AD-A048 600 DECISIONS AND DESIGNS INC MCLEAN VA F/G 5/10

NEW AND OLD BIASES IN SUBJECTIVE PROBABILITY DISTRIBUTIONS: DO --ETC(U)

AUG 77 T FUJII, D A SEAVER, W EDWARDS N00014-76-C-0074

UNCLASSIFIED NL

1 OF 1

ADA048 600

END

DATE FILMED

2-78

DDC

001855-4-T

social science research institute

UNIVERSITY OF SOUTHERN CALIFORNIA

USC

RESEARCH REPORT

New and Old Biases in Subjective Probability Distributions: Do They Exist and Are They Affected by Elicitation Procedures?

Tsuneko Fujii
David A. Seaver
Ward Edwards

Sponsored by:
Advanced Research Projects Agency
Department of Defense

Monitored by:
Engineering Psychology Programs
Office of Naval Research
Contract No. N00014-76-C-0074, ARPA

August 1977

SSRI Research Report 77-4

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied of the Advanced Research Projects Agency of the U.S. Government.

Social Science Research Institute
University of Southern California
Los Angeles, California 90007
213-741-6955

The Social Science Research Institute of the University of Southern California was founded on July 1, 1972 to permit USC scientists to bring their scientific and technological skills to bear on social and public policy problems. Its staff members include faculty and graduate students from many of the Departments and Schools of the University.

SSRI's research activities, supported in part from University funds and in part by various sponsors, range from extremely basic to relatively applied. Most SSRI projects mix both kinds of goals — that is, they contribute to fundamental knowledge in the field of a social problem, and in doing so, help to cope with that problem. Typically, SSRI programs are interdisciplinary, drawing not only on its own staff but on the talents of others within the USC community. Each continuing program is composed of several projects; these change from time to time depending on staff and sponsor interest.

At present, SSRI has six programs:

Program for research on crime control. Typical projects include evaluation of a federal program for decriminalization of juvenile status offenders; and development of an inventory of the contents and quality of the information held by criminal justice agencies in Los Angeles County.

Program for the study of dispute resolution policy. Typical projects include collection and analysis of national statistical data concerning the size, cost, and performance of present dispute resolution systems in six other countries; and detailed study of some 30 alternatives to present U.S. criminal justice procedures.

Program for research on desegregation. The present goal of this program is to study the effects of language, physical attractiveness, and community contact on acceptance of minority children in white schools and on their scholastic performance.

Program for research on decision analysis. Typical projects include study of elicitation methods for continous probability distributions; and development of a multi-attribute utility measurement method for evaluating social programs.

Program for research on rights of the mentally ill. This program is studying procedures used in Los Angeles Courts to determine whether a non-criminal mentally ill person is sufficiently dangerous to others or to himself to justify his involuntary custodial confinement.

Program for data research. Typical projects include development of techniques for estimating small-area population sizes between censuses; and development of crime indicators for use in criminal justice system planning.

SSRI anticipates that new programs will be added and old ones will be redefined from time to time. For further information, publications, and the like, write or phone the Director, Professor Ward Edwards, at the address given above.

Social Science Research Institute
Research Reports

76-1 William E. McGarvey. Can Adjustment Cause Achievement?: A Cross-Lagged Panel Analysis. March, 1976

76-2 Robert M. Carter, Cameron R. Dightman, and Malcolm W. Klein. The System Rate Approach to Description and Evaluation of Criminal Justice Systems. (Reprinted from *Criminology)*.

76-3 Ward Edwards. How to Use Multi-Attribute Utility Measurement for Social Decision-Making. August, 1976.

76-4 David A. Seaver. Assessment of Group Preference and Group Uncertainty for Decision-Making. August, 1976.

76-5 J. Robert Newman, David A. Seaver, and Ward Edwards. Unit Versus Differential Weighting Schemes for Decision Making: A Method of Study and Some Preliminary Results. July, 1976.

76-6 J. Robert Newman. Differential Weighting in Multi-Attribute Utility Measurement: When it should Not and When it Does Make a Difference. August, 1976.

76-7 Ward Edwards and David A. Seaver. Research on the Technology of Inference and Decision. October, 1976.

76-8 Detlof von Winterfeldt. Experimental Tests of Independence Assumptions for Risky Multiattribute Preferences. October, 1976.

77-1 J. Robert Newman. Differential Weighting for Prediction and Decision Making Studies: A Study of Ridge Regression. August, 1977.

77-2 Ward Edwards. Technology for Director Dubious; Evaluation and Decision in Public Contexts. August, 1977.

77-3 Lee C. Eils, III, David A. Seaver, and Ward Edwards. Developing the Technology of Probabilistic Inference: Aggregating by Averaging Reduces Conservatism. August, 1977.

77-4 Tsuneko Fujii, David A. Seaver, and Ward Edwards. New and Old Biases in Subjective Probability Distributions: Do They Exist and Are They Affected by Elicitation Procedures? August, 1977.

NEW AND OLD BIASES IN SUBJECTIVE PROBABILITY DISTRIBUTIONS: DO THEY EXIST AND ARE THEY AFFECTED BY ELICITATION PROCEDURES?

Research Report 77-4

August, 1977

Tsuneko Fujii
David A. Seaver
Ward Edwards

Social Science Research Institute
University of Southern California

Sponsored by

Defense Advanced Research Projects Agency

Summary

Past research indicates that people exhibit biases in assessing probability distributions on continuous variables. Three types of biases have been identified: too many true values falling into the extreme tails of the distributions, a displacement toward 50% for distributions assessed on percentages, and a general tendency to underestimate. This study explored the nature of these biases with particular emphasis on how they interact and how they are affected by the procedure used to elicit the distributions.

Two procedures were used to elicit subjective probability distributions on percentage variables. In a fractile procedure, subjects were asked to judge values of the unknown percentage that corresponded to fixed levels of their cumulative subjective probability distributions, while in an odds procedure, subjects judged the cumulative odds for fixed values of the unknown percentages. For all the unknown percentages, p%, distributions were assessed for both p% and 1-p%. The extent to which these assessments summed to less than 100% indicated a bias toward underestimation.

Underestimation was generally found when the fractile elicitation was used, but not when the odds procedure was used. Also, too many true values fell into the extreme tails of the distributions elicited by the fractile procedure, but no similar bias was found in distributions elicited by the odds procedure. The displacement toward 50% was found in distributions elicited by both procedures. This bias also appeared to be the cause of a considerable number of the true values in the extreme tails of the distributions. Many of the differences in the biases found when different elicitation procedures were used can probably be accounted for by subjects avoiding extreme responses and odds judgments between 1:1 and 2:1.

Contents

Figures

Tables

Acknowledgment

This research was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Office of Naval Research under Contract N00014-76-C-0074 under subcontract from Decisions and Designs, Inc.

Disclaimer

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Advanced Research Projects Agency or of the United States Government.

I. Introduction

Opinions, beliefs, and judgments are most often expressed in non-numerical statements such as "It probably is...", "I am pretty sure...", "He is not likely...", and so forth. We also use the expressions such as "million to one shot" and "fifty-fifty chance" which are semi-numerical expressions not ordinarily used in the literal sense. Beliefs concerning the likelihood of uncertain events are a central basis for making decisions. For good and consistent decisions to be made, these beliefs need an accurate and precise representation which is not provided by such statements. Decision analysis has been developed in the last two decades as a method to assist decision makers in making decisions under uncertain conditions. One of the tools of decision analysis is the use of subjective probability as a numerical expression of uncertainty about relevant variables. This expression of uncertainty provides a precise representation, but the accuracy is open to discussion and can be determined only by empirical testing.

