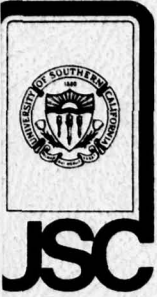


AD A 0 48570

DDC FILE COPY



UNIVERSITY OF SOUTHERN CALIFORNIA

# social science research institute

001855-5-T

*(Handwritten scribble)*

FG.

*(Handwritten number 2)*

## RESEARCH REPORT

THE EFFECTS OF RESPONSE SCALES ON  
LIKELIHOOD RATIO JUDGMENTS

WILLIAM G. STILLWELL  
DAVID A. SEAVER  
WARD EDWARDS

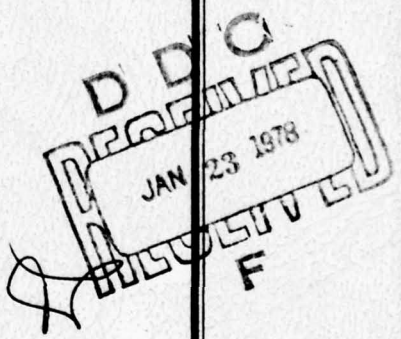
SPONSORED BY:  
ADVANCED RESEARCH PROJECTS AGENCY  
DEPARTMENT OF DEFENSE

MONITORED BY:  
ENGINEERING PSYCHOLOGY PROGRAMS  
OFFICE OF NAVAL RESEARCH  
CONTRACT No. N00014-76-C-0074, ARPA

APPROVED FOR PUBLIC RELEASE;  
DISTRIBUTION UNLIMITED;  
REPRODUCTION IN WHOLE OR IN PART IS PERMITTED  
FOR ANY USE OF THE U.S. GOVERNMENT

AUGUST 1977

SSRI RESEARCH REPORT 77-5



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 continuous 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.

Research Report 77-5

6 THE EFFECTS OF RESPONSE SCALES ON LIKELIHOOD RATIO JUDGMENTS.

9 Technical rept. Oct 76 - Sep 77, 1977

10 William G./Stillwell,  
David A./Seaver  
Ward/Edwards

Social Science Research Institute  
University of Southern California

11 Aug ~~1977~~ 1977

12 38 p.

DDC  
JAN 23 1978  
RESERVED  
F

Sponsored by

Defense Advanced Research Projects Agency

15 N00014-76-C-0074

390 664

1473  
mt

## Summary

Different methods of eliciting responses to the same question often produce different responses. In order to systematically study how response scales affect likelihood ratio judgments, two experiments were conducted. Experiment I manipulated two independent variables: the endpoints of the response scales (100:1, 1000:1, 10,000:1) and the spacing of the scales (logarithmic versus linear). Results compared the veridicality of responses on the six scales produced by crossing these factors plus another response mode in which subjects simply wrote their judgment in a blank (no scale).

Logarithmic scales produced responses that were both more veridical and more consistent than responses on linear scales which were, in turn, better than simple written responses. Measures of the effect of the endpoints were somewhat inconsistent and probably interacted with the range of veridical likelihood ratios. Judgments of relatively small likelihood ratios were affected by the spacing: linear spacing caused overestimation. Judgments of relatively large likelihood ratios were controlled more by the endpoints: higher endpoints produced larger judgments. Apparently, subjects use the range of the scale as information about the range of true likelihood ratios.

Experiment II manipulated two additional variables, data diagnosticity and the values of the true likelihood ratios. The results of Experiment I were confirmed while neither of the additional variables radically changed the effects of endpoints or spacing.

ACCESSION NO.	
NTIS	<input checked="" type="checkbox"/> Full Text
DDC	<input type="checkbox"/> B.M. Section
UNANNOUNCED	<input type="checkbox"/>
JUSTIFICATION	
BY _____	
DISTRIBUTION/AVAILABILITY CODES	
Dist:	SP. CIVIL
A	

## Contents

	Page
Summary .....	i
Figures .....	iii
Tables .....	iv
Acknowledgment .....	v
Disclaimer .....	vi
I. Introduction .....	1
II. Experiment I .....	3
1. Method .....	3
1. Subjects	
2. Apparatus	
3. Procedure	
2. Results .....	5
3. Discussion .....	13
III. Experiment II .....	16
1. Method .....	16
1. Subjects	
2. Procedure	
2. Results .....	18
3. Discussion .....	24
IV. References .....	27

## Figures

	Page
Figure 1: Scatterplots and Regression Lines of Log Responses Versus Log True Likelihood Ratios (Experiment I)	9

## Tables

	Page
Table 1: Correlations Between True Likelihood Ratios and Responses for Individual Subjects (Experiment I)	6
Table 2: Slopes and Intercepts of Responses Versus True Likelihood Ratios for Individual Subjects (Experiment I)	7
Table 3: Mean Absolute Deviations Between Log Responses and Log True Likelihood Ratios (Experiment I)	12
Table 4: Correlations Between True Likelihood Ratios and Responses for Individual Subjects (Experiment II)	19
Table 5: Average Slopes and Intercepts of Responses Versus True Likelihood Ratios (Experiment II)	20
Table 6: Correlations, Slopes, and Intercepts Median Responses Versus True Likelihood Ratios (Experiment II)	22
Table 7: Mean Absolute Deviations Between Log Responses and Log True Likelihood Ratios (Experiment II)	23

### 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

Judgments change in response to the information provided in the surrounding environment, regardless of whether or not the information is relevant to the judgment. Changing judgments in response to irrelevant information will usually lead to inconsistencies among judgments. Such inconsistencies pose a particular problem when the judgments serve as the basis for making decisions, as in decision analysis. Subjective judgments of both probability and utility are required for decision analysis--judgments which are known to be inconsistent in certain situations. For example, different methods of eliciting subjective probability distributions will produce different distributions (Seaver, von Winterfeldt, & Edwards, 1975; Stael von Holstein, 1971; Winkler, 1967). The questions asked to determine subjective probability distributions include no information that should cause the subjective distributions to change, yet consistent differences do occur.

