ARI Research Note 87-68

MEASURES OF SIMILARITY BETWEEN FUZZY CONCEPTS: A COMPARATIVE ANALYSIS

Rami Zwick, Edward Carlstein
and David Budescu
Thurstone Psychometric Laboratory

for

Contracting Officer's Representative
Michael Drillings

BASIC RESEARCH LABORATORY
Michael Kaplan, Director

U. S. Army

Research Institute for the Behavioral and Social Sciences

December 1987

U. S. ARMY RESEARCH INSTITUTE

FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the

Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

WM. DARRYL HENDERSON
COL, IN
Commanding

Research accomplished under contract
for the Department of the Army

University of North Carolina

Technical review by

Dan Ragland

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARI Research Note 87-68	2. GOVT ACCESSION NO. A187435	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* Measures of Similarity between Fuzzy Concepts: A Comparative Analysis		5. TYPE OF REPORT & PERIOD COVERED Interim Report August 85 - August 86 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Rami Zwick, Edward Carlstein, and David Budescu		8. CONTRACT OR GRANT NUMBER(s) MDA903-83-K-0347
9. PERFORMING ORGANIZATION NAME AND ADDRESS Thurstone Psychometric Laboratory Davie Hall, 013-A, University of North Carolina Chapel Hill, North Carolina 27514		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2Q161102B74F
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Institute for the Behavioral and Social Sciences, 5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600		12. REPORT DATE December 1987 13. NUMBER OF PAGES 33
14. MONITORING AGENCY NAME & ADDRESS(*if different from Controlling Office*) - -		15. SECURITY CLASS. *(of this report)* Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE n/a

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

Michael Drillings, contracting officer's representative

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

Fuzzy Set Theory, Cognition.
Similarity Judgement,
Understanding,

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

This research note discusses the selection of an appropriate index to measure the similarity between fuzzy subsets (concepts). The RN reviews 19 measures of linguistic approximation procedures, and compares their performance in a behavioral experiment. For categorizing pairs of fuzzy concepts as "similar" or "dissimilar" all measures performed well. For distinguishing degrees of similarity or dissimilarity, however, certain measures were superior. The best measures focus on only one "slice" of membership function.

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

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

INTRODUCTION

Giles [10] has described the current character of research in fuzzy reasoning as follows:

> "A prominent feature of most of the work in fuzzy reasoning is its ad hoc nature. ...If fuzzy reasoning were simply a mathematical theory there would be no harm in adopting this approach; ...However, fuzzy reasoning is essentially a practical subject. Its function is to assist the decision-maker in a real world situation, and for this purpose the practical meaning of the concepts involved is of vital importance" (Giles [10], pp. 263).

Fuzzy set theory would benefit from becoming a behavioral science, having its assumptions validated, and having its results verified by empirical findings (Kochen [19]). Unfortunately, not too many researchers in this field have adopted this philosophy. Except for a few major experimental works (Hersh et. al. [13,14]; Oden [23,24,25,26]; Wallsten et al. [31,32,33]; Rapoport et al. [28]; Zimmer [36]; Kochen [19]; Zysno [37]), most other works center around axiomatic treatment of the subject. As noted by Zeleny [35], it happens many times that as the axiomatic precision of a theory increases, it becomes less significant and relevant in its impact on the practice of human decision making and judgment.

This paper addresses this problem in the area of linguistic approximation- specifically, we focus on the question of selecting an appropriate index for measuring the similarity between fuzzy subsets. Several methods have been suggested for the process of linguistic approximation (Bonissone [2]; Eshragh and Mamdani [9]; Wenstop [34]). Each of them suggested a different measure of similarity, and each suffers from the same