Two types of accuracy are involved in subjective probability statements: the correspondence of the probability statement to the true beliefs of the assessor, and the correspondence of the statement to what actually happens. Murphy and Winkler (1970) discussed two aspects of the latter type of accuracy; primary and secondary validity. "Primary validity refers to the correspondence between the statement and the relevant observation on an <u>individual</u> basis, while secondary validity refers to the correspondence between collections of identical (or similar) statements and the relevant observed relative frequencies on a <u>collective</u> basis" (p. 281). Primary validity is related to the accuracy of the subject's judgment about the occurrence of a particular event. Prediction of rain tomorrow with probability of .85 is more accurate than prediction of rain with probability .55 if rain actually occurs. That is, probabilities should be extreme in favor of the event that actually occurs. Secondary validity is related to the bias of the assessor's statements. It typically is determined by comparing the observed relative frequencies of occurrence with the stated subjective probabilities. For example, an assessor is said to be biased if the relative frequency of events judged to have a probability of .30 of occurring differs substantially for 30%. Secondary validity, also referred to as calibration, realism, and external validity, has been the subject of considerable research (for a review, see Lichtenstein, Fischhoff, and Phillips, in press).

Probabilistic judgments are of two kinds. Assessments may be made for discrete categories, such as "success or failure", "rain or no rain", "win or lose". Or assessments may be made on continuous variables such as "the lowest temperature tomorrow", "the price of oil in 1980", "the number of murders in Los Angeles during 1977". This paper is concerned with biases (calibration) in the latter type of assessments.

Probably the best known and most extensively studied bias of this type is what has been described as the tendency of subjectively assessed distributions to be too tight. That is, a comparatively high percentage of true values fall into the tail areas of the subjective probability distributions which are assigned relatively low probabilities. This bias which seems to indicate the assessed distributions express more certainty than is justified by the knowledge of the assessor was first demonstrated by Alpert and Raiffa (1969). The distributions assessed on a variety of unknown quantities by students in Harvard's MBA program showed that 42.6% of the true values fell into the extreme tails (less than the .01 fractile or more than the .99 fractile) of the subjective distributions where only 2% should have been if the assessors were perfectly calibrated. This finding subsequently has been confirmed by additional experiments (Schaefer and Borcherding, 1973; Seaver, von Winterfeldt, and Edwards, 1975; Selvidge, 1975) although Seaver, et al. found the existence of this bias could be attributed at least in part to the method used for assessing the subjective probability distributions.

These results point rather strongly to a lack of calibration in the tails of assessed subjective probability distributions, but is this lack of calibration also exhibited in the *middle range of the distributions*? The results bearing on this question are less persuasive. Although Alpert and Raiffa (1969) and Shaefer and Borcherding (1973) found the interquartile ranges of the assessed distributions contained substantially fewer true values than the expected 50%; Seaver et al. (1975) and Selvidge (1975) found both too few and too many true values in the interquartile ranges depending on the uncertain quantity and the assessment procedure used. Thus, the extensiveness of this bias deserves further investigation.

A second bias that often appears in subjective probability distributions when the uncertain quantities are percentages is a tendency to overestimate small percentages and underestimate large percentages. In this paper we call this bias "conservatism" referring to the tendency to avoid extremes. Conservatism was originally used to describe the phenomenon typically found in probability revision experiments where after observing a set of data, subjects do not revise their posterior probability as much as the normative model, Bayes' Theorem (Phillips and Edwards, 1966; Wheeler and Edwards, 1975). Thus, the definition of conservatism used in this paper is expanded from the original definition.

This bias has most typically been studied in the context of assessing discrete categories of events (Lichtenstein et al., in press). However, it has also been found in assessing complete distributions on continuous events where the entire distribution is displaced toward 50%. A typical method of showing this bias is to determine the number of true values falling above and below the medians of the subjective probability distributions. For well-calibrated assessors, 50% of the true values should fall above the assessed median and 50% should fall below, regardless of the true value. Conservatism will be exhibited by more than 50% of the true values falling below the medians for true percentages less than 50% and vice versa for true percentages greater than 50%. Selvidge (1975) obtained exactly these results, suggesting that conservatism exists in the assessment of continuous variables as well as discrete variables. This consistent pattern of results, however, is not apparent in the Alpert and Raiffa study. And Schaefer and Borcherding and Seaver et al., although showing some forms of median displacement, do not present the data from individual questions necessary to examine this bias. Consequently, the evidence showing the existence of the conservatism bias is also rather inconclusive suggesting the need for additional research.

The possible existence of an additional bias has been indicated by Seaver et al. They found a general tendency to underestimate unknown percentages with substantially more true values falling above the medians of the assessed distributions than below. Although the data from individual questions are not reported, the results seem striking enough to imply the possible existence of a bias toward underestimation in addition to the conservatism bias. Notice, however, that these two biases will conflict when the true percentage is less than 50%, so some method of separating the influence of these two biases is needed. One possible method is to assess distributions for both p% and 100-p%. Assuming the conservatism bias is equally strong for percentages above and below 50%, the sum of percentages corresponding to fixed points in the two distributions will be less than 100 if this bias exists.

The research reported in this paper examines the extent of each of these three biases, particularly the underestimation bias which has not previously been systematically investigated. In addition, the study by Seaver et al. (1975) suggested that the procedure used to elicit the distributions has an effect on the bias leading to lack of calibration in the tails of the distributions. This study also examines how the elicitation procedure interacts with the displacement biases of conservatism and underestimation in addition to this dispersion bias.

One final possible relationship among these biases is also studied. As pointed out by Schaefer and Bordherding (1973), biases in displacement may lead to results that can be interpreted as a dispersion bias. For example, if the conservatism bias is present in subjects assessing unknown percentages with extreme true values, the high percentages will tend to be underestimated leading to a large number of true values falling in the upper extreme tails of the subjective distributions, and the low percentages will be overestimated leading to a large number of true values falling in the lower extreme tails of the subjective distributions. This will occur regardless of the relative tightness of the assessed distributions and will make the subjective distributions look particularly tight if statistics are summed over unknown percentages with a wide range of true values. To effectively assess the degree of these biases, data from subjective distributions assessed on single unknown percentages or on groups of percentages with similar true percentages must be considered, something that past studies have not done.

Knowledge of the existence of these biases and the degree to which the evidence of their existence depends on the procedure used to elicit the subjective probability distributions is particularly important if such probabilities are to be used for normative decision making. "Good" decisions must be based on "good" information and when that information is in the form of subjective probabilities, the probabilities should accurately reflect the opinions of the assessor. Or, if biases persist, they should be recognized and taken into account.

II. Method

II.1. Subjects. The subjects were 66 undergraduate students who were taking an introductory course in psychology at the University of Southern California. Subjects participated in the experiment on a voluntary basis.