If different elicitation methods lead to differences in the assessed probabilities, these differences need to be eliminated or taken into account. One way to approach this problem is to learn what causes these inconsistencies. For example, the type of response required affects the judgments. Responses to the same questions in odds and probabilities will typically not be equivalent as has been shown both in probabilistic inference tasks (Fujii, 1967; Phillips & Edwards, 1966) and in the assessment of subjective probability distributions (Seaver et al., 1975). In fact, even if the same type of response is required, the way in which it is recorded seems to systematically change the responses. Posterior odds judgments in probabilistic inference tasks have usually been larger when recorded on a logarithmic

scale than when simply written (Fujii, 1967; Phillips & Edwards, 1966). A similar difference has been shown between likelihood ratio judgments recorded on logarithmic scales and those that were written (Domas, Goodman, & Peterson, 1972). Goodman (1973) in reanalyzing the data from several experiments (including Domas et al.) to determine the effects of several variables on judgments of uncertainty, concluded that judgments recorded on logarithmic scales were generally larger regardless of accuracy. In some instances the larger responses were more veridical, while other times they were less veridical.

An unpublished pilot study by Seaver and von Winterfeldt conducted prior to the Seaver et al. experiment also suggested that another response scale variable--the upper endpoint--may affect odds or likelihood ratio judgments. Although the scale endpoints were not systematically manipulated, subjects' responses were apparently influenced by the endpoints. When subjects were very certain, they tended to respond with the scale endpoint regardless of its value, even though they had been instructed to respond off the scale if necessary.

The current experiments were undertaken to systematically explore how variations in the response scale affect likelihood ratio judgments. In particular, we were interested in the differences between responses on logarithmic scales, linear scales, and no scales; and in how the upper endpoints of the scales affect the responses. Knowledge of such differences should be of practical use to those who seek accurate quantification of uncertainty.

## II. Experiment I

### II.1. Method

II.1.1. Subjects. The subjects were 74 undergraduate students at the University of Southern California enrolled in an introductory psychology course. Participation in several experiments throughout the semester was required for credit in the course.

II.1.2. Apparatus. Stimuli for the experiment were seven inch (17.78 cm) sticks with one end painted red and the remainder of the stick painted white. Each stick represented a sample from one of two populations of sticks, each normally distributed with mean red lengths of five inches (12.7 cm) and two inches (5.08 cm) respectively and a common standard deviation of one inch (2.54 cm). The lengths of red and white were varied to produce true likelihood ratios from 2:1 to 12,000:1. Each of twenty-five different normal deviates were used to produce two sticks, one with more red than white and one with more white than red.

The population characteristics of the sticks were displayed to the subjects by two histograms, each a representative sample of one hundred sticks from one of the populations. These sticks were selected from a normal distribution and were spaced equidistant on the distribution function from minus to plus three standard deviations. The sticks from each population were randomly arranged to form the respective histograms. The displays were the actual size and color of the original stick populations with the population mean displayed by a heavy yellow horizontal line. These displays were visible to the subjects throughout the experiment.

Seven different response scales were used: three with logarithmically spaced markings and upper endpoints of 100:1, 1000:1, and 10,000:1; three with

linearly spaced markings and the same endpoints; and one with simply a blank to fill in. Henceforth these scales will be referred to as log100, log1000, log10000, lin100, lin1000, lin10000, and open. Each individual recorded responses, one to a page, in a booklet containing only a single type of response scale.

II.1.3. Procedure. Subjects received written instructions explaining the nature of the task and the experimental stimuli. The display histogram were described as random samples from the two populations. The written instructions further directed subjects that certainty was to be expressed in likelihood ratios and explained the concept of likelihood ratios.

Following the review of the written instructions, a short example of the two-hypothesis likelihood ratio estimation procedure was explained verbally. Both written and verbal instructions emphasized that when subjects' likelihood ratio estimates were greater than those provided on the scale, they were to make a mark at the top of the scale and simply write their numerical judgment.

Subjects then viewed the 50 sticks, one at a time, and responded with likelihood ratio judgments on the appropriate scales. The subjects were allowed to pick up the sticks or move them to get a better perspective, but were not allowed to compare them with previous sticks. For each stick the subjects chose which population was more likely to have produced the stick and indicated a likelihood ratio corresponding to their certainty.

The sticks were presented in four different randomized orders. Subjects were run in self-selected groups of from three to seven persons based on the time for which they registered on a sign-up sheet. Different response scales were assigned randomly to groups. The number of subjects using each

of the response scales was 11, 14, 9, 10, 9, 9, and 12 for the lin100, lin1000, lin10000, log100, log1000, log10000 and open scales respectively. Unequal numbers resulted from the failure of some subjects to follow directions properly in making their responses.

## II.2. Results

The data were subjected to a logarithmic transformation and all analyses were performed on the transformed data. The likelihood ratio responses were regressed on the true likelihood ratios for each individual subject. Table 1 shows the individual correlations from these analyses and the mean correlations for each response scale calculated using the Fisher-z transformation. The relatively large number of subjects with nonsignificant correlations ( $p > .05$ ) suggests considerable unreliability in some subjects' responses. This unreliability is more pronounced in subjects responding on linear scales (9 out of 34 subjects) than in subjects responding on logarithmic scales (1 out of 28 subjects). With the unreliability due to subjects with nonsignificant correlations removed, little, if any, difference exists among mean correlations.

Table 2 shows the slopes and intercepts of the individual regression analyses. The mean slopes and intercepts for each response scale were calculated excluding the subjects with nonsignificant correlations. A perfect correspondence between responses and true likelihood ratios would result in a slope of 1.0 and an intercept of 0.0. The most striking result is the difference in intercepts between linear and logarithmic scales. Intercepts on the logarithmic scales are consistently lower (closer to 0.0) than intercepts on the linear scales. The slopes also tend to increase as the endpoint of the scales increased with the exception of the lin1000 response

TABLE 1  
 Correlations Between True Likelihood Ratios and  
 Responses for Individual Subjects  
 (Experiment I)

		RESPONSE MODE							
		lin100	log100	lin1000	log1000	lin10000	log10000	log100000	Open
		.829	.481	.663	.769	.822	.726	.883	
		.846	.658	.811	.758	.531	.804	.742	
		.817	.914	.507	.738	.668	.634	.804	
		.911	.740	.502	.473	.518	.785	.701	
		.852	.589	.469	.800	.775	.765	.526	
		.397	.635	.485	.757	(-.005)	.708	.738	
		.473	.833	.370	.786	(-.138)	.710	.406	
		.603	.676	.785	.755	(.025)	.629	.619	
		.719	.522	.426	.770	(.215)	(-.278)	.809	
		(-.038)	.634	.644				.766	
		(.262)		.392				(.178)	
				(.195)				(-.220)	
				(.041)					
				(-.202)					
N	11 (2)	10 (0)	14 (3)	9 (0)	9 (4)	9 (1)	12 (2)		
$\bar{r}$	.68 (.75)	.69 (.69)	.47 (.57)	.74 (.74)	.44 (.68)	.54 (.73)	.60 (.72)		

Note: Non-significant correlations are in parentheses. N in parentheses is the number of subjects in the given response mode with non-significant correlations. Mean correlations in parentheses are calculated for response mode groups with non-significant correlations removed.

