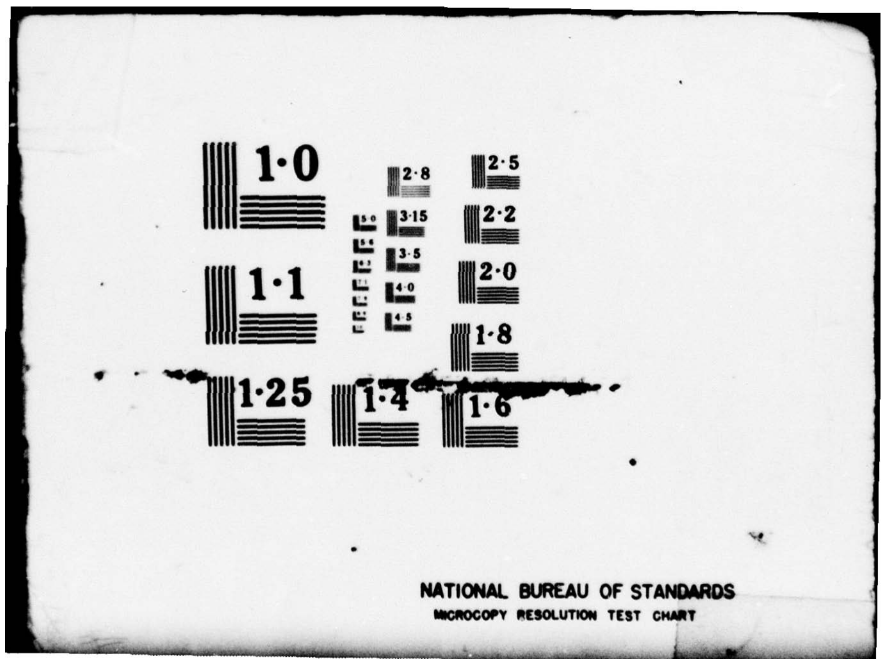
AD-A071 812 DECISIONS AND DESIGNS INC MCLEAN VA F/G 5/10
PRELIMINARY INVESTIGATIONS INTO THE PSYCHOLOGICAL FOUNDATIONS O--ETC(U)
JUN 79 J J WEISS, P J STICHA, M L DONNELL N00014-78-C-0651
UNCLASSIFIED PR-79-9-90 NL

1 OF 1
AD
A071812

END
DATE
FILMED
8-79
DDC


1·0
2·8
2·5
3·15
2·2
3·5
1·1
2·0
4·0
4·5
1·8
1·25
1·4
1·6
NATIONAL BUREAU OF STANDARDS
MICROCOPY RESOLUTION TEST CHART

June 1979

INTERIM TECHNICAL REPORT PR 79-9-90

Preliminary Investigations Into the Psychological Foundations of Fuzzy Reasoning

Jonathan J. Weiss
Paul J. Sticha
Michael L. Donnell

LEVEL II

DECISIONS and DESIGNS, INC.

Prepared for
Office of Naval Research
Contract No. N00014-78-C-0651

Decisions and Designs, Incorporated

Suite 600, 8400 Westpark Drive, P.O. Box 907
McLean, Virginia 22101

(703) 821-2828

23 July 1979

MEMORANDUM TO THE ADDRESSEES:

The research reflected in the attached report was conducted by Decisions and Designs, Incorporated and was supported by the Office of Naval Research under contract N00014-78-C-0651. This report is forwarded to your attention in accordance with the designated Distribution List.

We hope this report will prove useful to you or to your agency. Any comments or suggestions you might choose to offer regarding this report are welcomed.

Charles P. Annis
Charles P. Annis
Publications Manager

CPA:rh

Enclosure: Preliminary Investigations into the Psychological Foundations of Fuzzy Reasoning, Interim Report PR 79-9-90.

INTERIM TECHNICAL REPORT PR 79-9-90

PRELIMINARY INVESTIGATIONS INTO THE PSYCHOLOGICAL FOUNDATIONS OF FUZZY REASONING

by

Jonathan J. Weiss, Paul J. Sticha, and Michael L. Donnell

Prepared for

Office of Naval Research
Contract N00014-78-C-0651

June 1979

DECISIONS and DESIGNS, INC.
Suite 600, 8400 Westpark Drive
P.O.Box 907
McLean, Virginia 22101
(703) 821-2828

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: PR-79-9-90	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): PRELIMINARY INVESTIGATIONS INTO THE PSYCHOLOGICAL FOUNDATIONS OF FUZZY REASONING	5. TYPE OF REPORT & PERIOD COVERED: Interim Technical
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Jonathan J. Weiss, Paul J. Sticha, Michael L. Donnell	8. CONTRACT OR GRANT NUMBER(s): N00014-78-C-0651
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Decisions and Designs, Inc., P.O. Box 907, McLean, Virginia 22101	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, 800 North Quincy Street, Arlington, Virginia 22217	12. REPORT DATE: June 1979
	13. NUMBER OF PAGES: 45
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	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)

Fuzzy Set Theory
Psychological Measurement & Scaling
Reasoning
Logic
Mathematical Psychology
Psychophysics

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

Although fuzzy set theory has been rapidly gaining popularity as a rigorous framework for incorporating imprecision into quantitative reasoning, a brief survey of the existing literature indicates the lack of any consistently applied operational definition of the fundamental concept of membership function. In the present case, the membership of an element x in a set S is defined as the degree of truth of the state-

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ment "x is a member of S." It then becomes necessary to develop an empirically valid scale of truth which allows not only the binary extremes of "true" and "false," but also the continuum of intermediate values.

In a series of experiments, subjects performed two tasks: pairwise comparison, in which they selected the truer of two sentences; and direct numerical scaling, where they rated truth on a 0-to-100 scale. Results indicated that (1) the high degree of transitivity in each subject's paired-comparison judgments leads us to reject the hypothesis of a two-valued true-false logic in favor of a continuum of values; (2) ability to discriminate, as judged by the consistency between direct ratings and paired-comparison judgments, seems to be uniform along the true-false continuum, again favoring the hypothesis of a continuum of truth values over that of a binary categorical judgment; and (3) the high correlation between an item's aggregate binary preference score for a given subject and that subject's direct rating for the item indicates that at least two different methods of inferring degree of truth are highly consistent. Significance of these results and plans for further efforts to develop scaling methods are discussed.