ad hoc nature as most of the other work in this field. There is no serious attempt to validate the techniques through behavioral experiments. Some authors have mentioned that their techniques work very well, but did not provide the appropriate data to support their claim. For example, Bonissone [2] in his pattern recognition approach to linguistic approximation writes "This new distance reflects very well the semantic distance among fuzzy sets... This distance has been applied in the implementation and has provided very good results," but no results are reported, and it is not clear what criterion are used to make such a statement. Similarly, no serious attempts have been made by Wenstop [34] to validate details of his semantic model. Neither do Eshragh and Mamdani [9] behaviorally validate their approach, although they claim that "... the results obtained from 'LAM5' are quite encouraging and also considering the number of previous attempts and difficulties involved, one can say that 'LAM5' has proved workable," but again no supporting data were supplied. More importantly, no attempt has been made to compare the performances of the various different indexes of distance that could be used in these applications.

Overall, the lack of behavioral validation for any similarity index is disturbing because of (i) the crucial role (translation) that this index plays in any implementation of fuzzy reasoning theory and, (ii) the relative ease by which any proposed index may be validated. Regarding the second point, any successful distance measure should be able to account for and predict a subject's similarity judgment between fuzzy concepts, based on his separate membership functions of each concept.

The notion of similarity plays a fundamental role in theories of knowledge and behavior, and was dealt with broadly in the psychology literature (Gregson [12]). Overall, the theoretical analysis of similarity relations has been dominated by the geometric model. The geometric models represent objects as points in some coordinate space such that the observed dissimilarity between objects corresponds to the metric distance between the respective points.

The similarity indexes used in the linguistic approximation techniques adopted this approach. Bonissone [2] located each concept initially in four dimensional space, where the dimensions are : Power, Entropy, First moment, and Skewness of the membership function, and defined the distance between two concepts as the regular weighted Euclidean distance between the points representing these concepts. Wenstop [34] located the concepts in a two dimensional space. The two dimensions are : location (center of gravity) and imprecision (fuzzy scalar cardinality) of the membership function. The distance between any two concepts in this space is the regular Euclidean distance. The same geometrical distance philosophy was adopted by Eshragh and Mamdani [9] and by Kacprzyk [16].

Most conclusions regarding the appropriate distance metric have been based on studies using judgment of similarity between stimuli which can be located a-priori along (objectively) distinguishable dimensions (such as colors, tones, etc.). The question of integral vs. separable dimensions is crucial. Separable dimensions remain subjectively distinct when in

combination. By contrast, integral dimensions combine into a subjectively nondecomposable whole. There is an extensive literature supporting the idea that the Euclidean metric may be appropriate for describing psychological distance relationships among integral-dimensions stimuli, while something more along the lines of the city-block metric is appropriate for separable-dimensions stimuli (Attneave [1]).

As noted by Tversky [30], both dimensional and metric assumptions are open to questions. It has been argued that dimensional representations are appropriate for certain stimuli (those with a-priori objective dimensions), but for others, such as faces, countries, and personality, a list of qualitative features is appropriate. Hence the assessment of similarity may be better described as a comparison of features rather than as a computation of metric distance between points. Furthermore, various studies demonstrated problems with the metric assumption. Tversky [30] showed that similarity may not be a symmetric relation (violating the symmetry axiom of a metric), and also suggested that all stimuli may not be equally similar to themselves (violating the minimality axiom). Therefore, similarity may be better modeled by a function which is not conceptually a geometric distance (e.g. a set-theoretic function instead).

The plan for the rest of this paper is as follows: We first review the various distance indexes suggested in the literature, in the general case, and as adapted to fuzzy sets. Second, our experimental design will be presented. Finally we will discuss

the results and the implications of the results to the process of linguistic approximation.

Geometric Distance Models.

