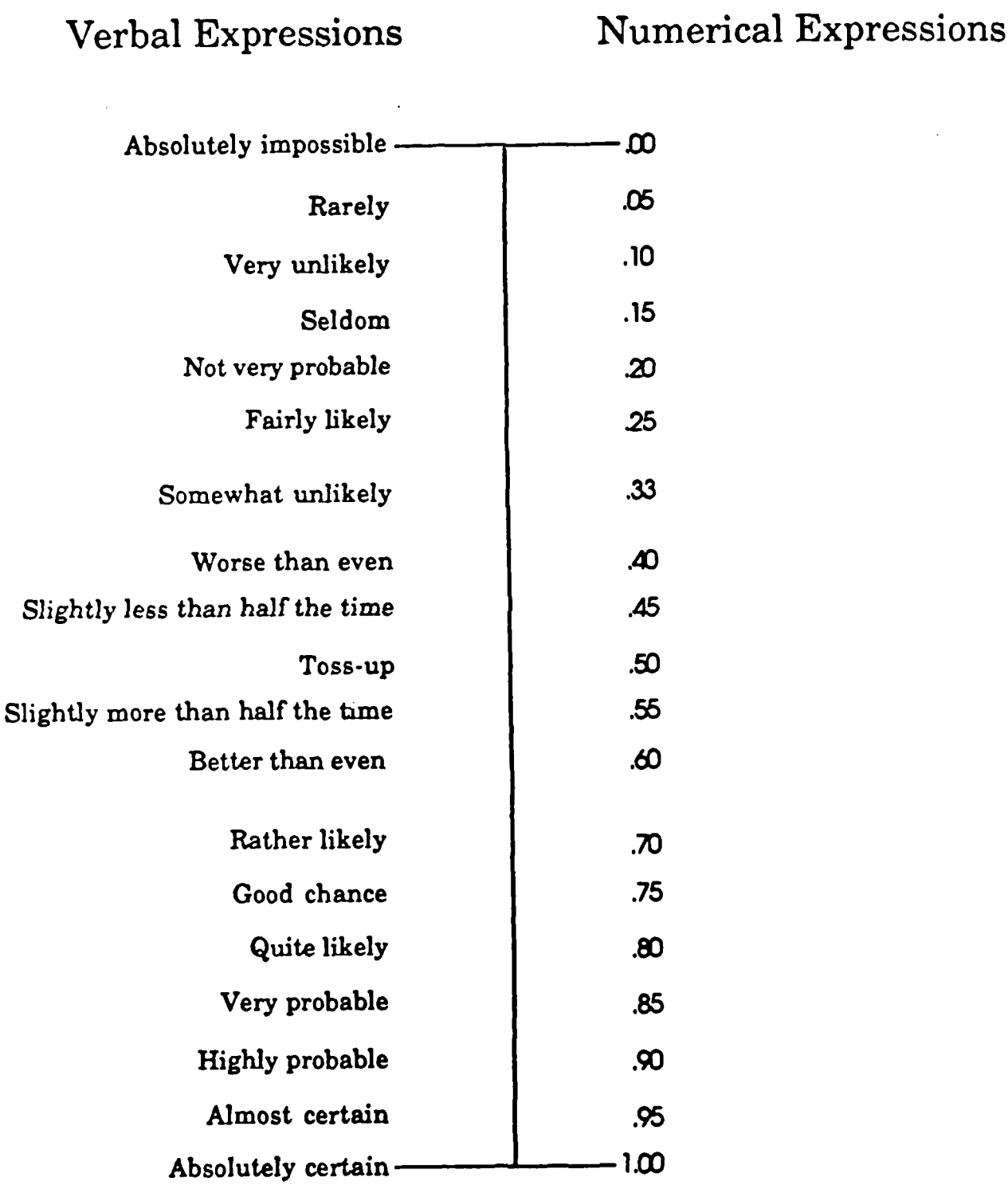


Figure 1. Probability Response Scale

accuracy in 8 of 10 problems, but this cost additional subject time. The need for such a procedure would presumably fade if a set of conventional meanings would become accepted.

A list that displayed both the verbal expressions of probability and their numerical interpretations (as in Figure 1) could be useful in this context. People would be free to use the mode they found more fitting to the problem and to their cognitive style. The two modes of expression would mutually define each other, so that people's interpretation of each would be more constrained. Finally, use of the scale would train people to associate the verbal and numerical expressions, promoting the acceptance and use of the new convention.