CONTENTS

	Page
DD FORM 1473	ii
FIGURES	v
TABLES	vi
1.0 INTRODUCTION	1
2.0 REVIEW OF EMPIRICAL LITERATURE	3
2.1 Methods of Assessment	3
2.2 Experimental Findings	5
3.0 THE EXPERIMENT	8
3.1 Method	9
3.1.1 Subjects	9
3.1.2 Stimuli	9
3.1.3 Procedure	10
3.2 Results	11
3.2.1 Transitivity of paired comparison judgments	11
3.2.2 Categorical perception of truth	17
3.2.3 Consistency of paired-comparison choices and direct estimates	18
3.3 Discussion	21
4.0 OUTLINE OF FUTURE RESEARCH PLANS	23
4.1 Plans for Experiment Number 2	23
4.2 Long-Range Planning and Evaluation	24
REFERENCES	28
CONTRACT DISTRIBUTION LIST	30
SUPPLEMENTAL DISTRIBUTION LIST	31
APPENDIX A. LIST OF STIMULUS SENTENCES	A-1
APPENDIX B. A BRIEF EXPLANATION OF "STEM-AND-LEAF" DIAGRAMS	B-1

FIGURES

Figure		Page
3-1	TEST OF CATEGORICAL PERCEPTION OF TRUTH	19

TABLES

Table		Page
3-1	STEM-AND-LEAF DISPLAY OF NUMBER OF INTRANSITIVE TRIPLES	12
3-2	CONSISTENCY MEASURES FOR PAIRED-COMPARISON DATA	13
3-3	STEM-AND-LEAF DISPLAY OF INFERRED PROBABILITY OF PAIR REVERSAL	16
3-4	STEM-AND-LEAF DISPLAY OF CORRELATION BETWEEN PAIRED-COMPARISON SCORE AND DIRECT TRUTH ESTIMATE	20

PRELIMINARY INVESTIGATIONS INTO THE
PSYCHOLOGICAL FOUNDATIONS OF FUZZY REASONING

1.0 INTRODUCTION

This technical report contains a summary of the empirical research designed and performed by members of the technical staff of Decisions and Designs, Incorporated (DDI). The research reported here is the first part of a more extensive program whose overall purpose is to define a firm empirical and theoretical foundation for the proper use of the concepts and operations associated with the theory of fuzzy sets. Such a basis should include operational definitions of the concepts involved; experimental validation of the correspondence between the empirical system thus described and some symbolic or numerical system; and specification of the permissible domain for each operation, as well as its implications in terms of scaling, measurement, and invariance under transformations.

Fuzzy set theory is a comparatively new field, originally developed by systems engineers as a method of dealing approximately with highly complex systems. The progress made during the initial decade of fuzzy set development has been strongest, therefore, in the areas of practical implementation and mathematical representation. However, in the area of empirical behavioral validation of the theory, the literature has been sparse, incomplete, and inconsistent. Because the mathematical properties of a system that would satisfy our naive intuitive definitions are not uniquely determined, and because the goals and backgrounds of fuzzy set researchers differ widely, there have been to date no definitive, standard set of notations and formulas, and no consistent framework in which to compare and reconcile results from different researchers. This inconsistency has been somewhat of an embarrassment to those attempting to assimilate the new technology, often leading to the conclusion that the concept of fuzziness is inherently inconsistent or illogical. The present studies are designed to demonstrate that there does exist an operational, empirically valid set of definitions, formulas, etc., which can act as a unifying framework for fuzzy set theory.

The remaining three sections of this report deal with completed and ongoing DDI research into the empirical foundations of fuzzy set theory and its related logic. Section 2.0 contains a brief survey of the (relatively sparse) current literature concerning empirical psychological studies of reasoning with imprecisely defined concepts. Section 3.0 describes the first completed experiment in a series developed and conducted by Drs. Jonathan J. Weiss, Paul J. Sticha, and Michael L. Donnell, all affiliated with DDI. Finally, Section 4.0 outlines our plans for future research, including one experiment currently in progress and several other possible future continuations concluding with an evaluation of fuzzy set theory in the light of the evidence thus far collected, with implications for future research into the theory, methodology, and applications of fuzzy sets.

2.0 REVIEW OF EMPIRICAL LITERATURE

Despite much recent theoretical interest in fuzzy set theory, and frequent assertions that human reasoning may be organized by imprecise rules, there has been little empirical research to verify the consistent use by individuals of fuzzy reasoning processes. The little research already conducted does not present a coherent picture of the human as a processor of fuzzy logical information because there has been no agreement about the types of reasoning tasks to which fuzzy set theory applies, and because basic questions about consistency in the use of imprecise information have not been addressed. This section reviews the limited empirical literature testing fuzzy set theories of human reasoning, paying particular attention to the points mentioned above.

2.1 Methods of Assessment

Although the axioms of fuzzy set theory have the potential for application to a variety of human activities, care must be taken that results in one area are not generalized unjustifiably to different reasoning tasks. Recent empirical research has investigated fuzzy reasoning in a variety of contexts. Graded-set membership functions have been inferred from confidence judgments (Macvicar-Whelan, 1978), choice proportions over individuals (Hersh & Caramazza, 1976), paired comparison and direct judgments of degree of truth of statements (Oden, 1977a, 1977b), degree of agreement with statements (Dreyfuss, Kochen, Badre, & Robinson, 1975), judgments of how well objects exemplify concepts (Rosch, 1975), and estimates of the range of possibility (Hersh, Note 1), the fuzzy analogue to the statistical concept of a confidence interval.

Potentially, of course, any task provides a set of behavioral data which may be modeled by fuzzy set theory. However, not all tasks pertain to questions of whether individuals perceive environmental partial truth or have imprecise internal representations of truth or set membership. For example, Macvicar-Whelan (1978) inferred membership functions from confidence judgments on statements about height. Although inferences were made about the representation of height, the judgments had nothing at all to do with height; rather, they were measures of confidence. Individuals may be more or less confident about statements, even if all their perceptions and internal representations are precise. Fuzzy set theory may provide a model of confidence, but this experiment does not tell us how we understand sentences about height.

A similar problem occurs when membership functions are inferred from group response proportions. This method was used by Hersh and Caramazza (1976), who had subjects judge the applicability of statements about the sizes of various squares. The membership function was taken to be the proportion of 'yes' responses within the group. Rationale for this definition is Zadeh's (1968) assertion that the probability of an event is the expected value of its membership function. However, this assertion cannot be used to infer individual membership functions from group response probabilities. The probability merely provides a measure of the degree of consensus among respondents. Furthermore, additional assumptions which allow the inference of individual membership functions must deny individual differences in membership functions. A somewhat similar method was used by Thole, Zimmermann, and Zysno (1979), who inferred differences in set membership from group choice proportions.

Other investigations have been somewhat more direct in their measurement of individual membership functions.