TABLE 2

Slopes and Intercepts of Responses Versus True Likelihood Ratios  
for Individual Subjects  
(Experiment I)

		RESPONSE MODE													
		lin100		log100		lin1000		log1000		lin10000		log10000		Open	
Slope	Inter.	Slope	Inter.	Slope	Inter.	Slope	Inter.	Slope	Inter.	Slope	Inter.	Slope	Inter.	Slope	Inter.
.05	1.78	.24	1.18	.08	2.63	1.26	-1.12	1.07	.50	.37	.00	.30	-.21		
.89	1.04	.43	.18	.65	.91	.61	1.01	.44	2.64	.37	.01	.62	.46		
1.02	.13	.45	.37	.51	1.29	.53	.37	.95	.30	.68	1.36	.16	.42		
.08	1.71	.54	.63	.34	1.86	.45	.42	.65	1.81	.64	1.77	1.11	.42		
1.05	-.14	.23	1.14	.10	2.57	.61	.83	.69	.63	1.20	-.70	.05	1.75		
.48	1.36	.41	.34	.05	2.82	.53	.90	(.00)	(.70)	.59	2.17	.30	1.92		
.26	.29	.46	.32	.08	2.64	.20	1.47	(-.18)	(3.79)	.92	1.05	.13	.70		
.32	1.04	.75	.26	.07	2.73	.75	.55	(.04)	(2.32)	.42	-.13	.73	1.28		
.08	1.76	.21	.92	.09	2.65	.68	.36	(.33)	(2.42)	(-.20)	(2.2)	.17	.38		
(-.02)	(1.01)	.30	.99	.65	-.08							.40	-.33		
(.06)	(1.70)		.29	.29	1.21							(.06)	(.72)		
			(.04)	(.04)	(2.62)							(-.22)	(2.70)		
			(.01)	(.01)	(2.78)										
			(-.11)	(-.11)	(1.66)										
(.47)	.997	(.402)	.633	(.265)	1.93	(.624)	.532	(.760)	1.18	(.649)	.691	(.397)	.679		
N	11	(2)	(0)	14	(3)	9	(0)	9	(4)	9	(1)	12	(2)		

Note: Parentheses indicate correlation for subject was non-significant. N in parentheses is the number of subjects in the given response mode with non-significant correlations. Mean slopes and intercepts in parentheses are calculated for response mode groups with individuals with non-significant correlations removed.

scale.

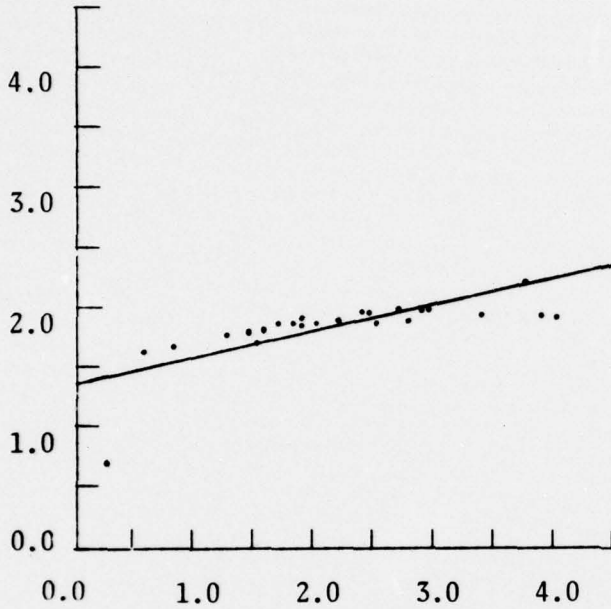
To provide numbers that represent each response scale without being influenced by the unreliability of the data, median responses were computed across subjects for each response scale at each of the 25 true likelihood ratios. Subjects with nonsignificant correlations were removed from this computation. The individual judgments used to calculate these medians were the arithmetic means of the responses to the two sticks with the same true likelihood ratio, but favoring different populations. Scatterplots of these medians and the regression lines and statistics are shown in Figure 1.

The dependence of subjects' likelihood ratio judgments on response scales is evidenced in several ways. Providing any scale for responses seems to increase the reliability of subjects' judgments as shown by the lower correlation for the open scale compared with correlations for five of the other six response scales: only lin100 has a lower correlation. In addition, all the correlations for logarithmic scales are noticeably higher than any of the correlations for the linear scales indicating that logarithmic spacing increases reliability. The slopes of the logarithmic scales are also generally higher (closer to 1.0) than the linear or open scales and the intercepts indicate that the logarithmic scales are superior to the linear or open scales. Thus, all three statistics favor the logarithmic scales over the linear and open scales.

The overall effects of the endpoints are less clear. The slopes obtained in this analysis confirm the tendency found in the individual data for the slopes to increase as the endpoints increase. No systematic effects on the correlations or intercepts are apparent. Not surprisingly, the scatterplots show that the endpoints clearly function as an upper bound for responses.

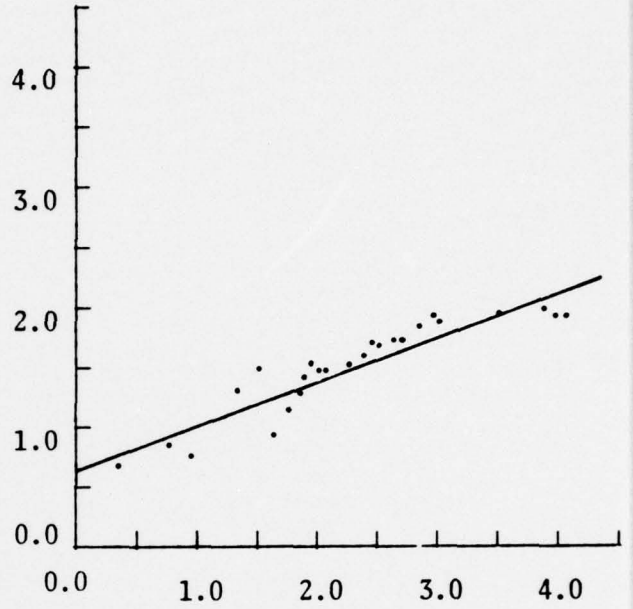
Scatterplots and Regression Lines of Log Responses Versus  
Log True Likelihood Ratios  
(Experiment I)