A particular class of distance functions that has been investigated by psychologists is known as the Minkowski r-metric. This metric is a one-parameter class of distance functions defined as follows:

$$d_r(x,y)=[\sum_{i=1}^{n} |x_i-y_i|^r]^{1/r}, \quad r\geq 1 \tag{1}$$

where x and y are two points in an n-dimensional space with components (x_i,y_i) i=1,2,...,n. Let us consider some special cases that are of particular interest. Clearly, the familiar Euclidean metric is the special case of r=2. The other special cases of interest are r=1 and r=ω. The case of r=1 is known as the "city block" model. As r approaches ω, Eq. (1) approaches the "dominance metric" in which the distance between stimuli x and y is determined by the difference between coordinates along only one dimension-that dimension for which the value $|x_i-y_i|$ is greatest. That is:

$$d_\omega(x,y)=\max_i |x_i-y_i|. \tag{1.1}$$

Each of the three distance functions, r=1,2, and ω, are used in psychological theory (Hull [15]; Restle [29]; Lashley [20]).

Generalizing the geometric distance models to fuzzy subsets.

Let E be a set and let A and B be two fuzzy subsets of E. Define the following family of distance measures between A and B:

$$d_r(A,B)=\left(\sum_{i=1}^{n} |\mu_A(x_i)-\mu_B(x_i)|^r/n\right)^{1/r} \quad r\geq 1, \quad (1.2)$$

or, if E=R

$$d_r(A,B)=\left(\int_{-\omega}^{+\omega} |\mu_A(x)-\mu_B(x)|^r dx\right)^{1/r} \quad r\geq 1, \quad (1.3)$$

and $$d_\omega(A,B)=\sup_x |\mu_A(x)-\mu_B(x)|. \quad (1.4)$$

The cases r=1 and 2 were studied by Kaufman [17]. Kacprzyk [16] proposed the distance measure $(d_2)^2$, and d_ω was proposed by Nowakowska [22]. Our empirical evaluation will consider d_1, d_2, $(d_2)^2$, and d_ω.

Hausdorff metric.

The Hausdorff metric is a generalization of the distance between two points in a metric space to two compact nonempty subsets of the space. If U and V are such compact nonempty sets of real numbers, then the Hausdorff distance is defined by:

$$q(U,V)=\max\{\sup_{v\in V}\inf_{u\in U} d_2(u,v), \sup_{u\in U}\inf_{v\in V} d_2(u,v)\} \quad (2)$$

where d_2 is as defined in Eq. (1).

In the case of real intervals A and B, the Hausdorff metric

is described by:

$$q(A,B)=\max\{|a_1-b_1|,|a_2-b_2|\} \qquad (2.1)$$

where $A=[a_1,a_2]$ and $B=[b_1,b_2]$.

Generalizing the Hausdorff metric to fuzzy subsets.

Let $F(\mathbb{R})$ be the set of all fuzzy subsets of the real line. There is no unique metric in $F(\mathbb{R})$ which extends the Hausdorff distance. Ralescu and Ralescu [27] proposed the following generalizations:

$$q_1(A,B)=\int_{\alpha=0}^{1} q(A_\alpha,B_\alpha)\,d\alpha \qquad (2.2)$$

$$q_\infty(A,B)=\sup_{\alpha>0} q(A_\alpha,B_\alpha). \qquad (2.3)$$

where A_α is the α-level-set of the fuzzy set A.
We propose the Hausdorff distance between the intervals with the highest membership grade:

$$q_*(A,B)=q(A_{1.0},B_{1.0}). \qquad (2.4)$$

If A and B are real intervals then

$$q_1(A,B)=q_\infty(A,B)=q_*(A,B)=q(A,B).$$

Goetschel and Voxman [11] suggested a different generalization of the Hausdorff metric. Let A and B be two fuzzy numbers (for the exact definition of fuzzy numbers in this context, which is slightly different from the usual definition, see Goetschel and Voxman [11]) and let $\text{supp}A=[a_A,b_A]$ and

$\text{supp}B=[a_B,b_B]$ and let $a=\min\{a_A,a_B\}$ and $b=\max\{b_A,b_B\}$, and set

$$A^*=\{(x,y)\mid a\leq x\leq b,\ 0<y\leq\mu_A(x)\}$$

and

$$B^*=\{(x,y)\mid a\leq x\leq b,\ 0<y\leq\mu_B(x)\}$$

Then their distance is:

$$Q(A,B)=q(A^*,B^*). \qquad (2.5)$$

Dissemblance index.