However, since there seems to be a tendency for investigators to use personal measurement schemes, it is important to avoid overgeneralizing results of a particular scheme to all schemes. The reader is advised to pay careful attention to measurement techniques when considering the results presented below.

2.2 Experimental Findings

Researchers seeking some empirical verification for fuzzy set theory have been principally concerned with three questions:

1. Is it possible to obtain quantifiably imprecise judgments from individuals, or other judgments from which the degree of imprecision may be inferred?

2. Does individual understanding of the logical operations ('not,' 'and,' 'or,' etc.) correspond to the conjectures of Zadeh (1965) and others?

3. What function is served by linguistic hedges, such as the word 'very'?

It should not be at all surprising that individuals occasionally produce imprecise judgments. What is surprising is that this fact has been used as evidence for the "psychological reality of fuzzy sets." Macvicar-Whelan (1978), using confidence judgments, found imprecise boundaries between those squares judged "large," for example, and those not judged "large." Kochen and Badre (1974), and Dreyfuss et al. (1975) investigated the effect of context on imprecision in judgment, and found that context information tends to increase judgmental precision when context information is related in a sensible way to the judgment, while nonsensical context information decreases precision.

Zadeh (1965) has suggested the following representation for the logical functions negation, conjunction, and disjunction: The truth value of the negation of a statement is assumed to be the complement of the truth value of the statement; that is, the sum of the truth values of a statement and its negation is 1. The truth value of the conjunction of two statements is the minimum of the truth values of the individual statements. The truth value of the disjunction of two statements is the maximum of the truth values of the two statements. Alternative formulations have been suggested by others (Goguen, 1969), and no consensus has yet been reached as to which of these formulations is "correct."

Hersh and Caramazza (1976) have investigated negation by using group probabilities and confidence judgments to assess truth functions. The data they obtained seem to support Zadeh's conjecture. However, the results should be interpreted in light of the assessment method used. Oden (1977b) used functional measurement techniques (Anderson, 1974) to investigate the role of conjunction and disjunction in imprecise reasoning. The results indicate that conjunction may be represented by the product of individual truth functions, and disjunction by the sum of the truth values less their product. Judgments of truth for conjunctions and disjunctions directly contradicted the minimum and maximum functions suggested by Zadeh. On the other hand, Thole et al. (1979) found evidence supporting the minimum operation for set intersection.

The linguistic hedge 'very' may have two functions: it may serve to translate the boundary between true and false sentences; or it may serve as an intensifier, making judgments more precise. Kochen and Badre (1974) found evidence for the use of 'very' as an intensifier. Macvicar-Whelan (1978), and Hersh and Caramazza (1976), on the other hand, found that 'very' did not serve to increase precision.

The empirical literature presents an incomplete picture of individuals' usage of imprecise reasoning. In particular, there is some confusion about the types of tasks in which imprecise heuristics are used. There are also some conflicting results on certain aspects of imprecise reasoning. Most important, basic questions remain unanswered, the most fundamental of which is consistency. That is, do people respond to imprecision in a consistent and coherent manner? And by what rules are complex judgments derived from their simpler constituents? The question of consistency, therefore, is the principal concern of DDI's current investigation.

3.0 THE EXPERIMENT

The first experiment in DDI's research effort was designed to investigate the consistency with which individuals make judgments about the truth of sentences. Only individual results were used, because it is important to distinguish the true fuzziness inherent in a subject's knowledge and beliefs from the less interesting phenomenon of "noise" due to the aggregation of subjects whose perceptions do not coincide. Two methods of measurement were used: subjects first rated the relative truth of sentences in paired-comparison judgments; later, they made direct numerical estimates of the degree of truth attributed to each sentence. Consistency was tested with respect both to the transitivity of the paired comparisons, and to the compatability of those paired judgments with the direct scaling data.

In interpreting the results of this and subsequent experiments, the reader should be aware of the natural correspondence between truth value and set membership. In standard two-valued logic, a statement is either "true" or "false," just as in classical set theory an element is either a member or a nonmember of a set. Thus, the truth value of the statement, "x is an element of the set S," is by definition equivalent to the <u>degree</u> <u>of</u> <u>membership</u> of x in S. Similarly, the truth value of any sentence Y can be identified with the degree of membership of Y in the set of true sentences. When extending these classical two-valued notions to the realm of fuzzy set membership and multi-valued logic, we may simply replace the binary set {0,1} by the interval [0,1]. Thus, 0 represents the truth value of a "false" statement, or the degree of membership for an item which is completely excluded from a given set; 1 represents the truth value of a completely "true" sentence, or the degree of membership of an item which is unambiguously a member of a given set; and intermediate values represent sentences which are judged only partly true, or elements which are only partially included in a given set.

Just as "truth" (in the classical sense) and "probability" (in Bayesian statistics) may be viewed as subjective, primitive concepts, we may stipulate that "degree of truth" or "level of membership" is also such a concept. In other words, although there is no a priori way of objectively defining the "real" truth value of a sentence (other than self-contraditions or tautologies), we may productively use the concept of subjectively defined partial truth if we can verify that people deal with it (at least ideally) in the same way. In other words, we assume, in a manner entirely analogous to the classical logician or the Bayesian, that individual differences in assigned truth values represent differences in information states and in personal values, but not fundamental variations in the subjects' essential concept of "degree of truth."

3.1 Method

3.1.1 Subjects - Subjects were 11 male and 14 female employees of Decisions and Designs, Incorporated. Subjects were not familiar with the purpose of the experiment. The experiment took up to an hour and was conducted during office hours. The subjects received no pay other than their regular salary for their participation in the experiment.

3.1.2 Stimuli - Stimuli were a set of 50 sentences. Sentences were limited to those for which truth could be evaluated, and selected to cover as uniformly as possible the range from "true" to "false." Sentences were non-controversial and were written to avoid dependence on subjects' personal values or on special knowledge. Stimuli for any single subject were 20 sentences randomly chosen from the entire set. Each subject received a different set of sentences. A list of the 50 stimuli sentences appear in Appendix A.

3.1.3 Procedure - The experiment had two parts. In the first part of the experiment, subjects viewed all pairs of the 20 sentences, and indicated for each pair the sentence they felt was truer. Subjects were presented all 190 pairs for a single replication, along with the following instructions:

> Often in real life we encounter statements which seem neither entirely true nor entirely false. Given two of these statements, we may be able to say that one or the other is more true, although we could not say that either is absolutely true or absolutely false.
>
> On the following pages are several pairs of statements labeled A and B. For each pair, circle the letter of the statement you feel is more true.
>
> There are no right or wrong answers to these questions, and this test in no way measures your intelligence or personality. In fact, your responses will not be compared to the individual responses of any other subjects. There are a lot of questions; work quickly but carefully. Your help in this research is much appreciated.
>
> Remember, for each pair, circle the letter of the sentence you feel is more true. Please answer every question, even though it may be difficult on some problems for you to decide which letter to circle. When you finish all sentence pairs, continue with the brief second part of the experiment. If you have no questions, you may begin.