lin100



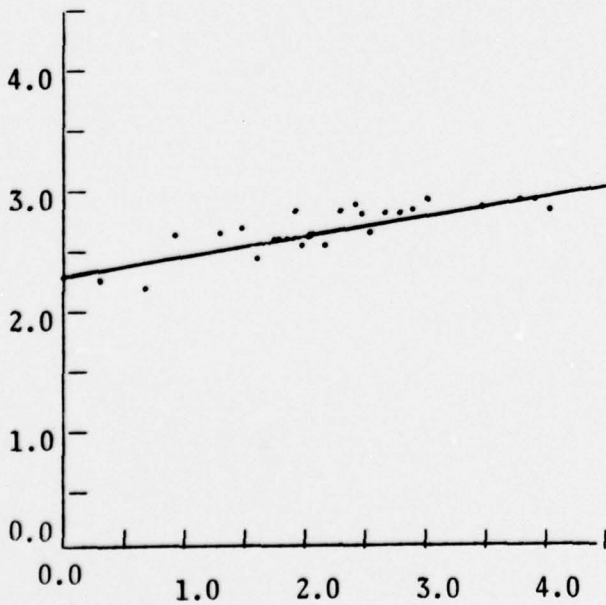
Correlation = .714  
Intercept = 1.421  
Slope = .204

log100



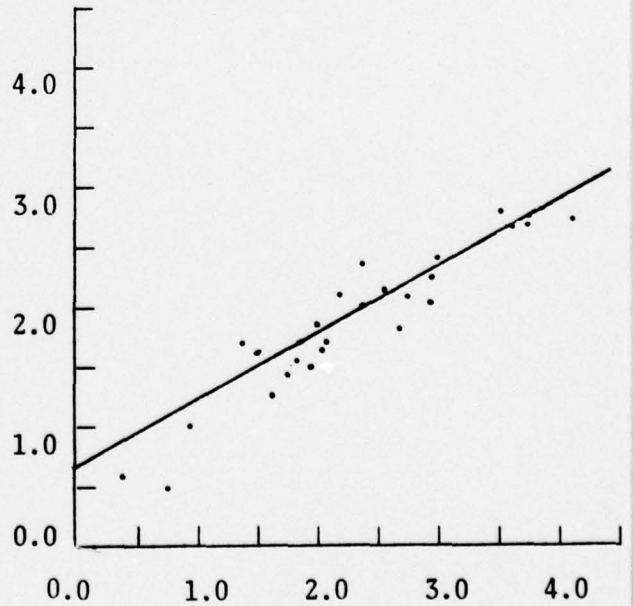
Correlation = .934  
Intercept = .693  
Slope = .372

lin1000



Correlation = .856  
Intercept = 2.443  
Slope = .148

log1000

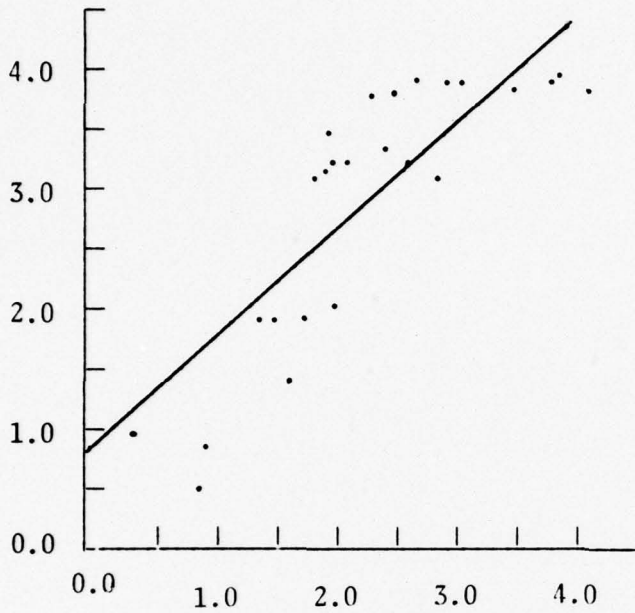


Correlation = .937  
Intercept = .696  
Slope = .571

Figure 1

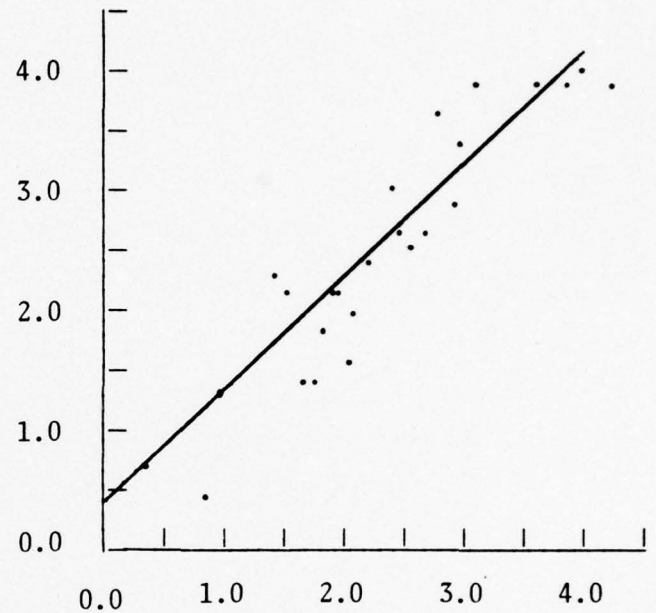
Scatterplots and Regression Lines of Log Responses Versus  
Log True Likelihood Ratios  
(Experiment I)

lin10000



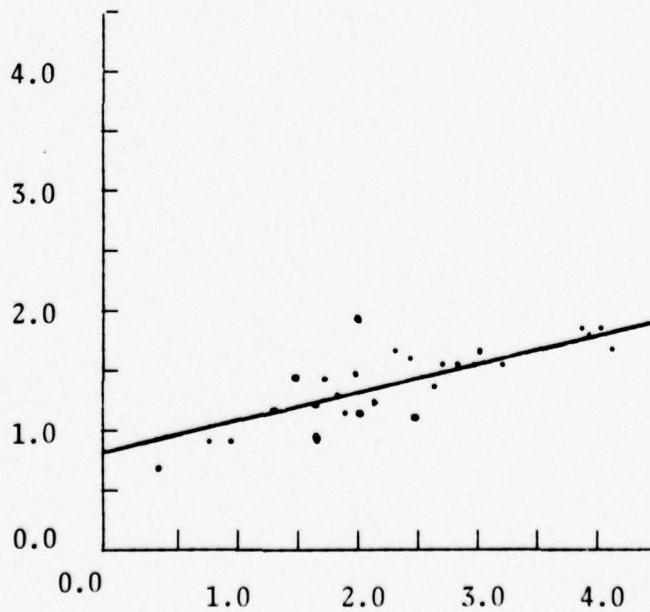
Correlation = .849  
Intercept = .789  
Slope = .957

log10000



Correlation = .930  
Intercept = .381  
Slope = .950

Open



Correlation = .771  
Intercept = .845  
Slope = .268

Figure 1 (cont.)