II.2. Questionnaires. Four questionnaires of twenty items each were developed for four groups of subjects. Items in the questionnaires were almanac questions of the type used in the experiments by Alpert and Raiffa (1969) and Seaver, von Winterfeldt and Edwards (1975), with all the uncertain quantities being percentages. Twenty items were selected so that the true percentages would vary over the full range of percentages: each 5% range contained one true percentage. Two items were subsequently eliminated due to some ambiguity in wording. A complete list of the questions and true percentages can be found in the appendix. Each item asking about a true percentage, p, had its counterpart which asked about 1-p. The item asking about p was called the positive item while the item asking about 1-p was called the negative item. Two questionnaires were made for each of two assessment procedures. Each questionnaire consisted of ten positive and ten negative items with one true percentage in each 5% range. The questions were randomized for the questionnaires.

II.3. Elicitation procedures. Two methods for eliciting subjective probability distributions on percentage variables were used. The groups using the fractile elicitation procedure, FRAC, assessed fractiles of the subjective probability distribution at five odds levels, 1:99, 1:3, 1:1, 3:1, and 99:1 for each question. These fractiles were elicited in the following form, "What percentage of the total population of California lived in Los Angeles County according to the 1970 census? Give the percentage such that your odds are 3:1 that the true percentage is less than that number." Subjects simply wrote in the percentages for the five required fractiles.

Subjects in the group using the second elicitation procedure, ODDS, were given five percentages of the variable, asked whether the true percentage was more likely to be above or below each given percentage, and asked to give the odds corresponding to their certainty. The five percentages used were selected separately for each question and included the true percentage and four percentages selected randomly between 1 and 99.

III. Results

For data analyses, all positive items on the two questionnaires used by the FRAC groups were grouped into FRAC+, while all negative items were grouped into FRAC-. Similarly, items on the two ODDS questionnaires were grouped into ODDS+ and ODDS-.

The extent of the underestimation bias as represented by a tendency for the estimates of p and 100-p to sum to less than 100, can be seen pictorially by plotting the cumulative subjective probability distributions for both the positive and negative items. Figure 1 illustrates this with item 17. For the FRAC groups (panel a) each point is the median percentage given for that fractile. For the ODDS group (panel b) each point is the median subjective odds assessed for the specific percentage. The vertical line represents the true percentage. For the sum of p and 100-p to be 100, the plots of distributions of positive and negative items should coincide. Naturally, as with all judgmental processes, some error is expected. However, the extent to which the error is always in the same direction will indicate a bias. In plots of the type illustrated by Figure 1, a bias toward underestimation will be indicated by the cumulative distribution for the positive item always falling to the left of the distribution of the negative item. For example, in panel a, the 1:1 odds level was given a percentage of 51% for the positive item (read from horizontal scale at the bottom of the figure), and 32% for the negative item (read from horizontal scale at the top of the figure). These two percentages sum to only

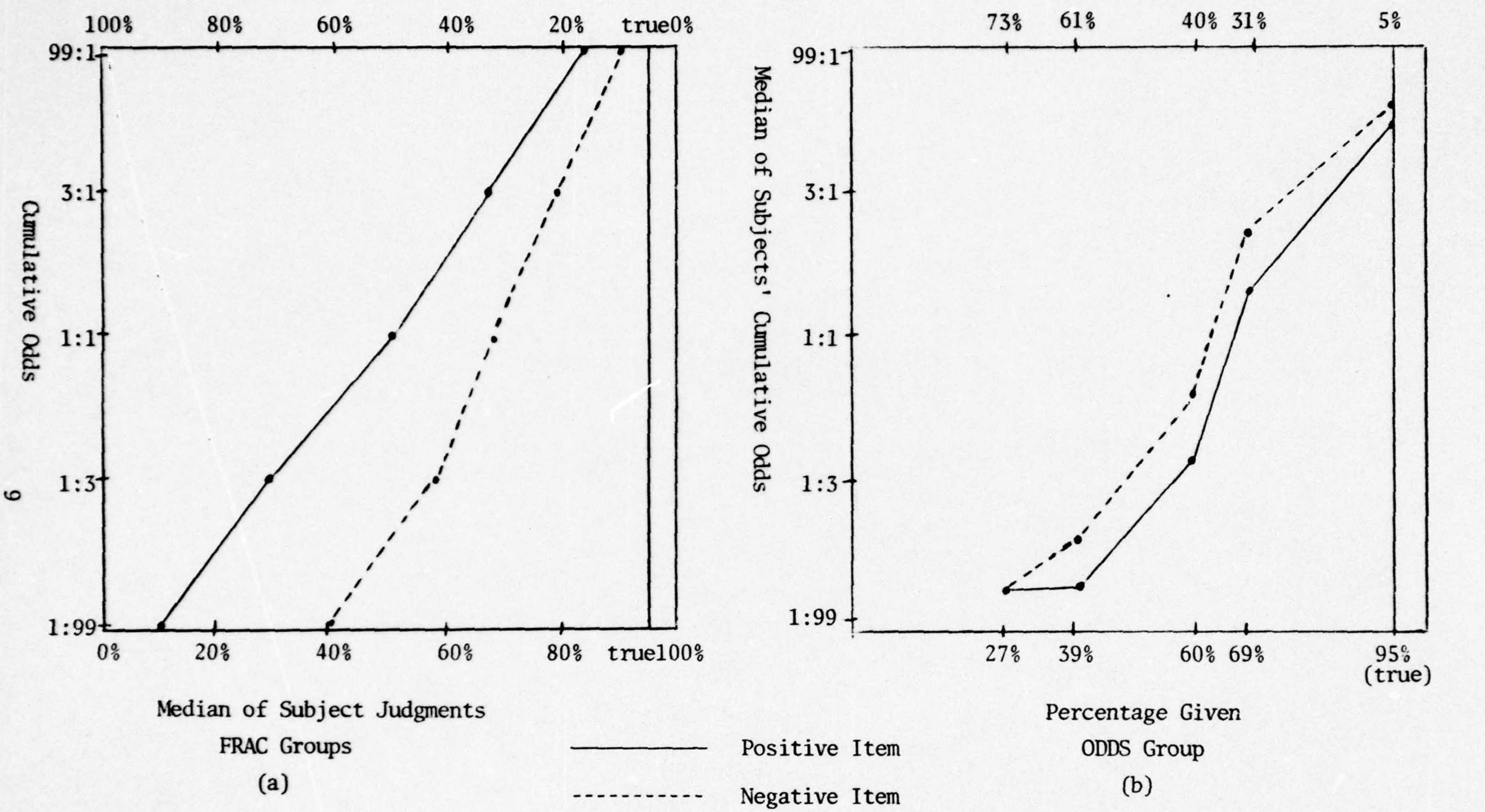

Figure 1: Comparison of Median Responses Between Positive and Negative Items (Item No. 17).

83% indicating underestimation. For this item, in the FRAC groups, the underestimation is consistent across all odds levels. For the ODDS groups, however, there is a tendency toward overestimation on this item. A count of the number of items showing overestimation, underestimation, and no overall bias (the two distributions cross) for the FRAC group showed 0, 14, and 4 items respectively in each category; while the count for the ODDS groups was 3, 5, and 10 items respectively. Thus, underestimation was quite apparent in the FRAC groups but not in the ODDS groups.

Figure 2, in which the medians of the subjective distributions are plotted as a function of the true percentages, also illustrates the underestimation bias in the FRAC groups (panel a). In Figure 2, this bias is shown if the median for the positive item is less than 100% minus the median for the negative item, 15 and 12 items respectively for the FRAC and ODDS groups. The average discrepancy between the positive and negative items is slightly higher for the FRAC groups (8.01%) than for the ODDS groups (4.31%), indicating a greater disposition toward underestimation for the FRAC groups than for the ODDS groups.