In the second part of the experiment, subjects individually estimated the truth of each sentence, assigning a number between 0 and 100 to each sentence in proportion to its perceived truth. Subjects received the following instructions:

For this part of the experiment, you will see the same sentences you saw before, now one at a time instead of in pairs.

Your task is to assign to each sentence a number between 0 and 100 in proportion to how true you think each sentence is. If you think the sentence is completely true, assign the number 100; if you think the sentence is completely false, assign the number 0; otherwise assign an intermediate number corresponding to the sentence's degree of truth. Put the number on the line to the left of each sentence. If you have no questions, you may begin.

Data consisted of paired-comparison judgments and direct estimates for each subject.

3.2 Results

3.2.1 Transitivity of paired-comparison judgments - For each subject's paired-comparison data, the number of intransitive triples was calculated. The distribution of the number of intransitive triples is presented in the stem-and-leaf display (Tukey, 1977) in Table 3-1. (Readers unfamiliar with stem-and-leaf displays can find a brief explanation in Appendix B.) The median number of intransitive triples is 17, and the first and third quartiles are given by 8 and 24, respectively. To help interpret these numbers, the maximum possible number of intransitive triples is 330, and the expected number given random choice on each pair would be 285. Clearly, there is quite a bit of regularity in the paired-comparison judgments.

Kendall (1955) has shown that under the assumption of random choice, the number of intransitive triples is distributed approximately χ^2, and has derived formulas for calculating a test statistic and appropriate degrees of freedom. The third column of Table 3-2 shows the value of χ^2 for each subject. Examination of this table shows that

0	2	3	4	4	5	5	8	8
10	1	3	5	5	7	7	8	9
20	3	4	4	5	8			
30	0	7	9					
•								
•								
•								
•								
110	7							

Table 3-1

STEM-AND-LEAF DISPLAY OF NUMBER OF INTRANSITIVE TRIPLES

Subject	Number of Intransitive Triples	χ^2 (Randomness) df ≈ 27	χ^2 (True-False) df ≈ 40	χ^2 (Three Group) df ≈ 75
1	6	166.5^t	113.3^t	127.0^t
2	15	162.0^t	101.3^t	100.0^f
3	5	167.0^t	114.7^t	130.0^t
4	39	150.0^t	69.3^h	28.0
5	18	160.5^t	97.3^t	91.0
6	28	155.5^t	84.0^t	61.0
7	23	158.0^t	90.7^t	76.0
8	6	166.5^t	113.3^t	127.0^t
9	15	162.0^t	101.3^t	100.0^f
10	8	165.5^t	110.7^t	121.0^t
11	24	157.5^t	89.3^t	73.0
12	2	168.5^t	118.7^t	139.0^t
13	17	161.0^t	98.7^t	94.0
14	24	157.5^t	89.3^t	73.0
15	17	161.0^t	98.7^t	94.0
16	30	154.5^t	81.3^t	55.0
17	37	151.0^t	72.0^h	34.0
18	3	168.0^t	117.3^t	136.0^t
19	11	164.0^t	106.7^t	112.0^h
20	19	160.0^t	96.0^t	88.0
21	8	165.5^t	110.7^t	121.0^t
22	117	111.0^t	−34.7	−206.0
23	25	157.0^t	88.0^t	70.0
24	5	167.0^t	114.7^t	130.0^t
25	13	163.0^t	104.0^t	106.0^f

$^f p < .05$ $^h p < .01$ $^t p < .001$

Table 3-2

CONSISTENCY MEASURES FOR PAIRED-COMPARISON DATA every subject had significantly ($p < .001$) fewer intransitive triples than would be predicted by chance.

A somewhat more interesting null hypothesis than that of random choice is the hypothesis that the set of sentences may be partitioned into two distinct subsets: "true" sentences and "false" sentences. Under this hypothesis, subjects will always say that a "true" statement is truer than a "false" statement, but will choose randomly between pairs of "true" statements or pairs of "false" statements. The expected number of intransitive triples under this hypothesis is smaller than under the hypothesis of random choice. Furthermore, the expected number depends on the number of sentences in the true and false sets. The expected number is lowest when half of the sentences are true and half are false, so this case gives, in some sense, the most stringent test of consistency compared to the true-false hypothesis. In the above case, it is possible to modify the test statistic to test the true-false hypotheses. χ^2 values are presented in the fourth column of Table 3-2. The expected and maximum number of intransitivities under the true-false hypothesis are 60 and 80, respectively. As illustrated in Table 3-2, all subjects but one had significantly fewer ($p < .01$) violations than would be expected under the true-false hypothesis. The remaining subject had more violations of intransitivity than would be possible under the true-false hypothesis.

It may be plausible to hypothesize that sentences are divided into three groups: "true" sentences, "false" sentences, and those sentences which are neither true nor false or for some reason cannot be evaluated as true or false. It is possible to derive a statistic to test this three-group hypothesis, but its approximation by a χ^2 distribution is only fair. Nevertheless, the fifth column of Table 3-2 shows that twelve of the twenty-five subjects had significantly fewer ($p < .05$) intransitive triples than would be predicted by the three-group hypothesis. These independent χ^2 variables may be summed to obtain an overall indication of consistency. The obtained overall χ^2 of 2080 is highly significant (df = 1860, $p < .001$).

To summarize, the consistency of paired comparison responses was much greater than would be expected by either random choice, or a two-group hypothesis. For twelve of the subjects, the three-group hypothesis could also be rejected; furthermore, the three-group hypothesis could also be rejected when individual data were aggregated over all subjects.

Another way of assessing the level of consistency is by estimating the probability of making an error on a pair which would produce the obtained rate of inconsistency. That is, suppose subjects were transitive, except that they reversed their choice randomly with probability p. Then, the probability of an intransitive triple, p_i is given by:

$$p_i = p(1 - p). \qquad [1]$$

If the obtained proportion of intransitive triples is p_o, then p may be estimated by:

$$\hat{p} = \frac{1 - \sqrt{1 - 4p_o}}{2}. \qquad [2]$$

Note that the expected proportion of intransitive triples reaches its maximum when p is 1/2.

The obtained distribution of $\hat{\underline{p}}$ is displayed in the stem-and-leaf display in Table 3-3. This distribution ranges from 0.18% to 11.61% and has a median of 1.51%. For all subjects except one, the inferred probability of paired comparison reversal is less then 3.6%. Clearly, the deviations from transitivity are minor, and consistent with a reasonable level of experimental error.