****

Insert Figure 1 about here.

****

Alternative lists of verbal expressions of very low or very high probabilities would be useful for making distinctions among degrees of near impossibility or near certainty. These lists should be based on research discovering the phrases people already use in contexts where these ranges of probability are pertinent, such as medicine (cf. Meyer and Pauker, 1987) or technological systems. A recent example highlights this need. To assess the overall risk of space shuttle failure, NASA engineers were asked to make verbal assessments of the reliability of space shuttle components. These were then translated into numbers, using an arbitrary code ("frequent" = .01; "reasonably probable" = .001; "occasional" = .0001; and "remote" = .00001) that was not used by the engineers in making their original assessments (Marshall, 1986). This poor risk assessment practice has given subjective judgment a bad name in the aerospace community: "the government is relying too much on subjective judgment and too little on statistical analysis in deciding which of thousands of safety problems on the space shuttle should get attention" (Marshall, 1988, p 1233). Codification of verbal expressions of probability would impose consistent interpretations on the phrases and allow experts' subjective judgment to makes its potentially crucial contribution.

5. Bibliography.

Beyth-Marom, R. (1982). How probable is probable? A numerical translation of verbal probability expressions. Journal of Forecasting, 1, 257-269.

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

Hamm, R.M. (1987). Diagnostic inference: People's use of information in incomplete Bayesian word problems. (Publication #87-11.) Institute of Cognitive Science, University of Colorado, Boulder.

Hamm, R.M. (1988). Accuracy of probabilistic inference using verbal and numerical probabilities. Institute of Cognitive Science, University of Colorado, Boulder.

Kong, A., Barnett, G.O., Mosteller, F., and Youtz, C. (1986). How medical professionals evaluate expressions of probability. New England Journal of Medicine, 315, 740-745.

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

Mapes, R.E.A. (1979). Verbal and numerical estimates of probability in therapeutic contexts. Social Science and Medicine, 13A, 277-282.

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

Marshall, E. (1988). Academy panel faults NASA's safety analysis. Science, 239, 1233.

Meyer, K.B., and Pauker, S.G. (1987). Screening for HIV: Can we afford the false positive rate? New England Journal of Medicine, 317, 238-241.

Shanteau, J. (1974). Component processes in risky decision making. J. Experimental Psychology, 103, 680-691.

Simpson, R.H. (1944). The specific meanings of certain terms indicating differing degrees of frequency. Quarterly J. of Speech, 30, 328-330.

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

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

Zimmer, A.C. (1983). Verbal *vs.* numerical processing of subjective probabilities. In R. W. Scholz (Ed.), Decision Making under Uncertainty. North-Holland: Elsevier Science Publishers, pp 159-182.

Zwick, R.. (1987). Combining stochastic uncertainty and linguistic inexactness: Theory and experimental evaluation. Ph.D. Dissertation, Psychology Department, University of North Carolina, Chapel Hill.

6. Appendix 1.

The "Doctor" problem, one of four probabilistic inference word problems used in the study.

In alternative versions of the problem, the probabilistic information was presented in either **verbal** or *numerical* form.

****

[0 pieces of key information.] The next word problem is about a doctor trying to figure out what disease a patient has. The patient is, undeniably, ill, but it is difficult to know what disease he has. You will be asked to estimate how likely it is that the patient has one of two diseases.

The patient comes in to the emergency room at night with a very unusual symptom - his eyes are bright yellow. The doctor knows that there are only two diseases that can produce this particular symptom - hepatitis and toxic uremia. People never contract both illnesses at the same time.

With what you know now, what is the probability that the patient has toxic uremia?

_____

[1 piece of key information.] A discussion with a colleague reminds the doctor that toxic uremia is a less common disease than hepatitis. He checks a textbook and finds that **[it is highly probable that people]** [*90% of people*] who present to their doctors with the symptom of yellow eyes *have* hepatitis, therefore, **[it is very unlikely that they]** [*only 10% of people with this symptom*] have toxic uremia.

With what you now know, what is the probability that the patient has toxic uremia?

_____

[2 pieces of key information.] The doctor orders the lab to do a Spock test on the patient's blood. In two hours the results are back - the Spock test indicates that the patient has toxic uremia.