The underestimation bias is also exhibited in the assessment of fractiles other than the median. Figure 3 shows the lines regressing median subjective responses on true percentage for all five fractiles. In all cases the responses to the positive items are less than 100% minus the responses to the negative items.

Conservatism is also evident in this data. The medians presented in Figure 2 suggest this tendency, but cannot actually confirm it, since if the individual distributions are skewed, as would be expected for the extreme percentages, the median of the distribution would be above the

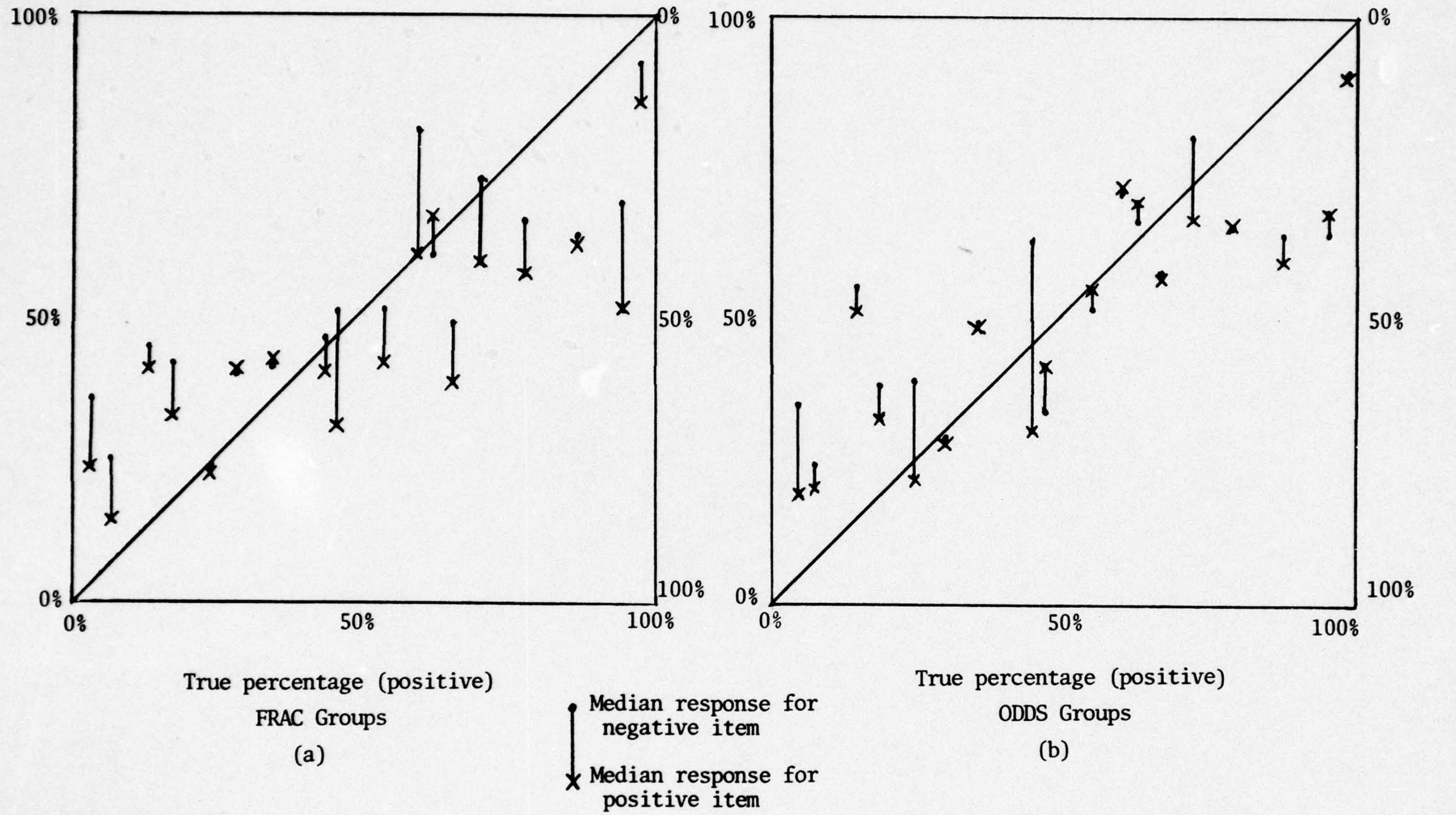

Figure 2: Comparison of Responses to Positive and Negative Items at 1:1 Level.