.00	2	3	4	4	5	5	7	7
.01	0	2	3	3	5	5	6	7
.02	1	2	2	2	5	7		
.03	4	6						
.04								
⋮								
.11	6							

Table 3-3

STEM-AND-LEAF DISPLAY OF INFERRED PROBABILITY OF PAIR REVERSAL

3.2.2 <u>Categorical perception of truth</u> - It seems clear that the subjects consistently respond to the truth differences between sentences. However, the sentences may be perceived to be either true or false. If this is the case, then subjects may have more difficulty distinguishing the relative truth of two sentences perceived as true or perceived as false, than distinguishing the relative truth of a true and false statement. This type of result has been used as evidence for categorical perception of speech sounds (Liberman, Harris, Hoffman, & Griffith, 1957); sounds which are placed in different categories are much easier to distinguish than sounds placed in the same category, even if the sounds in different categories are much more similar on some physical dimension than the sounds in the same category.

A fundamental difference between the sentences used in this investigation and computer-generated speech sounds is that there is no environmental measure of partial truth. However, it may seem reasonable to use the direct truth ratings as the objective measure of truth. Since the direct ratings were on a scale of 0 to 100, statements rated greater than 50 were interpreted as "true" statements; those rated less than 50 were interpreted as "false" statements. The statement pairs were then divided into four groups. The "true" group included those pairs in which both statements were "true," or one was "true" and the other was rated 50. The "false" group included those pairs in which both statements were "false," or one was "false" and one was rated 50. The mixed group included those pairs consisting of a "true" and a "false" sentence. The remaining pairs, consisting of two sentences rated 50, were not analyzed further (there were 65 such pairs, comprising 1.37% of the total sample).

For each group, the probability of choosing the sentence with the higher directly estimated truth value was calculated as a function of the difference in the estimated truth values. Direct estimate differences were grouped by tens so that groups consisted of pairs with differences between one and ten, eleven and twenty, and so forth. If perception of truth is categorical, the proportion of choices consistent with direct estimates should be greater for the mixed group than for the true or false group for any scale difference.

The results of this analysis are displayed in Figure 3-1. The obtained proportions are displayed along with 95% confidence intervals for the actual probability. Examination of Figure 3-1 shows that the probability of a choice consistent with the direct estimates indeed increases when the difference in the estimates increases. There appear to be no differences between the mixed group and the other two groups. An analysis of variance of the transformed proportions for the range of scale differences from one to fifty confirms what seems obvious from Figure 3-1. The effect of difference in scale value was large and significant ($F(4,8) = 9.99$, $p < .01$), accounting for 81% of the total sum of squares. The effect of pair group was not significant ($F(2,8) = 0.67$, NS), accounted for only 2.7% of the total sum of squares. Thus, there is no evidence for categorical perception of truth.

3.2.3 <u>Consistency of paired-comparison choices and direct estimates</u> - The results described in the previous section suggest that there is substantial agreement between the two methods of assessing partial truth. To check this directly, the product-moment correlation was calculated for each subject between the number of statements a sentence was judged truer than in the paired judgments and the sentence's directly estimated truth value. A stem-and-leaf display of the distribution of correlations is shown in Table 3-4. The correlations ranged from 0.737 to 0.977, with a median of 0.900. Clearly, there is almost complete agreement between the two assessment methods.

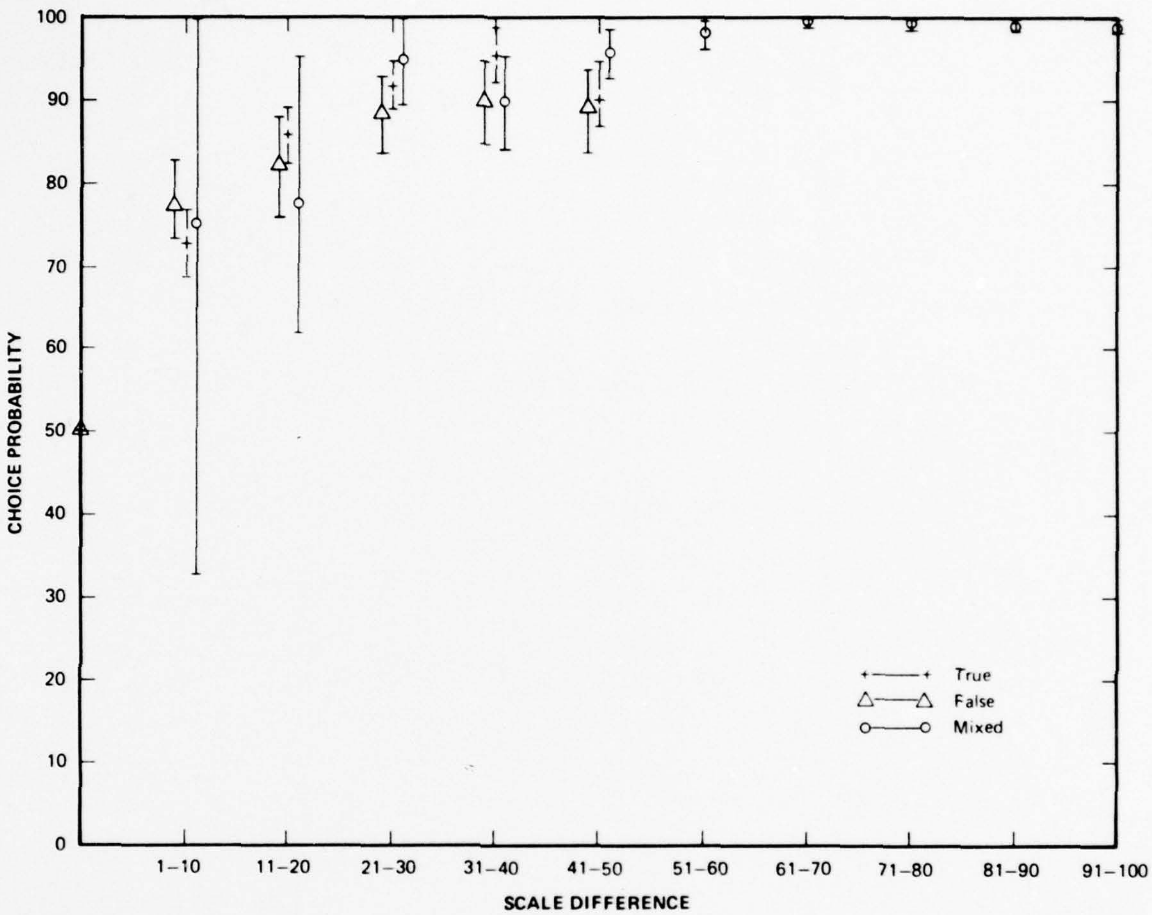


Figure 3-1

TEST OF CATEGORICAL PERCEPTION OF TRUTH

.700 – .749	37, 41
.750 – .799	76
.800 – .849	20, 46
.850 – .899	55, 58, 87, 88, 93, 94, 97
.900 – .949	00, 01, 02, 10, 12, 16, 19, 37, 38
.950 – .999	54, 59, 70, 77