Each scale with 100:0 as the endpoint has a maximum median response less than 100:1 (2.0 on the logarithmic scale). A similar effect is also apparent for the other endpoints. In this respect the open scale seems most similar to the scales with 100:1 endpoints.

Because the large number of true likelihood ratios greater than 100:1 may have unduly influenced these findings, similar regression analyses were performed on only the median responses to true likelihood ratios less than 100:1 (12 values). Although the differences are less dramatic, the same general effects were found in this restricted range. The only striking difference was in the lin10000 scales where the slope increased to about 1.5 and the intercept decreased to -0.1.

Two final analyses, consisting of six planned comparisons each, were performed to determine the effects of the response scales on the correspondence between individual subjects' responses and true likelihood ratios (Hays, 1973, chapter 14). The measures used were the absolute values of the differences between the logarithm of the response and the logarithm of the true likelihood ratio. The six comparisons were log versus linear, log versus open, linear versus open, 100:1 endpoints versus 1000:1 endpoints, 100:1 endpoints versus 10000:1 endpoints, and 1000:1 endpoints versus 10000:1 endpoints. These comparisons were made both on data from all subjects and on data from only those subjects with significant correlations (see Table 1). The measures of correspondence used in these comparisons were the absolute value of the difference between the logarithm of the response and the logarithm of the true likelihood ratio.

The means of this measure for each response scale and the marginal means used in the planned comparisons are presented in Table 3. Significant

Table 3  
 Mean Absolute Deviations Between Log Responses  
 and Log True Likelihood Ratios

		Endpoints			Marginal Means
		100:1	1000:1	10000:1	
Spacing	Linear	.8296 (.7694)	.8421 (.8185)	1.2083 (.9909)	.9350 (.8483)
	Logarithmic	.8547 (.8547)	.6670 (.6670)	1.1584 (1.1613)	.8920 (.8929)
	Marginal Means	.8416 (.8100)	.7736 (.7592)	1.1834 (1.0761)	
	Open	1.1823	(1.1579)		

Note: Numbers in parentheses exclude subjects with nonsignificant correlations.

differences ( $p < .01$ ) yielded the following orders (from best to worst) for data from all subjects.

1000:1 endpoints → 100:1 endpoints → 10000:1 endpoints  
logarithmic → linear → open

Comparisons using data from only those subjects with significant correlations resulted in the following orderings. (All differences were significant at the .01 level except the 1000:1 endpoints versus 100:1 endpoints which was significant at the .02 level.)

1000:1 endpoints → 100:1 endpoints → 10000:1 endpoints  
logarithmic ↕ linear → open

### II.3. Discussion

This study indicates the existence of consistent biases in subjects' likelihood ratio judgments that are dependent upon the scale on which the judgments are recorded. Apparently information from the response scales that should be irrelevant is not treated as such by the subjects when making their responses.

The two factors manipulated in this study affect different ranges of likelihood ratio judgments. The spacing of the scales seems to control responses to relatively small likelihood ratios, while the scale endpoints exert more control over larger likelihood ratio judgments.

The logarithmic scales facilitated responses at the lower end of the scales leading to consistently more veridical responses than the linear scales. Subjects may have had more difficulty responding with small likelihood ratios on the linear scales because the small likelihood ratios were physically close together relative to the same likelihood ratios on the

logarithmic scales. For example, the distance between 1:1 and 10:1 on the  $\log_{10}10000$  scale used in this study was approximately 7.85 cm, but was only about .031 cm on the  $\text{lin}10000$  scale. The physically small region available for low responses on the linear scales may well have led subjects to avoid responses in that region. The relatively high intercepts for responses on linear scales support this conjecture.

The obvious effect of the scale endpoints is that they serve as a ceiling for responses. The slopes of median responses also generally increased as the endpoints increased. Thus, the results of the analysis of differences between responses and true likelihood ratios showing responses on scales with 1000:1 endpoints to be more veridical are somewhat surprising. The conflict between these results is probably primarily due to the difference between the use of medians and means. The lack of a rationale for choosing between these statistics suggests that conclusions concerning the effect of scale endpoints on the veridicality of judgments should not be drawn without more research.

Use of the open scale seems inadvisable. The correlations between median responses and true likelihood ratios indicated the open scale may produce judgments less closely tied to the true likelihood ratios, while analysis of the differences between responses and true likelihood ratios showed the responses were less veridical on open scales than on either logarithmic or linear scales. This is not surprising since any type of judgment would be expected to be more consistent when responses are made on physical scales rather than simply written.

The findings of Experiment I are consistent with the results reported by Domas et al. (1972) in that the slopes of the regression lines comparing

responses with true likelihood ratios are larger for logarithmically spaced scales. However, the slopes are less than 1.0 rather than greater as found by Domas et al. This difference can be explained by the relatively large true likelihood ratios used in this study. Larger likelihood ratios typically result in a decrease in the slope of such regression lines. Certain other differences are also apparent. While Domas et al. attribute the larger slopes with logarithmic scales to a tendency to make larger judgments, in this study the larger slopes are probably at least partially due to the increased use of small odds, and, therefore, intercepts closer to 0.0. Domas et al. do not report the intercepts of their data for a similar comparison to be made. In this study any tendency to make larger judgments seems more likely to be the result of higher endpoints rather than logarithmic scales.