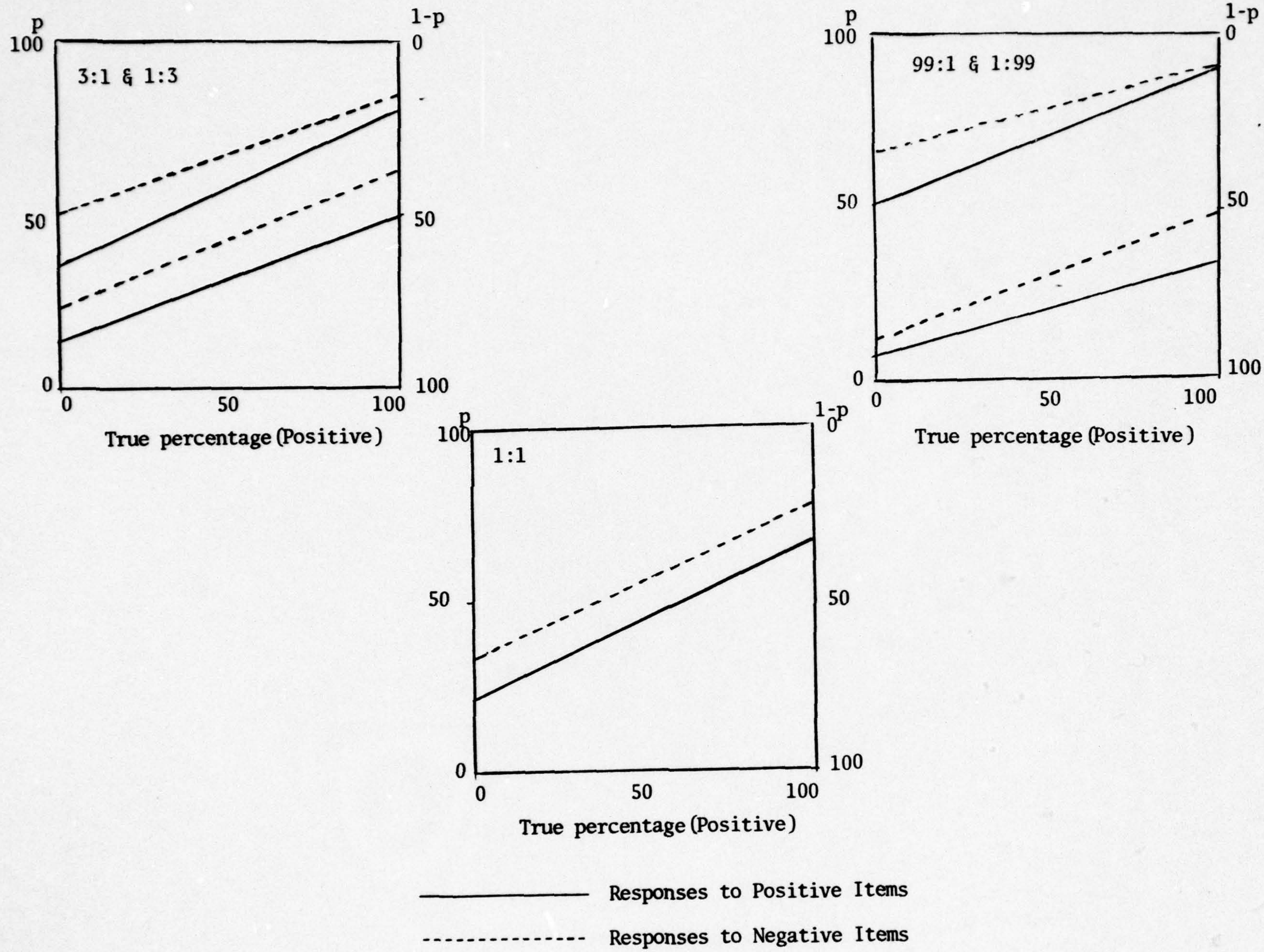

Figure 3: Comparison Regression Lines for Responses to Positive and Negative Items in FRAC Groups

true percentage for low true percentages and below for high true percentages.

To further examine this bias, the data were categorized according to where the true percentage fell in the subjective distributions. Six categories were defined by the five fractiles used by the FRAC groups:

1. Below the .01 fractile
2. Between the .01 and .25 fractiles
3. Between the .25 and .50 fractiles
4. Between the .50 and .75 fractiles
5. Between the .75 and .99 fractiles
6. Above the .99 fractile

Data from well-calibrated subjects should have approximately 1%, 24%, 25%, 25%, 24%, and 1% of the true values falling into the six categories respectively when grouped in this manner.

Conservatism is shown by a large number of true percentages falling into categories 1 through 3 when the true percentage is low and a large number of true percentages falling into categories 4 through 6 when the true percentage is high. The percentages of responses in each category are broken down by elicitation procedure and true percentage in Table 1.

Considerable conservatism is apparent. With true percentages less than 50%, the percentage of true percentages falling into categories 1 through 3 is 68%, 70%, 62%, and 72% for the FRAC+, FRAC-, ODDS+, and ODDS- groups respectively; well over the 50% expected from well-calibrated subjects. Similarly, for true percentages greater than 50%; 73%, 77%, 67%, and 73% of the true percentages fall into categories 4 through 6 for the four groups respectively.

Table 1 also illustrates the two measures usually used to indicate

TABLE 1

Percentage of Responses in Each Category

Item No	True %	FRAC+ Category 1	2	3	4	5	6	ODDS+ Category 1	2	3	4	5	6
5	4	.61	.32	.07				.03	.62	.26	.09		
4	7	.29	.17	.35	.12		.06	.16	.72	.13			
20	14	.33	.56		.11			.06	.71	.18	.06		
16	18	.28	44	.19	.03	.06		.03	.56	.31	.03	.06	
6	24		.23	.23	.15	.38			.09	.03	.29	.53	.06
13	29	.11	.28	.28	.28	.06			.25	.13	.38	.25	
3	35	.08	.38	.38	.15			.03	.56	.35		.06	
2	44		.15	.23	.46		.15		.09	.12	.24	.50	.06
1	46			.17	.22	.28	.33	.07		.07	.47	.33	.07
19	54	.08	.03	.22	.28	.17	.22		.31	.19	.13	.28	
14	59	.36	.14	.36	.14			.06	.59	.21	.03	.12	
18	62	.06	.06	.44	.19	.13	.13		.38	.28	.09	.25	
8	66			.08	.15	.62	.15		.15	.15	.18	.53	
10	71		.11	.17	.33	.11	.28		.25	.06	.25	.38	.06
11	78	.07		.14	.07	.43	.29		.12	.09	.29	.50	
12	87				.21	.21	.57				.12	.79	.09
17	95					.25	.75			.06	.16	.78	
15	98			.07	.07	.21	.64		.06	.06	.38	.50	
Total		.128	.164	.187	.166	.155	.201	.024	.303	.148	.175	.332	.019

Item No	True %	FRAC- Category 1	2	3	4	5	6	ODDS- Category 1	2	3	4	5	6
5	96				.06	.17	.78				.16	.84	
4	93				.07	.36	.57		.06		.06	.79	.09
20	86			.07		.43	.50		.13	.03	.09	.69	.06
16	82			.03	.23	.27	.47		.06	.06	.12	.76	
6	76			.50	.22	.17	.11	.06	.28	.09	.06	.50	
13	71			.17	.17	.42	.25	.03	.35	.21	.06	.35	
3	65	.09	.12	.21	.18	.24	.18			.06	.31	.63	
2	56		.05	.44	.22	.05	.22		.11	.14	.14	.57	.04
1	54		.14	.25	.25	.14	.21		.59	.21	.15	.06	
19	46	.13	.13	.40	.13	.13	.07	.06	.41	.15	.21	.18	
14	41			.11	.28	.06	.56	.03	.18	.13	.19	.47	.06
18	38	.31	.08	.15	.15	.08	.23	.25	.06	.38	.28	.03	
8	34	.17	.39	.28	.06		.11	.03	.66	.19	.03	.09	
10	29		.07	.50	.14	.21	.07		.26	.15	.15	.44	
11	22	.12	.38	.26				.16	.38	.06	.09	.31	
12	13	.28	.55	.17	.24			.20	.80				
17	5	.79	.11	.07	.04			.06	.68	.24	.03		
15	2	.47	.35	.06	.12			.03	.47	.41	.09		
Total		.129	.141	.208	.143	.141	.204	.050	.301	.139	.123	.373	.014

the relative tightness of assessed distributions. The percentage of true values falling into the extreme tails of the subjective distributions (categories 1 and 6), called the surprise score, SS, should be approximately 2%. For both the FRAC+ and FRAC- groups, the SS is well above this figure (approximately 33% for both groups). For the ODDS groups, however, the SS's are only slightly higher than 2%; approximately 4% and 6% for the ODDS+ and ODDS- groups respectively.

The second measure of tightness, the interquartile score, IS, (categories 3 and 4) should be approximately 50% for well-calibrated subjects. Interquartile scores less than 50% indicate too tight distributions while scores greater than 50% indicate too loose distributions. Table 1 shows the IS's of 35%, 35%, 32%, and 26% for the FRAC+, FRAC-, ODDS+, and ODDS- groups respectively were all well below 50%.

These measures are, however, based on data accumulated over all values of true percentages. The effect of the true percentage on these measures is evident in Table 1 and summarized in Table 2 where the SS's and the IS's are given for middle range true values (40-60%) and for extreme true values (1%-10% and 90%-99%). In the FRAC groups, the distributions assessed on the extreme percentages are much too tight while distributions assessed on the middle range of true percentages are too tight as measured by the SS, but not as measured by the IS. The ODDS groups show only a slight difference in the IS's with the distributions assessed on the middle range of true percentages being nearer to the expected 50%.