Table 3-4

STEM-AND-LEAF DISPLAY OF CORRELATION BETWEEN PAIRED-COMPARISON SCORE AND DIRECT TRUTH ESTIMATE

3.3 Discussion

In order to represent judgments of relative truth by an ordinal scale, it is necessary that these judgments be connected and transitive. The data seem unequivocal on transitivity. There were far fewer violations of transitivity than would be expected if subjects responded to sentences as though they were either true or false. The results also seem to indicate that individuals perceived the truth of the statements in a continuous rather than a categorical manner. Furthermore, two measurement techniques gave highly similar truth scales. Although the results seem clear, attention should be paid to limitations in the experiment.

There are two differences between the present analysis of categorical perception and those which have been done in other areas (e.g., Liberman et al., 1957). First, in the current analysis, there is no objective measure of partial truth. Since two subjective measures are being compared, both may be perceived the same way, and no categorization effect will be found, independent of the nature of subjects' perceptions. Second, categorical perception has been tested in a single context only. In the perception of speech sounds, the main strength of the findings that perception is categorical comes from the fact that in a different context, when individuals employ a physical rather than a linguistic interpretation of the sounds, perception is continuous. The existence of contextual dependencies corrects for possible confounding effects of nonlinear subjective scales. This result suggests that although perception of truth may be continuous in the context in which it was analyzed, it may be that categorical perception would occur in other contexts, such as in more complex judgment and decision-making tasks. Further research in this direction may serve to illuminate this problem.

It must be realized that both the task used and the type of consistency tested are somewhat limiting. Behavior in a simple paired-comparison task will not necessarily reflect the same processes as the behavior in a more complex decision-making task. Furthermore, the statements themselves were simple. A much stronger test of consistency would involve sentences formed by joining two or more simple statements with appropriate logical connectives. Tests of this stronger form of consistency would give knowledge of the processes by which the truth of combinations of statements is evaluated, a stronger representation of partial truth, as well as further verifying the validity of partial truth in human reasoning.

4.0 OUTLINE OF FUTURE RESEARCH PLANS

The encouraging results of the experiment described in this report indicate that further investigation of the psychological foundations underlying fuzzy set theory would be of value. DDI has designed a second experiment to extend the domain of application beyond simple sentences, and to test for scale properties stronger than the ordinal ranking demonstrated here. Furthermore, we envision a broader program of research based on the experiments performed under the current contract. The goal of such a program would be a complete, well-validated body of empirical evidence which could be used to justify the application of fuzzy set theory to human decision problems, and to determine appropriate methods for implementing fuzzy set techniques.

4.1 Plans for Experiment Number 2

Our second experiment will extend the domain of partial truth to include compound sentences formed by joining two sentences with logical operations, such as conjuction and disjunction. In the experiment, conjoint measurement techniques will be used to test the ordinal implications of alternative mathematical formulations of the logical operations. If the assumptions are satisfied, then a scale of truth can be determined which is consistent with a particular formulation of the logical operation; this scale may also have stronger uniqueness properties than the ordinal scale determined in the previous experiment. In addition, the scale obtained indirectly will be compared with direct estimates of truth.

The experiment will consist of two parts. In the first part, subjects will rank order compound sentences according to their perceived truth. Examples of these sentences are

"Ohio is in the midwest, and a porpoise is a typical mammal," and "Neither 'A man 5'11" tall is very tall,' nor 'Many stars are visible in the city sky at night,' is true." The sentences will form factorial designs which allow for testing using conjoint measurement. Product and minimum rules will be tested for conjunctions; inverse product and maximum rules will be tested for disjunctions.

In the second part of the experiment, direct estimates will be obtained of the truth of the simple and compound sentences ordered in the first part. Consistency of these estimates with the alternative combination rules and with the scale obtained from the rank order will be tested.

Analysis will concentrate on answering the following three questions: 1) Can a representation of truth of compound sentences be established which is consistent with a reasonable and coherent set of combination rules? 2) Can a scale be obtained for partial truth with ordered metric or stronger uniqueness? 3) Do direct estimates provide a reasonable scaling technique? The techniques of conjoint measurement have been extended in two ways to answer these questions. First, methods have been developed to test whether a rank order on a factorial design is consistent with minimum or maximum combination rules. Second, techniques are being developed to determine whether two orderings (i.e., on conjunction and disjunction) are simultaneously consistent with related combination rules (e.g., minimum and maximum). These techniques will be used along with other appropriate methods to seek the answers to the questions stated above.

4.2 Long-Range Planning and Evaluation

Based on the results of our first experiment, we can envision a number of directions for fruitful extensions of the currently planned research. Roughly, these can be divided into two categories: theoretical advances (including both mathematical and empirical research), and applications.

Theoretical advances which might be accomplished in a long-term research effort might include the following:

- o axiomatization and empirical testing of advanced logical and arithmetic operations (such as quantification, fully developed predicate calculus, and logical "types" or "categories" which may be thought of as an extension of "sets");

- o development of a standard set of replicable methods which may be used as operational definitions of the concepts of fuzzy set theory, and employed to assure consistency of results; and

- o extensions of fuzzy set theory itself to apply to unsolved problems in other areas of mathematics, engineering, or decision theory.

Among the applications which could be productively developed, two which have been previously investigated by DDI are fuzzy decision analysis (Watson, Weiss, and Donnell, 1979) and fuzzy policy capturing (Weiss, 1978). The preliminary investigations in both of these areas have indicated the potential for useful applications to practical problems, and demonstrated prototype systems for their implementation. However, before such methods can be applied with the confidence that now characterizes the probability and utility-theoretic models used in decision analysis, much refinement, testing, and implementation of methods remains to be done. Furthermore, once fuzzy-set-based decision aiding techniques have been developed, it is expected that the experience gained during the first applications will provide the impetus for both methodological improvements and further theoretical research.