Kaufman and Gupta [18] start with distance between intervals. Let $A=[a_1,a_2]$ and $B=[b_1,b_2]$ be two real intervals contained in $[\beta_1,\beta_2]$ and define:

$$\Delta(A,B)=(|a_1-b_1|+|a_2-b_2|)/2(\beta_2-\beta_1). \qquad (3.1)$$

Generalizing the dissemblance index to fuzzy subsets.

Now let A and B be two fuzzy numbers in $\mathbb{R}$. For each level α, we can consider $\Delta(A_\alpha,B_\alpha)$, where β_1 and β_2 are given by any convenient values which surround A_α and B_α for all $\alpha\in[0,1]$. Kaufman and Gupta [18] now define

$$\Delta_1(A,B)=\int_{\alpha=0}^{1}\Delta(A_\alpha,B_\alpha)d\alpha. \qquad (3.2)$$

As obvious analogies to q_∞ and q_*, we add

$$\Delta_\omega(A,B)=\sup_\alpha\Delta(A_\alpha,B_\alpha) \qquad (3.3)$$

$$\Delta_*(A,B)=\Delta(A_{1.0},B_{1.0}). \qquad (3.4)$$

Set Theoretic Approach.

In his well-known paper "Features of similarity," Tversky [30] described similarity as a feature-matching process. The similarity between objects is expressed as a linear combination of the measure of their common and distinct features. Let $\Delta=\{a,b,c,\ldots,\}$ be the domain of objects under study. Assume that each object in Δ is represented by a set of features or attributes, and let A, B, C denote the set of features associated with objects a,b,c respectively. In this setting, Tversky [30] derived axiomatically the following family of similarity functions:

$$s(a,b)=\theta f(A\cap B)-\alpha f(A-B)-\beta f(B-A)$$

for some $\theta,\alpha,\beta\geq 0$.

This model does not define a single similarity scale, but rather a family of scales characterized by different values of the parameters θ,α, and β, and by the function f.

If $\alpha=\beta=1$ and $\theta=0$ then $-s(a,b)=f(A-B)+f(B-A)$, which is the dissimilarity between sets proposed by Restle [29].

Another matching function of interest is the ratio model:

$$s(a,b)=\frac{f(A\cap B)}{f(A\cap B)+\alpha f(A-B)+\beta f(B-A)} \qquad \alpha,\beta\geq 0,$$

where similarity is normalized so that s lies between 0 and 1. Assuming that f is feature-additive (i.e. $f(A\cup B)=f(A)+f(B)$ for $A\cap B=\emptyset$), then the above model generalizes several set-theoretic models of similarity proposed in the literature. If $\alpha=\beta=1$, $s(a,b)$ reduces to $f(A\cap B)/f(A\cup B)$ (Gregson [12]). If $\alpha=\beta=\frac{1}{2}$ then

$s(a,b)=2f(A\cap B)/(f(A)+f(B))$ (Eisler and Ekman [7]). If $\alpha=1$ and $\beta=0$, $s(a,b)=f(A\cap B)/f(A)$ (Bush and Mosteller [3]). Typically the f function is taken to be the cardinality function.

Generalizing the set theoretic approach to fuzzy subsets.

Several authors have proposed similarity indexes for fuzzy sets which can be viewed as generalizations of the classical set theoretic similarity functions (Dubois and Prade [5]). These generalizations rely heavily upon the definitions of cardinality and difference in fuzzy set theory. Definitions of the cardinality of fuzzy subsets have been proposed by several authors. A systematic investigation of this notion was performed by Dubois and Prade [6]. For generalizing the set-theoretic approach to a similarity index between fuzzy subsets, the scalar cardinality measure will be adopted in the sequel. The scalar cardinality (power) of a fuzzy subset A of U is defined as:

$$|A|=\sum_{u\in U} \mu_A(u), \quad \text{(De Luca and Termini, [4]).}$$

When Support(A) is not finite, we define the power of A to be

$$|A|=\int_{-\infty}^{+\infty} \mu_A(x)dx.$$

Define the following operations between fuzzy subsets:

(a) $\forall x\in X,\ \mu_{A\blacklozenge B}(x)=|\mu_A(x)-\mu_B(x)|.$

$A\blacklozenge B$ is the fuzzy subset of elements that belong more to A than to B, or conversely.