The actual spread of the distributions elicited by the two procedures also can be compared by examining the slopes of median distributions of

Table 2

Surprise Scores and Interquartile Scores for Middle and Extreme True Percentages

	True Percentage			
	40%-60%		1%-10% 90%-99%	
Group	SS	IS	SS	IS
FRAC+	29%	52%	59%	17%
FRAC-	30%	52%	65%	10%
ODDS+	7%	34%	5%	29%
ODDS-	4%	33%	4%	25%

each item as presented in Figure 1. Two such comparisons were made: the interquartile range (IQR) and the maximum possible range (MPR). In the FRAC groups, the slope of the MPR was the slope between points at the 1:99 and the 99:1 odds levels. In the ODDS groups, the slope of the MPR was taken between points for the odds assessed for the lowest of the five percentages given for each item and the odds assessed for the highest percentage given. Interpolation was used to find the IQR in the ODDS groups. The results of these comparisons are given in Table 3, panel a. The symbol (≈) indicates that the difference between the average slopes was less than .10, while the inequality symbols show a difference greater than .10 in the indicated direction. Although .10 was an arbitrary criterion, the comparisons between the FRAC and ODDS groups suggest that

TABLE 3

Comparison of Average Slopes for Each Item

Between Group Comparison

FRAC Groups vs. ODDS Groups

Item	IQR Pos.	IQR Neg.	MPR Pos.	MPR Neg.
1	<	>	<	<
2	>	>	<	<
3	<	>	<	<
4	≃	>	<	>
5	≃	>	<	<
6	>	<	<	<
8	>	>	≃	≃
10	>	≃	<	<
11	≃	≃	≃	≃
12	>	>	<	<
13	>	≃	<	<
14	>	≃	<	<
15	≃	<	<	<
16	≃	>	<	<
17	>	<	≃	<
18	>	≃	≃	≃
19	>	>	<	<
20	≃	>	<	<

frequency

>	11	10	0	0
≃	6	5	4	3
>	1	3	14	15

Within Group Comparison

IQR vs. MPR

Item	FRAC Groups Pos.	FRAC Groups Neg.	ODDS Groups Pos.	ODDS Groups Neg.
1	≃	≃	<	<
2	≃	≃	<	<
3	≃	≃	<	<
4	≃	>	<	<
5	≃	≃	<	<
6	>	≃	<	<
8	≃	≃	<	<
10	≃	≃	<	<
11	≃	≃	≃	<
12	≃	≃	<	<
13	>	≃	<	<
14	≃	≃	<	<
15	≃	≃	<	<
16	≃	≃	<	<
17	≃	≃	<	<
18	≃	<	<	<
19	≃	≃	<	<
20	≃	≃	<	<

frequency

<	2	1	0	0
≃	16	16	1	0
>	0	1	17	18

Note:

(a)

> : Average slope in ODDS group is greater than FRAC group

≃ : Difference is less than .10

< : Average slope in FRAC group is greater than ODDS group

(b)

> : Average slope within MPR is greater than IQR

≃ : Difference is less than .10

< : Average slope within IQR range is greater than MPR.

the subjective probability distributions of the ODDS groups are tighter within the IQR and looser in the MPR, a finding consistent with the SS and IS data.

A qualitative idea about the shape of the subjective probability distributions can be determined by comparing the MPR slope with the IQR slope for each item. An IQR slope greater than the MPR slope shows that the subjective distribution has a higher density between the quartiles than in the rest of the distribution, while an MPR slope greater than the IQR slope shows lower density between the quartiles of the subjective distribution. Approximately equal slopes suggest a near uniform distribution. The results of these comparisons, shown in Table 3, panel b, show that for the most part, the distributions assessed by the FRAC groups are near uniform, while the distributions of the ODDS groups have higher density in the middle of the distribution, a more typical shape for probability distributions.

IV. Discussion

All three of the biases investigated were found to some degree in this study. In most cases, the extent of the bias depended on the value of the true percentage and/or the procedure used to elicit the subjective probability distributions.

The underestimation bias seemed to be much stronger when the fractile procedure was used to elicit the probability distributions than when the odds procedure was used. The extent of this bias has not been previously determined, although Seaver et al. (1975) suggested its possible existence.

Obviously, further study is warranted. However, should the underestimation bias prove to be rather common, a possible method for alleviating it comes to mind. Perhaps effects of the bias can be reduced or eliminated by obtaining both positive and negative assessments and combining them. The results of this study suggest that this may be a feasible approach to cancelling the effects of this bias.

The surprise scores and interquartile scores that are traditionally used to measure tightness seem to indicate that the distributions assessed in this study were too tight. But other factors suggest such a simple interpretation may be misleading. Can a distribution that covers a wide range of the possible percentages and is nearly uniform really be called tight? The answer to this question depends on the meaning given to the concept of "tightness". The tightness measured by surprise scores and interquartile scores is a relative tightness that compares assess distributions with actual occurrences. This type of tightness has received widespread attention in research not only because of the importance of knowing the correspondence between the subjective distributions and reality, but also because measures of this correspondence are easily available. Difficulties do exist with this concept of tightness. It only applies to collections of distributions, never to a single distribution. In this study, as in past studies, the tightness measures were determined not only across the distributions of a single assessor, but also across assessors.

The concept of tightness implied by examining the range and shape of single distributions is more absolute. Flat distributions covering a wide range of values would not normally be considered too tight. But questions such as "What is flat?" and "What is a wide range?" make this a difficult concept with which to deal. Therefore, although in an absolute sense the distributions assessed in this experiment may not be "too tight", the following discussion adopts the more traditional meaning of tightness for ease of discussion and comparison with previous findings.

The results of this experiment concerning both the tightness of distributions as measured by surprise and interquartile scores and the conservatism bias are generally in agreement with past results. Distributions elicited by both fractile and odds procedures exhibited conservatism. Tightness, as measured by surprise scores summed across all true values, was apparent in the FRAC groups but not in the ODDS groups, a finding consistent with the results of Seaver et al. The interquartile scores, again summed over all true percentages, were much lower than 50% for both groups also seemingly indicating the distributions were too tight. This result contrasts with the results of Seaver et al. who found interquartile scores near 50% for both elicitation procedures. No satisfactory explanation of this discrepancy seems to exist. However, the difference in interquartile scores should not necessarily be surprising, since the interquartile scores in this study are in the same general range as those obtained by Alpert and Raiffa (1969) and Schaefer and Borcherding (1973).

An important way in which this study differs from most past studies of the biases in subjective probability distributions is that the method used to document the biases allows some determination of the degree to which the conservatism bias may be partially responsible for surprise scores and interquartile scores that have traditionally been used to show the tightness of distributions. For example, when the true percentage is extremely low, conservatism will tend to displace the assessed distribution toward the right; so simply because of conservatism, the surprise score may be high.

The extent to which conservatism influences the measures traditionally used to assess the tightness of distributions depends both on the measure used (surprise score or interquartile score) and on the procedure used to assess the distributions. Although direct measurement of the relative contribution of conservatism to the surprise scores and interquartile scores was not possible, some inferences can be drawn. Conservatism could not account for the high surprise scores in the FRAC groups and the low interquartile scores in the ODDS groups when the true percentages were in the middle range (40%-60%). The similarity in surprise scores and interquartile scores between middle range and extreme true percentages for the ODDS groups suggests something other than conservatism produces these scores even with extreme true percentages when the odds elicitation procedure is used. The much higher surprise scores and lower interquartile scores for extreme true percentages than for true percentages in the middle range indicated considerable conservatism in the FRAC groups.