Further extensions of fuzzy set applications might include a number of other highly important fields, of which the following are a representative sample:

- consensus formation, group decision making, and multi-participant negotiation;

- pattern recognition by machine, using visual and linguistic information (e.g., recognition of faces from a verbal description);

- natural language processing (e.g., translation, data-base access by linguistic interaction, fuzzy "programming"); and

- microprocessor control of complex systems under multiple objectives (e.g., automobile engines regulated to combine fuel economy, emission control, and power response).

Each of these would, of course, represent a full-scale project of its own, but all would be based largely upon the fundamental research performed within the current project.

To summarize, our preliminary experimental investigations into the empirical basis for fuzzy set theory have been encouraging. We intend to continue our studies of the fundamental operations underlying fuzzy set theory and its reasoning, and to extend such investigations to apply to more phenomena than the simple sentences of the first experiment. Because of the encouraging nature of our results, it seems very likely that a consistent operational system can eventually be defined and validated, to serve as the basis for a fuzzy technology of reasoning and arithmetic, with productive applications to decision theory and related problem areas.

REFERENCES

Anderson, N. H. Information integration theory: a brief survey. In D. H. Krantz, R. C. Atkinson, R. D. Luce, and P. Suppes (Eds.), Contemporary Development in Mathematical Psychology (Vol. 2). San Francisco: Freeman, 1974.

Dreyfuss, G. R., Kochen, M., Robinson, J., & Badre, A. N. On the psycholinguistic reality of fuzzy sets. In R. E. Grossman, L. J. San, & T. J. Vance (Eds.) Functionalism. Chicago: University of Chicago Press, 1975, pp. 135-149.

Goguen, J. A. The logic of inexact concepts. Synthese, 1969, 19, 325-373.

Hersh, H. M. & Caramazza, A. A fuzzy set approach to modifiers and vagueness in natural language. Journal of Experimental Psychology: General, 1976, 105, 254-276.

Kendall, M. G. Rank Correlation Methods. New York: Hafner Publishing Company, 1955.

Kochen, M. & Badre, A. N. On the precision of adjectives which denote fuzzy sets. Journal of Cybernetics, 1974, 4, 49-59.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. The discrimination of speech sounds within and across phoneme boundaries. Journal of Experimental Psychology, 1957, 54, 358-368.

Macvicar-Whelan, P. J. Fuzzy sets, the concept of height, and the hedge "very". IEEE Transactions on Systems, Man, and Cybernetics, 1978, SMC-8, 507-511.

Oden, G. C. Fuzziness in semantic memory: Choosing exemplars of subjective categories. Memory and Cognition, 1977, 5, 198-204. (a)

Oden, G. C. Integration of fuzzy logical information. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 565-575. (b)

Rosch, E. H. Cognitive representation of semantic categories. Journal of Experimental Psychology: General, 1975, 104, 192-233.

Thole, U.; Zimmermann, H. J.; & Zysno, P. On the suitability of minimum and product operators for the intersection of fuzzy sets. Fuzzy Sets and Systems, 1979, 2, 167-180.

Tukey, J. W. Exploratory Data Analysis. Reading, Massachusetts: Addison-Wesley Publishing Company, 1977.

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

Weiss, J. J. Fuzzy Policy Capturing as a Decision Aid (Interim Report PR 78-30-83). McLean, VA: Decisions and Designs, Inc., 1978.

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

Zadeh, L. A. Probability measures of fuzzy events. Journal of Mathematical Analysis and Applications, 1968, 23, 421-427.

Reference Note 1

Hersh, H. M. Fuzzy reasoning: The integration of vague information. Unpublished manuscript, 1978. (Available from Harry M. Hersh, Digital Equipment Corp., 146 Main Street, ML3-2/E41, Maynard, MA 01754.)

CONTRACT DISTRIBUTION LIST

Director Engineering Psychology Programs Code 455 Office of Naval Research 800 North Quincy Street Arlington, VA 22217	5 copies
DCASMA Baltimore Hampton Plaza Building, Room 200 300 East Joppa Road Towson, MD 21204	1 copy
Director Naval Research Laboratory Technical Information Division Code 2627 Washington, DC 20375	6 copies
Defense Documentation Center Cameron Station, Bldg. 5 Alexandria, VA 22314	12 copies
Commanding Officer ONR Branch Office Attn: Dr. J. Lester Building 114, Section D 666 Summer Street Boston, MA 02210	1 copy

SUPPLEMENTAL DISTRIBUTION LIST

CDR Paul R. Chatelier
Military Assistant for Training and Personnel Technology
Office of the Deputy Under Secretary of Defense
OUSDRE (E&LS)
Pentagon, Room 3D129
Washington, D.C. 20301

Director
Analysis and Support Division
Code 230
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

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

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

Commanding Officer
ONR Branch Office
Attn: Dr. C. Davis
536 South Clark Street
Chicago, IL 60605

Director
Statistics and Probability Program
Code 436
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Commanding Officer
ONR Branch Office
Attn: Dr. E. Gloye
1030 East Green Street
Pasadena, CA 91106

Office of Naval Research
Scientific Liaison Group
American Embassy, Room A-407
APO San Francisco 96503

Dr. Robert G. Smith
Office of the Chief of Naval Operations, OP987H
Personnel Logistics Plans
Washington, DC 20350

Dr. Alfred F. Smode
Training Analysis and Evaluation Group
Naval Training Equipment Center
Code N-OOT
Orlando, FL 32813

Dr. Gary Poock
Operations Research Department
Naval Postgraduate School
Monterey, CA 93940

Dean of Research Administration
Naval Postgraduate School
Monterey, CA 93940

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

Navy Personnel Research and Development Center
Code 305
San Diego, CA 92152

Navy Personnel Research and Development Center
Management Support Department
Code 210
San Diego, CA 92152

Mr. J. Barber
HQS, Department of the Army
DAPE-PBR
Washington, DC 20546

Dr. Joseph Zeidner
Technical Director
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Director
Organizations and Systems Research Laboratory
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Edgar M. Johnson
Organizations and Systems Research Laboratory
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

U.S. Air Force of Scientific Research
Life Sciences Directorate, NL
Bolling Air Force Base
Washington, DC 20332

Dr. Donald A. Topmiller
Chief, Systems Engineering Branch
Human Engineering Division
USAF AMRL/HES
Wright-Patterson AFB, OH 45433

Air University Library
Maxwell Air Force Base, AL 36112

Dr. Gordon Eckstrand
AFHRL/ASM
Wright-Patterson AFB, OH 45433

Dr. Stephen J. Andriole
Director
Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Blvd.
Arlington, VA 22209

Dr. Judith Daly
Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Blvd.
Arlington, VA 22209

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

Dr. Meredith P. Crawford
Department of Engineering Administration
George Washington University
Suite 805
2101 L Street, NW
Washington, DC 20037