(b) $\forall x\in X,\ \mu_{A\square B}(x)=\max[\min(\mu_A(x),1-\mu_B(x)),\min(1-\mu_A(x),\mu_B(x))].$

$A\square B$ is the fuzzy subset of elements that approximately belong

to A and not to B, or conversely.

The following indexes have been proposed in the literature (Dubois and Prade [5]) as dissimilarity measures between fuzzy subsets:

$$S_1(A,B)=1-|A\cap B|/|A\cup B| \qquad (4.1)$$

is analogous to Gregson's [12] proposal for classical sets;

$$S_2(A,B)=|A\odot B| \qquad (4.2)$$

is analogous to Restle's [29] proposal for classical sets.

Also

$$S_3(A,B)=\sup_{x\in X}\mu_{A\odot B}(x) \qquad (4.3)$$

and finally a disconsistency index ("degree of separation", Enta [8])

$$S_4(A,B)=1-\sup_{x\in X}\mu_{A\cap B}(x). \qquad (4.4)$$

A pattern recognition approach (Bonissone [2]).

In this approach, the efficiency of the linguistic approximation process is of major importance. The process is composed of two stages. In the first stage, the set of possible labels is narrowed down by using a crude measure of distance that (hopefully) performs well on fuzzy sets that are far apart from each other. The idea is to represent each fuzzy set by a limited number of features so that the distance computation is simplified. Bonissone [2] chose four features: (i) the power of the set (area under the curve), (ii) a measure of the fuzziness of

the set (non-probabilistic entropy) defined by De Luca and Termini [4] as

$$\text{Entropy}(A)=\int_{-\omega}^{+\omega} S(\mu_A(x))dx$$

where $S(y)=-y\text{Ln}(y) - (1-y)\text{Ln}(1-y)$,

(iii) The first moment (center of gravity of the membership function) defined by

$$\text{FMO}(A)=\left(\int_{-\omega}^{+\omega} x\mu_A(x)dx\right) / \text{Power}(A),$$

and finally (iv) Skewness defined as

$$\text{Skew}(A)=\int_{-\omega}^{+\omega} (\mu_A(x)-\text{FMO}(A))^3\mu_A(x)dx.$$

Bonissone [2] defines the distance between two fuzzy sets as the weighted euclidean distance between the vectors (Power(A), Entropy(A), FMO(A), Skew(A)) and (Power(B), Entropy(B), FMO(B), Skew(B)). In what follows we will denote this distance by $V_1(A,B)$ (using equal weights). After narrowing down the set of possible labels, the second stage starts, in which a modified Bhattacharyya distance is computed. This distance should discriminate well between sets that are close to each other. The Bhattacharyya distance is defined as:

$$R(A,B)=\left[1-\int_{-\omega}^{+\omega} (\mu^*_A(x)\cdot\mu^*_B(x))^{1/2}dx\right]^{1/2} \qquad (5)$$

where the membership functions have been normalized, i.e.:

$$\mu^*_A(x)=\mu_A(x) / \text{Power}(A)$$

and similarly for μ_B.

Wenstop [34] adopted a similar approach. He represented each fuzzy set as the 2-vector: (Power(A), FMO(A)). The distance between two fuzzy sets is defined to be the regular euclidean distance between the two corresponding vectors. We will denote this distance by $V_2(A,B)$.

Correlation Index.