With what you know now, what is the probability that the patient has toxic uremia?

_____

[3 pieces of key information.] The doctor consults his diagnostic manual and discovers that the Spock test is the best way to find out whether a patient with yellow eyes has hepatitis or toxic uremia. However, the Spock test is not foolproof. When the patient has toxic uremia, **[it is rather likely that the Spock test will indicate that the patient has this illness. It is somewhat unlikely that the Spock test will indicate that the patient has hepatitis]** [*the Spock test correctly indicates this 70% of the time, but 30% of the time it falsely indicates that the patient has hepatitis*]. Similarly, when the patient actually has hepatitis, **[it is somewhat unlikely that the Spock test will indicate that the patient has toxic uremia]** [*the Spock test will indicate that the disease is toxic uremia approximately 30% of the time*].

With what you know now, what is the probability that the patient has toxic uremia?

_____

7. Appendix 2.

Instructions for questionnaire eliciting lower and upper bounds on the numerical meanings of each phrase.

[Two versions were prepared. One asked for upper bounds first, and the other asked for lower bounds first.]

People often use words or phrases such as "impossible" or "very likely" to express a degree of uncertainty or certainty. We are interested in the range of uncertainties for which you think it appropriate to use each of a number of words or phrases.

Think of a cafeteria tray that has 100 ping pong balls on it. Some of them are white and the rest are yellow. You can see every one of them clearly. You must convey to a friend how many of the balls are white. You want to tell him how likely it is that a white ball would be picked if they were thoroughly mixed up and someone were to draw one without looking. However, you are not allowed to tell the person the actual proportion of white ping pong balls. Rather, you are forced to use a non-numerical descriptive phrase.

We want to know the range of proportions of white ping pong balls, in the tray described above, for which you would consider each term to be appropriate. We will ask you to tell us this for each of 20 terms.

> The first term is **"about even"**. What is the highest[lowest] proportion of white balls (out of 100) for which you think it would be appropriate to use the term "about even", in trying to tell your friend the proportion of white and yellow ping pong balls? Write that number here:
>
> _____
>
> Now what is the lowest [highest] proportion of white balls for which you think it would be appropriate to use the term "about even"?
>
> _____

Look at your answers. You should have named two numbers somewhere between 0 and 100 (inclusive). The second number should have been lower [higher] than (or equal to) the first. Any number in between the two numbers would be a reasonable interpretation for your friend to make when you tell him that the chance of drawing a white ping pong ball is "about even". Any number higher [lower] than your first answer would not be a reasonable interpretation of "about even"; nor would any number lower [higher] than your second answer be reasonable. If these statements are not all true, you may wish to go back and change one or both of your answers.

On the next page is a list of words or phrases expressing degree of uncertainty. Assume that you are using each phrase to describe the chance of drawing a white ping pong ball from the tray of 100 balls. For each phrase, please express the upper and lower [lower and upper] numerical limits that you would expect your friend to use in interpreting it.

Please focus on each word or phrase by itself, rather than trying to compare it with your answers for other words or phrases.

	Upper [Lower] Limit	Lower [Upper] Limit
Uncertain	_____	_____
Rather likely	_____	_____
Somewhat unlikely	_____	_____
Rarely	_____	_____
Slightly less than half the time	_____	_____
Good chance	_____	_____
Fairly unlikely	_____	_____
Absolutely impossible	_____	_____
Toss-up	_____	_____
Quite likely	_____	_____
Not very probable	_____	_____
Absolutely certain	_____	_____
Slightly more than half the time	_____	_____
Very probable	_____	_____
Seldom	_____	_____
Almost certain	_____	_____
Better than even	_____	_____
Highly probable	_____	_____
Very unlikely	_____	_____

Notes

[1]The applicable t-test is:

$$t = \frac{\bar{X} - \mu_o}{\frac{s}{\sqrt{N}}} = \frac{\bar{X} - \mu_o}{\frac{sd\sqrt{\frac{N}{N-1}}}{\sqrt{N}}} = \frac{\bar{X} - 0}{\frac{sd\sqrt{\frac{18}{17}}}{\sqrt{18}}} = \frac{\bar{X} \times 4.0069}{sd}$$

where "s" is the unbiased estimate of the standard deviation and "sd" is the measured standard deviation.