Because the procedures used to elicit the subjective probability distributions seem to have an effect on the extent of these biases, it becomes important to know which elicitation procedures may reduce or eliminate biases. The results of this study suggest that with regard to most of the biases, the odds procedure is better than the fractile procedure. The odds procedure does not produce the underestimation that the fractile method does. The surprise scores and interquartile scores are less dependent on the true percentage when the odds procedure rather than the fractile procedure is used. The odds procedure produces many fewer surprises than the fractile method. Only on the interquartile score does the fractile elicitation procedure seem to lead to better calibrated distributions than the odds procedure.

What causes the differences in the distributions elicited by the two methods? A rather simple phenomena may explain the difference in surprise scores. A tendency to avoid extreme responses would lead to both the large number of surprises in the FRAC groups and the small number of surprises in the ODDS groups. Since the responses of the FRAC groups are percentages, avoiding extreme percentages could lead to the .01 and .99 fractiles being too close together. However, in the odds elicitation procedure the responses are odds. A surprise can only occur if a subject assigns odds of at least 99:1 that the true percentage is greater or less than the given percentage when the given percentage is the true percentage. If the subjects avoid extreme responses, very few responses will be as large as 99:1, so there is little chance of a surprise occurring.

Consideration of the odds that must be assigned to the true percentage in the odds elicitation procedure for the true percentage to fall within the interquartile range also may explain why the interquartile scores were so low. Only if the odds assigned are 3:1 or less will the true percentage fall within the interquartile range. The high number of true percentages falling into categories 2 and 4 (see Table 1) may simply reflect the fact that subjects are more likely to make responses between 3:1 and 99:1. Perhaps suitable training in the use of smaller odds, particularly fractional values between 1:1 and 2:1, would help eliminate this bias. Use of probabilities rather than odds as the measures of uncertainty might also help with this problem in the interquartile range. Seaver et al. found the interquartile scores were higher when probabilities were used rather than odds, but in that case the interquartile scores were too high.

Since this experiment was conducted primarily to explore several possible biases and their relationship to the procedures used to elicit the distributions and the true values of the unknown percentages, most of the findings are suggestive rather than conclusive. The results do strongly suggest the existence of a previously unconfirmed bias, underestimation. They are also generally consistent with previous results with respect to conservatism and the tightness of distributions, but suggest possible interactions between the measurements of these biases. Following Seaver et al. (1975), there are also indications that the procedure used to elicit the subjective probability distributions also influences the extent of the biases. Generally, the ODDS procedure leads to less biased distributions than the FRAC procedure. All of these suggestive findings deserve further exploration in attempts to discover the processes by which people assign subjective probability distributions to unknown variables.

V. References

Alpert, M. and Raiffa, H. A progress report on the training of probability assessors. Unpublished manuscript, Harvard University, 1969.

Lichtenstein, S., Fischhoff, B., and Phillips, L. Calibration of Probabilities: The State of the Art. In Jungermann, H. and G. de Zeeuw (Eds.), Proceedings of the Fifth Research Conference on Subjective Probability, Utility, and Decision Making, in press.

Murphy, A.H., and Winkler, R.L. Scoring rules in probability assessment and evaluation. Acta Psychologica, 1970, 34, 273-286.

Phillips, L.D. and Edwards, W. Conservatism in a simple probability inference task. Journal of Experimental Psychology, 1966, 72, 346-354.

Schaefer, R.E. and Borcherding, K. The assessment of subjective probability distributions: A training experiment. Acta Psychologica, 1973, 37, 117-129.

Seaver, D.A., von Winterfeldt, D. and Edwards, W. Eliciting subjective probability distributions on continuous variables. SSRI Research Report 75-8, Social Science Research Institute, University of Southern California, August, 1975.

Selvidge, J. Experimental comparison of different methods for assessing the extremes of probability distributions by the fractile method. Management Science Report Series, Graduate School of Business Administration, University of Colorado, Boulder, Colorado, Report #75-13, 1975.

Wheeler, G.E., and Edwards,W. Misaggregation explains conservative inference about normally distributed populations. SSRI Research Report 75-11, Social Science Research Institute, University of Southern California, August, 1975.

VI. Appendix

Questions (Positive Items)

1. What percentage of the total world water area is contained in the Pacific Ocean? 46%

2. In 1971, what percentage of the total U.S. electrical energy was produced by coal? 44%

3. What percentage of all U.S. military personnel were in the Army in 1973? 35%

4. What percentage of all U.S. coastline is in California? 7%

5. What percentage of the total world gold production was produced in the U.S. in 1970? 4%

6. As of April, 1973, what percentage of all U.S. Federal employees were employed by the U.S. Postal Service? 24%

8. What percentage of all natural gas marketed in the world was produced in North and Central America during 1971? 66%

10. What percentage of all members of the Muslim religion lived in Asia in 1972? 71%

11. During the period from 1870 to 1971, what percentage of immigrants to the U.S. came from Europe? 78%

12. What percentage of registered voters voted in the 1972 U.S. general (Presidential) election? 87%

13. What percentage of the total U.S. advertising expenditures went to newspapers in 1970? 29%

14. During the period from 1950 to 1972, what percentage of all people examined for entry into the U.S. armed forces were found acceptable? 59%

15. As of 1973, what percentage of U.S. households had television? 98%

16. What percentage of the population of the city of Los Angeles was black in 1970? 18%

17. During 1971, what percentage of all television air time in France was occupied by programs produced in France? 95%

18. What percentage of U.S. Presidents have been lawyers? 62%

19. In 1972, what percentage of all automobiles produced in the U.S. were

manufactured by General Motors?   54%

20. What percentage of 1973 U.S. Federal revenue was produced by Corporation Income Tax?   14%

Research Distribution List

Department of Defense

Assistant Director (Environment and Life Sciences)
Office of the Deputy Director of Defense Research and Engineering (Research and Advanced Technology)
Attention: Lt. Col. Henry L. Taylor
The Pentagon, Room 3D129
Washington, DC 20301

Office of the Assistant Secretary of Defense (Intelligence)
Attention: CDR Richard Schlaff
The Pentagon, Room 3E279
Washington, DC 20301

Director, Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, Program Management Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209
(two copies)

Administrator, Defense Documentation Center
Attention: DDC-TC
Cameron Station
Alexandria, VA 22314
(12 copies)

Department of the Navy

Office of the Chief of Naval Operations (OP-987)
Attention: Dr. Robert G. Smith
Washington, DC 20350

Director, Engineering Psychology Programs (Code 455)
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217
(three copies)

Assistant Chief for Technology (Code 200)
Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

Office of Naval Research (Code 230)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Naval Analysis Programs (Code 431)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Operations Research Programs (Code 434)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research (Code 436)
Attention: Dr. Bruce McDonald
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Information Systems Program (Code 437)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research (ONR)
International Programs (Code 1021P)
800 North Quincy Street
Arlington, VA 22217

Director, ONR Branch Office
Attention: Dr. Charles Davis
536 South Clark Street
Chicago, IL 60605

Director, ONR Branch Office
Attention: Dr. J. Lester
495 Summer Street
Boston, MA 02210

Director, ONR Branch Office
Attention: Dr. E. Gloye and Mr. R. Lawson
1030 East Green Street
Pasadena, CA 91106
(two copies)