Murthy, Pal, and Majumder [21] defined a correlation-like index that reflects the similarity in behavior of two fuzzy sets. The measure is actually a standardized squared euclidean distance between two fuzzy sets as defined by d_2.

Let

$$X_A = \int_{-\omega}^{+\omega} (2\mu_A(x)-1)^2 dx$$

and define

$$CORR(A,B) = 1-(4/X_A+X_B)\cdot(d_2)^2. \qquad (6)$$

In what follows we will use the index $\rho(A,B)=1-CORR(A,B)$.

METHOD

Subjects: 15 native speakers of English were recruited by placing notices in graduate students' mailboxes in the business school and the departments of anthropology, economics, history, psychology, and sociology at the University of North Carolina at Chapel Hill. We assumed that they would represent a population of people who think seriously about communicating "degrees of uncertainty," and who generally do so with non-numerical phrases. The general nature of the study was described and subjects were

promised $25 for three sessions of approximately an hour and a half each.

General procedure: Subjects were run for a practice session and then two data sessions. The experiment was controlled by an IBM PC with the stimuli presented on a color monitor and responses made using a joystick. During the data sessions, subjects worked through 4 types of trials: (1) linguistic probability scaling trials, (2) Trials in which subjects integrated two probability terms connected by AND, (3) Similar trials using the connector OR, and (4) Similarity judgment trials. This paper is concerned only with tasks (1) and (4). For more details about the other two tasks see Wallsten et. al. [33].

(1) **Linguistic probability scaling trials.**

The objective of these trials is to establish the subject's membership function for various linguistic probability phrases. Recently, Wallsten et. al. [31] have developed a method for empirically establishing the membership functions of fuzzy concepts, based on conjoint measurement and utilizing a graded pair-comparison technique. Rapoport, Wallsten, and Cox [28] further established that the methods of direct magnitude estimation and graded pair-comparison yield similar membership functions. In this study we have adopted the direct magnitude estimation technique. On such a trial a red and white radially divided spinner appeared on the screen, as shown in Fig. 1.

****

Insert Fig. 1 about here

****

Instructions for this task said in part:

> Imagine that you cannot see the spinner, but you have to predict whenever it will land on white on the next random spin. A friend of yours can see the spinner, although not too well because it is rotating at a moderate rate. Your friend is going to give you his or her best opinion about the chances of the spinner landing on white. However, this person does not tell you a probability... Rather, the person may use any of a large number of nonnumerical probability phrases... We are interested in your interpretation of the probability phrases as they apply to the spinner context. Assume that your friend tells you that it is doubtful that the spinner will land on white. This gives you some basis for judging the probability of that event. Now, consider the spinner on the screen. How close is that spinner's probability of landing on the white to the judgment you had formed upon hearing that it is doubtful that the spinner will land on white?

The subject then moved the cursor on a line to indicate how close the particular displayed spinner came to the opinion that he or she had formed on the basis of the phrase doubtful. The cursor could be moved from not at all close (low membership) to absolutely close (high membership).

Six phrases were employed, three representing lower probabilities and three representing higher probabilities: doubtful, slight chance, improbable, likely, good chance, and fairly certain. In the direct estimation task, each phrase was presented with 11 spinner probabilities: 0.02, 0.12, 0.21, 0.31, 0.40, 0.50, 0.60, 0.69, 0.79, 0.88, and 0.98.

Subjects judged each combination of phrase and probability numbers twice, once in each session.

(2) **Similarity judgment trials.**

Instructions for this task said in part:

> ... Two non-numerical probability phrases will be printed on the screen on each trial. We are interested in how similar, or synonymous, you consider the two phrases to be with respect to describing the probability of a spinner landing on white.

The subject then moved the cursor on a line to indicate how similar the phrases are. The cursor could be moved from not at all similar to absolutely similar. Each subject judged the similarity between all possible pairs (15) (excluding a phrase and itself) twice in each session.