Several conclusions tentatively can be drawn from this study: (1) any scale is better than no scale; (2) logarithmic scales are better than linear scales; (3) the absolute magnitude of responses depends heavily on the endpoint of the response scale. If these conclusions remain valid, they have considerable practical implications for the elicitation of subjective likelihood ratios. However, because of the apparent dependency of effects of the values of likelihood ratios, the true likelihood ratios of the stimuli used in this study may have been a critical factor in determining the overall effects. The stimuli used had a large  $d'$  and a wide range of likelihood ratios with relative emphasis on large likelihood ratios. Thus, they are quite dissimilar to stimuli used in other laboratory experiments which typically have lower values of  $d'$  (usually 2.2 or less) and true likelihood ratios more concentrated in a lower range. In order to explore

how  $d'$  and the range of true likelihood ratios affect these results, a second study was undertaken.

### III. Experiment II

Experiment II examined two factors which would extend knowledge of the nature of the response mode phenomenon. A less extreme level of data diagnosticity represented by a  $d'$  of 1.5, was used along with the original level of 3.0. Also, the method of selection of true likelihood ratios was varied: both the method used in Experiment I resulting in true likelihood ratios of 2:1 to 12,000:1 and a more typical method of generation by a normal random process were used. Both of these factors had led to the selection of generally large likelihood ratios in Experiment I which may have biased the results in favor of logarithmic scales with large endpoints.

#### III.1. Method

III.1.1. Subjects. One hundred and ninety-two undergraduates at the University of Southern California served as subjects for this experiment as a requirement for an introductory psychology class. Subjects were each paid \$3.00 for participation in the experiment.

III.1.2. Procedure. The normal process underlying the generation of data was the same as used in Experiment I, but the stimuli were changed. Subjects were told that samples were taken from a series of lakes and that the growth of a certain red algae was chemically analyzed. This red algae was said to be indicative of the likelihood that the sampled lake was polluted at the time of the sample. Subjects were told that, on the average, polluted lakes contained 38 parts per million red algae growth, while nonpolluted lakes averaged 32 parts per million. The standard deviations were 2.0 and 4.0 to produce the two levels of  $d'$ .

The original range of likelihood ratios was produced as in Experiment I and again they ranged from 2:1 to 12,000:1 with the same intermediate values as in Experiment I. The other range of likelihood ratios, termed

normal range likelihood ratios, was selected by a computer utility program for the generation of normal deviates that produced a series of 25 deviates from a normal population with a mean of zero and standard deviation of 1.0. These deviates were then converted to the population parameters defined in the study and the likelihood ratios were calculated. The resultant likelihood ratio ranges varied from 1.13:1 to 55.8:1 with  $d'=1.5$ , and from 1.62:1 to 28,566:1 with  $d'=3.0$ .

Written instructions explained the nature of the task and the experimental stimuli. Subjects were instructed to circle the more likely hypothesis and express certainty on the scale provided as a likelihood ratio between the two competing hypotheses. The concept of likelihood ratios was explained in more specific detail than in the first experiment. The experimenter explained that the midpoint between the two means should be the cutoff between samples favoring either hypothesis and that the more extreme the sample from this midpoint, the higher the likelihood ratio should be in favor of the hypothesis on that side of the midpoint. As in the first experiment, subjects were told that if the likelihood ratio judgments were larger than provided for on the scale, they were to mark the top of the scale and write their numerical judgment. Subjects then made fifty likelihood ratio judgments for samples from fifty hypothetical lakes. The order of presentation of these samples came in three different random sequences.

Subjects were run one to four at a time in self-selected groups based on the time for which they registered on a sign-up sheet. Twelve subjects were run in each of the 16 cells of a completely crossed  $2 \times 2 \times 2 \times 2$  design. The factors in addition to  $d'$  and the selection procedure for stimuli were again spacing (logarithmic and linear) and endpoints (100:1 and 10,000:1).

Judgments were made in booklets containing response scales similar to those used in Experiment I. The sample result from the red algae test appeared in the upper left corner of the response sheet with the words "The designated lake contains Red Algae (*Soficticus Grahamae*) tested at (sample result) parts per million. It is more likely to be (polluted or not polluted) with a likelihood of:".

### III.2. Results

All data were again transformed logarithmically and all analyses were performed on the transformed data. Likelihood ratio responses were regressed on the true likelihood ratios for each subject. Table 4 shows the individual correlations from these analyses and the mean correlations, calculated using the Fisher-z transformation, for each of the 16 cells in the design. Comparing across all levels of other factors, these correlations show no systematic differences between logarithmically spaced scales and linearly spaced scales. Also, no systematic differences are apparent for scales with endpoints of 100:1 versus scales with endpoints of 10,000:1. Despite the much more specific instructions and detailed explanation of the method for judging likelihood ratios, the relative number of nonsignificant ( $p > .05$ ) and negative correlations differs little from Experiment I (16.2% in Experiment I and 12.5% in Experiment II), although the difference is in the expected direction. Subjects with nonsignificant and negative correlations were removed from all subsequent analyses.

Table 5 shows the mean slopes and intercepts from the individual regression analyses. Again, as in Experiment I, the intercepts differ greatly between logarithmically and linearly spaced scales with the intercepts of log scales being closer to the correct 0.0. This is true regardless

Table 4  
Correlations Between True Likelihood Ratios  
and Responses for Individual Subjects  
(Experiment II)

<u>Linear</u>								
	100:1				10,000:1			
	RANGE= Normal		RANGE= 2:1 to 12,000:1		RANGE= Normal		RANGE= 2:1 to 12,000:1	
d' =	1.5	3.0	1.5	3.0	1.5	3.0	1.5	3.0
.778	.832	.845	.850	.854	.603	.674	.848	
.922	.812	.879	.804	.688	.881	.777	.830	
.839	.637	.862	.887	.701	.827	.870	.715	
.631	.821	.711	.787	.914	.578	.691	.790	
.913	.756	.908	.813	.881	.728	.961	.511	
.933	.661	.777	.697	.821	.757	.774	.754	
.861	.925	.825	.920	.513	.897	.707	.877	
.901	.971	.733	.802	.838	.637	.760	.831	
.950	(.094)	.838	(.135)	.806	.836	.879	.439	
.957	(-.772)	.603	(.236)	.539	.826	.869	.439	
.767	(.151)	.881	(-.239)	.695	.760	.605	(.227)	
(.012)	(-.898)	.816	(-.537)	.660	(.323)	(.290)	(.358)	
Averages	.885	.840	.820	.831	.770	.753	.781	.684
<u>Logarithmic</u>								
	100:1				10,000:1			
	RANGE= Normal		RANGE= 2:1 to 12,000:1		RANGE= Normal		RANGE= 2:1 to 12,000:1	
d' =	1.5	3.0	1.5	3.0	1.5	3.0	1.5	3.0
.941	.740	.836	.486	.973	.420	.802	.617	
.848	.783	.902	.774	.830	.827	.879	.622	
.914	.398	.873	.585	.907	.817	.913	.834	
.946	.523	.791	.807	.862	.958	.727	.950	
.986	.927	.835	.890	.754	.619	.559	.835	
.867	.936	.842	.723	.874	.557	.907	.948	
.846	.782	.928	.894	.829	.811	.887	.827	
.495	.930	.562	.893	.968	.611	.807	.887	
.735	.934	.917	.725	.799	.927	.852	.798	
.720	.881	.663	.673	.922	.844	.824	(.031)	
.402	(.145)	.674	.828	(-.795)	.480	.789	(-.540)	
(-.635)	(.319)	(-.612)	(.320)	(.290)	.680	(.160)	(.097)	
Averages	.861	.781	.827	.755	.865	.765	.813	.846