Dr. M. Bertin
Office of Naval Research
Scientific Liaison Group
American Embassy – Room A-407
APO San Francisco 96503

Director, Naval Research Laboratory
Technical Information Division (Code 2627)
Washington, DC 20375
(six copies)

Director, Naval Research Laboratory (Code 2029)
Washington, DC 20375
(six copies)

Scientific Advisor
Office of the Deputy Chief of Staff
for Research, Development and Studies
Headquarters, U.S. Marine Corps
Arlington Annex, Columbia Pike
Arlington, VA 20380

Headquarters, Naval Material Command (Code 0331)
Attention: Dr. Heber G. Moore
Washington, DC 20360

Headquarters, Naval Material Command (Code 0344)
Attention: Mr. Arnold Rubinstein
Washington, DC 20360

Naval Medical Research and Development Command (Code 44)
Naval Medical Center
Attention: CDR Paul Nelson
Bethesda, MD 20014

Head, Human Factors Division
Naval Electronics Laboratory Center
Attention: Mr. Richard Coburn
San Diego, CA 92152

Dean of Research Administration
Naval Postgraduate School
Monterey, CA 93940

Naval Personnel Research and Development Center
Management Support Department (Code 210)
San Diego, CA 92152

Naval Personnel Research and Development Center (Code 305)
Attention: Dr. Charles Gettys
San Diego, CA 92152

Dr. Fred Muckler
Manned Systems Design, Code 311
Navy Personnel Research and Development Center
San Diego, CA 92152

Human Factors Department (Code N215)
Naval Training Equipment Center
Orlando, FL 32813

Training Analysis and Evaluation Group
Naval Training Equipment Center
(Code N-00T)
Attention: Dr. Alfred F. Smode
Orlando, FL 32813

Department of the Army

Technical Director, U.S. Army Institute for the Behavioral and Social Sciences
Attention: Dr. J.E. Uhlaner
1300 Wilson Boulevard
Arlington, VA 22209

Director, Individual Training and Performance Research Laboratory
U.S. Army Institute for the Behavioral and
and Social Sciences
1300 Wilson Boulevard
Arlington, VA 22209

Director, Organization and Systems Research Laboratory
U.S. Army Institute for the Behavioral and
Social Sciences
1300 Wilson Boulevard
Arlington, VA 22209

Department of the Air Force

Air Force Office of Scientific Research
Life Sciences Directorate
Building 410, Bolling AFB
Washington, DC 20332

Robert G. Gough, Major, USAF
Associate Professor
Department of Economics, Geography and Management
USAF Academy, CO 80840

Chief, Systems Effectiveness Branch
Human Engineering Division
Attention: Dr. Donald A. Topmiller
Wright-Patterson AFB, OH 45433

Aerospace Medical Division (Code RDH)
Attention: Lt. Col. John Courtright
Brooks AFB, TX 78235

Other Institutions

The Johns Hopkins University
Department of Psychology
Attention: Dr. Alphonse Chapanis
Charles and 34th Streets
Baltimore, MD 21218

Institute for Defense Analyses
Attention: Dr. Jesse Orlansky
400 Army Navy Drive
Arlington, VA 22202

Director, Social Science Research Institute
University of Southern California
Attention: Dr. Ward Edwards
Los Angeles, CA 90007

Perceptronics, Incorporated
Attention: Dr. Amos Freedy
6271 Variel Avenue
Woodland Hills, CA 91364

Director, Human Factors Wing
Defense and Civil Institute of
Environmental Medicine
P.O. Box 2000
Downsville, Toronto
Ontario, Canada

Stanford University
Attention: Dr. R.A. Howard
Stanford, CA 94305

Montgomery College
Department of Psychology
Attention: Dr. Victor Fields
Rockville, MD 20850

General Research Corporation
Attention: Mr. George Pugh
7655 Old Springhouse Road
McLean, VA 22101

Oceanautics, Incorporated
Attention: Dr. W.S. Vaughan
3308 Dodge Park Road
Landover, MD 20785

Director, Applied Psychology Unit
Medical Research Council
Attention: Dr. A.D. Baddeley
15 Chaucer Road
Cambridge, CB 2EF
England

Department of Psychology
Catholic University
Attention: Dr. Bruce M. Ross
Washington, DC 20017

Stanford Research Institute
Decision Analysis Group
Attention: Dr. Allan C. Miller III
Menlo Park, CA 94025

Human Factors Research, Incorporated
Santa Barbara Research Park
Attention: Dr. Robert R. Mackie
6780 Cortona Drive
Goleta, CA 93017

University of Washington
Department of Psychology
Attention: Dr. Lee Roy Beach
Seattle, WA 98195

Eclectech Associates, Incorporated
Post Office Box 179
Attention: Mr. Alan J. Pesch
North Stonington, CT 06359

Hebrew University
Department of Psychology
Attention: Dr. Amos Tversky
Jerusalem, Israel

Dr. T. Owen Jacobs
Post Office Box 3122
Ft. Leavenworth, KS 66027

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 001855-4-T	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) New and Old Biases in Subjective Probability Distributions: Do They Exist and are They Affected by Elicitation Procedures[redacted].		5. TYPE OF REPORT & PERIOD COVERED Technical rept. 10/76-9/77
		6. PERFORMING ORG. REPORT NUMBER 77-4
7. AUTHOR(s) Tsuneko Fujii, David A. Seaver, Ward Edwards		8. CONTRACT OR GRANT NUMBER(s) Prime Contract N00014-76-C-0074 Subcontract 76-0308-0715
9. PERFORMING ORGANIZATION NAME AND ADDRESS Social Science Research Institute University of Southern California		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Advanced Research Projects Agency 1400 Wilson Blvd Arlington, Virginia 22209		12. REPORT DATE Aug[redacted]977
		13. NUMBER OF PAGES 36p.
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Decisions and Designs, Inc. Suite 100, 7900 Westpark Drive McLean, Virginia 22101 (under contract from Office of Naval Research)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Subjective probability distribution
Conservatism
Biases
Underestimation
Calibration
Validity

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

Past research indicates that people exhibit biases in assessing probability distributions on continuous variables. Three types of biases have been identified: too many true values falling into the extreme tails of the distributions, a displacement toward 50% for distributions assessed on percentages, and a general tendency to underestimate. This study explored the nature of these biases with particular emphasis on how they interact and how they are affected by the procedure used to elicit the distributions.

Two procedures were used to elicit subjective probability distributions on percentage variables. In a fractile procedure, subjects were asked to judge values of the unknown percentage that corresponded to fixed levels of their cumulative subjective probability distributions, while in an odds procedure, subjects judged the cumulative odds for fixed values of the unknown percentages. For all the unknown percentages, p%, distributions were assessed for both p% and 1-p%. The extent to which these assessments summed to less than 100% indicated a bias toward underestimation.

Underestimation was generally found when the fractile elicitation was used, but not when the odds procedure was used. Also, too many true values fell into the extreme tails of the distributions elicited by the fractile procedure, but no similar bias was found in distributions elicited by the odds procedure. The displacement toward 50% was found in distributions elicited by both procedures. This bias also appeared to be the cause of a considerable number of the true values in the extreme tails of the distributions. Many of the differences in the biases found when different elicitation procedures were used can probably be accounted for by subjects avoiding extreme responses and odds judgments between 1:1 and 2:1.