Membership function evaluation: Based on previous research (Wallsten et. al. [31]; Rapoport et. al. [28]) we have concluded that a cubic polynomial can accurately represent the membership functions for the six phrases. Note that a cubic polynomial resembles the 'S' and 'Π' functions that have been proposed in the literature in this context (Eshragh and Mamdani [9]). A cubic polynomial was fit to the 22 points representing each phrase within a subject, using the least squares technique. Each function was then normalized to attain the value 1 on the interval [0,1]. In defining the membership functions, any value less than zero was redefined to equal 0, and similarly any value greater than 1 was redefined to equal 1. These adjustments were generally quite minor. Examples of the membership functions for the six phrases for one subject are shown in Fig. 2. All membership functions for all subjects were either nondecreasing, nonincreasing, or single-peaked.

****

Insert Fig. 2 about here

****

RESULTS AND DISCUSSION

For each subject and each pair of words, all nineteen distance measures were calculated. (At times it was necessary to discretize one axis, using a 100-point grid, in order to calculate a distance measure.) To evaluate the performance of a particular distance measure, we compare its computed values to the "true" distance ratings as given directly by the subject in the similarity judgment trials. This evaluation is done on two levels: First we can ask if the distance measure correctly categorizes a "similar" pair of words by returning a "small" distance, and if it correctly categorizes a "dissimilar" pair of words by returning a "large" distance. This crude evaluation is, in practice, independent of the subject-specific "true" distance rating, because the subjects generally agree that the pairs p_1=(doubtful, improbable), p_2=(doubtful, slight chance), p_3=(improbable, slight chance), p_4=(fairly certain, good chance), p_5=(fairly certain, likely), p_6=(likely, good chance) are each composed of two "similar" words; likewise the subjects generally agree that the pairs q_1=(doubtful, fairly certain), q_2=(doubtful, good chance), q_3=(doubtful, likely), q_4=(improbable, fairly certain), q_5=(improbable, good chance), q_6=(improbable, likely), q_7=(slight chance, fairly certain), q_8=(slight chance, good chance), q_9=(slight chance, likely) are each composed of two

"dissimilar" words. For this task of dichotomous categorization, essentially all the distance measures were successful across all subjects (see Figure 3, for example). This is testimony to the intuitive base upon which each distance definition rests: They are designed to indicate gross differences between membership functions, if and only if such differences actually exist. The practical implication is simply that if linguistic approximation or concept-clustering is to be carried out in two stages, then any of these distance measures may be used for the first stage.

****

Insert Figure 3 about here

****

The second level of our evaluation asks whether the distance measure reflects the correct degree of similarity within "similar" pairs of words, and whether the distance measure reflects the correct degree of dissimilarity within "dissimilar" pairs of words. In answering this more subtle question, intersubject variability must be acknowledged: Each subject will have his own membership functions for the words in pair p_i. These two membership functions are "similar" in the gross sense, but the similarity between them is different than the similarity between his membership functions for the words in pair p_j. The degree of similarity within each pair is given, for that subject, by his "true" distance rating. If the distance measure works well in the context of fuzzy sets, it should yield distances for pairs p_i and p_j that "agree" with the corresponding "true" distance ratings given by the subject. To quantify the amount of

agreement between a particular distance measure and the "true" distance we compute the correlation between these two quantities, over all pairs $\{p_i : 1 \leq i \leq 6\}$, for a given subject (see Figure 3). Thus our criterion for agreement is linear association. Now, the same considerations apply to the "dissimilar" pairs. Here we compute the correlation between the particular distance measure and the "true" distance over all pairs $\{q_i : 1 \leq i \leq 9\}$, for a given subject. By analyzing the p_i's and q_i's separately, we allow for the possibility that a particular distance measure may be quite accurate in modeling fine variations in similarity (i.e. small distances), but may be quite inaccurate in modeling fine variation among pairs that are each composed of two "dissimilar" words. Furthermore, in practical applications one may only need to find a distance measure that is sensitive to the degree of similarity in pairs of "similar" words (e.g. in linguistic approximation). The separate analyses also give a distance measure the opportunity to be linearly related to "true" distance with two (locally) different slopes (see Figure 3).