Dr. Charles Gettys
Department of Psychology
University of Oklahoma
455 West Lindsey
Norman, OK 73069

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

Dr. Kenneth Hammond
Institute of Behavioral Science
University of Colorado
Room 201
Boulder, CO 80309

Dr. Ronald Howard
Department of Engineering-Economic Systems
Stanford University
Stanford, CA 94305

Dr. William Howell
Department of Psychology
Rice University
Houston, TX 77001

Journal Supplement Abstract Service
American Psychological Association
1200 17th Street, NW
Washington, DC 20036 (3 copies)

Dr. Robert R. Mackie
Human Factors Research, Inc.
5775 Dawson Avenue
Goleta, CA 93017

Dr. Gary McClelland
Institute of Behavioral Sciences
University of Colorado
Boulder, CO 80309

Dr. Miley Merkhofer
Stanford Research Institute
Decision Analysis Group
Menlo Park, CA 94025

Dr. Terence R. Mitchell
University of Washington
Seattle, WA 98195

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

Professor Judea Pearl
Engineering Systems Department
University of California-Los Angeles
405 Hilgard Avenue
Los Angeles, CA 90024

Professor Howard Raiffa
Graduate School of Business
Administration
Harvard University
Soldiers Field Road
Boston, MA 02163

Dr. Paul Slovic
Decision Research
1201 Oak Street
Eugene, OR 97401

Dr. Amos Tversky
Department of Psychology
Stanford University
Stanford, CA 94305

Dr. Gershon Weltman
Perceptronics, Inc.
6271 Variel Avenue
Woodland Hills, CA 91364

Dr. David Dianich
Chairman, Department of
Business and Economics
Salisbury State College
Salisbury, MD 21801

North East London Polytechnic
The Charles Myers Library
Livingstone Road
Stratford
London E15 2LJ
UNITED KINGDOM

Director, Human Factors Wing
Defence & Civil Institute of
Environmental Medicine
Post Office Box 2000
Downsville, Toronto, Ontario
CANADA

Dr. A. D. Baddeley
Director, Applied Psychology Unit
Medical Research Council
15 Chaucer Road
Cambridge, CB2 2EF
UNITED KINGDOM

Dr. J. Baal Schem
Acting Director
Tel-Aviv University
Interdisciplinary Center for Technological
Analysis and Forecasting
Ramat-Aviv, Tel-Aviv
ISRAEL

APPENDIX A

LIST OF STIMULUS SENTENCES

LIST OF STIMULUS SENTENCES

1. Ohio is in the Midwest.
2. It's a long way from Pennsylvania to Canada.
3. Physicians make a lot of money.
4. A penguin is like a fish.
5. California is in the northern United States.
6. Four inches is a heavy snowfall.
7. A porpoise is a typical mammal.
8. Four pancakes is a lot for a lumberjack's breakfast.
9. A man five feet, eleven inches tall is very tall.
10. Mathematics is a difficult subject.
11. It is inappropriate to be barefoot in a restaurant.
12. Shoplifting is a serious crime.
13. Mexican food is very spicy.
14. Jockeys are very short.
15. President Carter has a strong accent.
16. Hamsters are popular pets.
17. It is uncomfortable to wear your right shoe on your left foot.
18. An umbrella is effective in keeping one dry in the rain.
19. Many stars are visible in the city sky at night.
20. It is not expensive to call someone in Europe from the United States.
21. Most Americans are overweight.
22. Cheesecake is very sweet.
23. Athletes are overpaid.

24. Alcohol excites people.

25. Ringo Starr (of the Beatles) is a great drummer.

26. It is unsafe to live in large American cities.

27. Bob Hope is very funny.

28. The Internal Revenue Service (IRS) is reasonably fair.

29. The United Nations is an effective organization.

30. Most lawyers act as though they have no conscience.

31. Los Angeles has clean air.

32. Chevrolet makes excellent cars.

33. White lies do no harm.

34. Medicine tastes terrible.

35. Women and men are treated equivalently in our society.

36. Greece is a nice place to go for a vacation.

37. Lake Erie is a large body of water.

38. The Olympics is an example of international cooperation.

39. Rhode Island has a lot of people.

40. West Germany is a highly industrialized country.

41. Fish have a strong odor.

42. The color of goldfish is yellow.

43. Skiing is a dangerous sport.

44. Tennis balls are fuzzy.

45. A lemon is a large fruit.

46. Smoking tobacco is beneficial to one's health.

47. Alaska is north of Canada.

48. A 35-year-old man is middle-aged.

49. Denver is near the Pacific Ocean.

50. Potato chips are salty.

APPENDIX B

A BRIEF EXPLANATION OF "STEM-AND-LEAF" DIAGRAMS

A BRIEF EXPLANATION OF "STEM-AND-LEAF" DIAGRAMS

The "stem-and-leaf" diagram is a recently developed means of displaying raw or calculated data in a manner which permits either rapid visual scanning or detailed study. It is essentially a kind of histogram, where equal intervals are labeled along the left-hand margin, and each occurrence of a value within a given interval corresponds to one token on the appropriate line of the main figure. Typically, the left-margin label will be either the interval's range or the lower bound of that range (choice of scale is at the pleasure of the user, but is typically designed to segment the sample into groups of no more than ten or fifteen). The tokens themselves are the least significant digits of the corresponding numbers. For example, to represent the set of observations (2, 4, 7, 7, 12, 15, 17, 19, 21, 28, 43), we might use the five intervals [0-9], [10-19], [20-29], [30-39], and [40-49]. Thus, the stem-and-leaf diagram would look like this:

0	2	4	7	7
10	2	5	7	9
20	1	8		
30				
40	3			

The reader can then read the diagram either as a gross histogram of the scores (4 in the 0-9 interval, 4 in the 10-19 interval, etc.), or else look in more detail at the specific scores (for example, the entry "7" on the line labeled "10" corresponds to the observation "17" that was listed in the original sample.

As a further example, suppose you must represent given the scores: .0037, .0042, .0059, .0059, .0062, .0064, .0076, .0078, .0091, .0099, .0120, .0145, .0187, .0277, .0291. The stem-and-leaf representation of this set of data might look like this:

.0000-.0049	37	42						
.0050-.0099	59	59	62	64	76	78	91	99
.0100-.0149	20	45						
.0150-.0199	87							
.0200-.0249								
.0250-.0299	77	91						

Here again, the essential shape of the distribution may be read simply and directly, while each individual data item is also represented with no loss of information (for example, the entry "91" on the ".0050-.0099" line represents the observed value of .0091.

This simple but useful format is gaining popularity as a means of displaying data, and the reader should find it worth the small effort it takes to get comfortably acquainted with it.