Note: Nonsignificant correlations are removed from averages.

Table 5  
Average Slopes and Intercepts of Responses  
Versus True Likelihood Ratios  
(Experiment II)

		Linear		Logarithmic	
		100:1	10,000:1	100:1	10,000:1
RANGE=Normal	d'=1.5	b= .851 a= .576	b=1.000 a=2.624	b= .674 a= .490	b=1.503 a= .625
	d'=3.0	b= .270 a=1.048	b= .266 a=2.720	b= .265 a= .619	b= .363 a=1.903
RANGE=2:1 to 12,000:1	d'=1.5	b= .227 a= 1.100	b= .369 a=2.466	b= .389 a= .547	b= .679 a=1.037
	d'=3.0	b= .348 a= .887	b= .251 a=2.976	b= .249 a= .881	b= .542 a= .889

Note: Subjects with nonsignificant correlations between true likelihood ratios and response likelihood ratios are not represented in the calculations in this table. Slopes are represented by b, intercepts by a.

of the range of the true likelihood ratios or the  $d'$  condition in which the subject responded. The slopes of responses on logarithmically spaced scales and linearly spaced scales also differ with the average slope for logarithmically spaced scales closer to the optimal value of 1.0. Endpoints also affected slopes: scales with an upper endpoint of 10,000:1 have an average slope closer to 1.0.

Medians were calculated across subjects for response mode groups at each level of likelihood ratio and these medians were regressed on true likelihood ratios. These correlations, slopes and intercepts are broken down by factors in Table 6. Logarithmic scales seem to be superior to linear scales as evidenced by higher correlations, slopes closer to 1.0 and intercepts closer to zero, but these criteria may not completely reflect the accuracy of the judgments. A question arises in the evaluation of the regression analysis in the case where either the slope is less than 1.0 and the intercept greater than 0.0, or the slope is greater than 1.0 and the intercept is less than 0.0. In either case, the subject may be making responses in the correct range of true values, but the deviation might reflect some specific bias such as avoidance of high and low range responses. Scales with upper endpoints of 10,000:1 had a somewhat higher correlation between response likelihood ratios and the true likelihood ratios, but the superiority of the slope of scales with either endpoint was not definitive in the light of the extremely high intercepts for those scales. Subjects could well be radical in their judgments when using the higher endpoint, despite the slope being less than 1.0.

To investigate this possibility, an analysis of variance was done on difference scores calculated as in Experiment I. Table 7 shows the means for this ANOVA. Significant differences were found for both endpoints and

Table 6

Correlations, Slopes, and Intercepts  
Median Responses versus True Likelihood Ratios  
(Experiment II)

		Linear		Logarithmic	
		100:1	10,000:1	100:1	10,000:1
RANGE=Normal	d'=1.5	r= .898 b=1.009 a= .537	r= .846 b= .770 a=2.950	r= .891 b= .749 a= .419	r= .962 b=1.404 a= .479
	d'=3.0	r= .842 b= .207 a=1.205	r= .877 b= .291 a=2.836	r= .940 b= .345 a= .522	r= .789 b= .432 a=1.995
RANGE=2:1 to 12,000:1	d'=1.5	r= .857 b= .201 a=1.236	r= .836 b= .334 a=2.860	r= .893 b= .375 a= .541	r= .950 b= .646 a= .938
	d'=3.0	r= .886 b= .275 a=1.114	r= .848 b= .217 a=3.206	r= .784 b= .274 a= .945	r= .945 b= .522 a=1.018

Note: Subjects with nonsignificant correlations between true likelihood ratios and response likelihood ratios are not represented in the calculations in this table. Correlations are represented by r, slopes by b, and intercepts by a.

Table 7

Mean Absolute Deviations Between Log Responses  
and Log True Likelihood Ratios  
(Experiment II)

		Linear		Logarithmic		Marginal Means
		100:1	10,000:1	100:1	10,000:1	
RANGE=Normal	d'=1.5	.618	2.628	.385	1.083	1.179
	d'=3.0	.925	1.171	1.236	1.077	1.102
RANGE=2:1 to 12,000:1	d'=1.5	.836	1.340	.945	.830	.988
	d'=3.0	.743	1.369	.922	.624	.915
Marginal Means		.781	1.627	.872	.904	

Note: Subjects with nonsignificant correlations between Log Response and Log True are not included in this table.

spacing ( $p < .001$ ). Logarithmic scales and scales with endpoints of 100:1 result in responses which are significantly closer to true. No significant difference was found for subjects under differing  $d'$  conditions, but subjects' assessments were more veridical when responding to a normal range of true likelihood ratios than when the likelihood ratios were arbitrarily chosen to cover the range from 2:1 to 12,000:1.

Several interactions were significant but the magnitude of the effects was generally minimal except for the endpoint by spacing interaction which accounted for 10.3% of the variance. Other factors which accounted for appreciable amounts of the variance were endpoint (13.1%), spacing (7.6%) and the  $d'$  by endpoint interaction (7.8%). The magnitude of these effects may be contrasted with the main effect of the range which, although significant ( $p < .001$ ), accounted for only 2.7% of the variance.

### III.3. Discussion

Response-mode-produced biases in subjects' likelihood ratio judgments appear to be pervasive. The amount and specific dimensions of the biases are primarily dependent upon the characteristics of the response mode as well as the exact nature of the task and data generator. Logarithmically spaced scales generally seem to result in responses being significantly closer to the true response. This may be because logarithmic scales facilitate the use of responses near 1:1. Or, subjects may use (probably unconsciously) the fact that distances on logarithmic scales should be linearly related to the value of the random variable serving as the stimulus. This follows from the true likelihood being an exponential function of the random variable.

Differences in responses resulting from upper endpoints of 100:1 and 10,000:1 reflect a general tendency for subjects to maintain a larger magnitude of response when a larger upper endpoint is used. The upper endpoint may serve as a ceiling for responses, for example, producing judgments on the 100:1 scales which would never exceed that upper bound. Such a simple explanation cannot, however, explain why responses on scales with 100:1 endpoints are more accurate than responses on scales with 10,000:1 endpoints, even with  $d'=3.0$  and/or true likelihood ratios ranging from 2:1 to 12,000:1. In these conditions, the relatively large number of true likelihood ratios larger than 100:1 would suggest that scales with endpoints of 10,000:1 should lead to more accurate responses.

On the other hand, larger upper endpoints could be perceived by the subjects as conveying information as to the range of likely values in which their judgments should fall. Larger endpoints may suggest generally larger likelihood ratios, thus leading to considerable overestimation of small and middle range true likelihood ratios. The larger intercepts of responses on scales with 10,000:1 endpoints exemplifies this possibility.

As in the first experiment, findings are consistent with the results of Domas et al. (1972) in that slopes of the regression lines comparing response likelihood ratios with true likelihood ratios are larger for scales with logarithmic spacing. Still, despite the addition of a less extreme  $d'$  in Experiment II, slopes remain less than 1.0 in most cases, as opposed to the Domas et al. study where slopes were generally greater than 1.0. Still,  $d'$  cannot be ruled out completely as a contributing factor since Domas et al. used levels of  $d'$ , .46 to 1.14, which reflected relatively undiagnostic data.

In summary, response scales have been shown to be a consistent factor when subjects are making likelihood ratio judgments. Although logically irrelevant to the judgments being made, both the magnitude of likelihood ratios presented on the scale and the spacing of those ratios contribute to systematic biases in the subjects' responses. The results of Experiment II substantiated the findings of Experiment I as to the effects of endpoint and spacing of response scales. Experiment II went further to show that these results could not be attributed to either the effect of the extreme  $d'$  or the extreme nature of the true likelihood ratios in Experiment I. Generally, subjects were better able to estimate the likelihood ratios when they were responding on logarithmically spaced scales. Further, subjects' performance was somewhat improved when the upper endpoint was less than the highest one presented in these two studies (10,000:1).

When the types of judgments involved in these studies are necessary inputs to decision making, the biases encountered here should be taken into account when deciding how the judgments are to be elicited. The results of these two studies show that consideration should be given to the diagnosticity of the data with which the person making the judgment will be dealing as well as the range of the true likelihood ratios he or she is likely to encounter.

#### IV. References

- Domas, P., Goodman, B., & Peterson, C. Bayes's Theorem: Response scales and feedback. Technical Report No. 037230-5-T, Engineering Psychology Laboratory, University of Michigan, September, 1972.
- Fujii, T. Conservatism and discriminability in probability estimation as a function of response mode. Japanese Psychological Research, 1967, 9, 42-47.
- Goodman, B. Direct estimation procedures for eliciting judgments about uncertain events. Technical Report No. 011313-5-T, Engineering Psychology Laboratory, University of Michigan, 1973.
- Hays, W. Statistics for the Social Sciences, 2nd Edition, New York: Wiley, 1973.
- Phillips, L., & Edwards, W. Conservatism in a simple probability inference task. Journal of Experimental Psychology, 1966, 72, 346-352.
- Seaver, D., von Winterfeldt, D., & Edwards, W. Eliciting subjective distributions on continuous variables. Research Report No. 75-8, Social Science Research Institute, University of Southern California, August, 1975.
- Stael von Holstein, C. Two techniques for assessment of subjective probability distributions--An experimental study. Acta Psychologica, 1971, 35, 478-494.
- Winkler, R. The assessment of prior distributions in Bayesian analysis. Journal of the American Statistical Association, 1967, 62, 776-800.

## 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 (ONRT  
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. Glove 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  
Naval 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. Uhlener  
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 Orlandy  
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 001855-5-T	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
✓ 4. TITLE (and Subtitle) The Effects of Response Scales on Likelihood Ratio Judgments	5. TYPE OF REPORT & PERIOD COVERED Technical 10/76-9/77	6. PERFORMING ORG. REPORT NUMBER SSRI 77-5
7. AUTHOR(s) William G. Stillwell, David A. Seaver, and Ward Edwards	8. CONTRACT OR GRANT NUMBER(s) Prime Contract N00014-76-C-0074 Subcontract 76-030-0715	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
9. PERFORMING ORGANIZATION NAME AND ADDRESS Social Science Research Institute University of Southern California Los Angeles, California 90007	11. CONTROLLING OFFICE NAME AND ADDRESS Advanced Research Projects Agency 1400 Wilson Blvd. Arlington, Virginia 22209	12. REPORT DATE August, 1977
✓ 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)	13. NUMBER OF PAGES	15. SECURITY CLASS. (of this report) Unclassified
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 Likelihood ratio Logarithmic scale		
✓ 20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Different methods of eliciting responses to the same question often produce different responses. In order to systematically study how response scales affect likelihood ratio judgments, two experiments were conducted. Experiment I manipulated two independent variables: the endpoints of the response scales (100:1, 1000:1, 10,000:1) and the spacing of the scales (logarithmic versus linear). Results compared the veridicality of responses on the six scales produced by crossing these factors plus another response mode in →		

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

which subjects simply wrote their judgment in a blank (no scale).

Logarithmic scales produced responses that were both more veridical and more consistent than responses on linear scales which were, in turn, better than simple written responses. Measures of the effect of the endpoints were somewhat inconsistent and probably interacted with the range of veridical likelihood ratios. Judgments of relatively small likelihood ratios were affected by the spacing: linear spacing caused overestimation. Judgments of relatively large likelihood ratios were controlled more by the endpoints: higher endpoints produced larger judgments. Apparently, subjects use the range of the scale as information about the range of true likelihood ratios.

Experiment II manipulated two additional variables, data diagnosticity and the values of the true likelihood ratios. The results of Experiment I were confirmed while neither of the additional variables radically changed the effects of endpoints or spacing.

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

**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.
- 77-5 William G. Stillwell, David A. Seaver, and Ward Edwards. The Effects of Response Scales on Likelihood Ratio Judgments. August, 1977.