****

Insert Figures 4 & 5 about here

****

For each distance measure, its p_i-correlations for the 15 subjects are summarized by a line-plot. The 19 line-plots (one for each measure) appear in Figure 4. Analogous line-plots of the q_i-correlations appear in Figure 5. It is desirable for a measure to have high mean and median correlation, to have small

dispersion among its correlations (i.e. interquartile range), and to be free of extremely low (i.e. negative) correlations.

Several trends are clear from these displays:

(i) There is a great deal more variability between the performances of the various distance measures on "dissimilar" pairs (Fig. 5) than on "similar" pairs (Fig. 4): the means, medians, and interquartile ranges are much more homogeneous in Figure 4 than in Figure 5. (Note that statistical fluctuation would actually work in the opposite direction: the correlations for the "dissimilar" pairs are calculated from 9 data points, while those for "similar" pairs are calculated from 6 data points.) This immediately suggests that more caution must be exercised when selecting a distance measure to distinguish between varying degrees of dissimilarity.

(ii) On the "dissimilar" pairs (Fig. 5), those measures which perform the worst $(d_2, (d_2)^2, d_1, d_\omega, S_2, S_\infty, \rho)$ are measures which ignore the ordering on the x-axis (the base variable axis). Conversely, those measures which perform the best $(q_\omega, q_*, \Delta_\omega, \Delta_*)$ are measures which do account for the distances on the x-axis by looking at α-level-sets. This distinction is quite logical: When measuring the distance between words that are essentially "dissimilar" (i.e. have nearly disjoint supports), it is the x-axis that carries all the information regarding the degree of dissimilarity between the membership functions. Distance measures that ignore the x-ordering have the advantage of being unambiguously defined even for membership functions over abstract (i.e. unordered) spaces, but such measures have the disadvantage of being insensitive to varying degrees of dissimilarity (e.g. as

in pairs q_i). In the "similar" pairs (Fig. 4), the membership functions within a pair (p_i) have nearly identical supports. Hence the x-distance is not critical, and we find both types of distance measures doing well- those that look at α-level-sets (notably $q_*, \Delta_\omega, \Delta_*$), and those that ignore the ordering on x (notably S_4).

(iii) Among those measures accounting for x-ordering ($q_1, q_\omega, q_*, \Delta_1, \Delta_\omega, \Delta_*, Q$), q_1 and Q are especially susceptible to having extremely poor correlation with "true" similarity ratings. This occurs for both q_i-correlations and p_i-correlations. Note that Q is conceptually different from the other 6 such measures, possibly accounting for the difference in performance.

(iv) Measure S_2 is arguably the worst both for "similar" pairs and for "dissimilar" pairs.

(v) Measures S_1 and S_4 are clearly the best in terms of q_i-correlations, among those measures which ignore the x-ordering. Their superiority in the "dissimilar" setting is noteworthy because, again, x-distance <u>is</u> relevant in this setting. Furthermore, measure S_4 performs reasonably well (among all measures) in the "similar" setting also.

(vi) Quite surprisingly, all of the measures with consistently good performance ($S_4, q_\omega, q_*, \Delta_\omega, \Delta_*$) share the following property: they concentrate their attention on a single value, rather than performing some sort of averaging or integration. In the case of S_4, attention focuses on the particular x-value where the membership function of $A \cap B$ is largest; in q_ω and Δ_ω attention focuses on the α-level-set where the x-distance is largest; in q_*

and Δ_* attention focuses on the x-distance at the highest membership grade. Such measures are generally considered unstable (and hence suspect) in many mathematical analyses. Yet, here is strong empirical evidence that subjects actually behave this way: reduction of complicated membership functions to a single "slice" may be the intuitively natural way for human beings to combine and process fuzzy concepts.

(vii) The consistently good performance of q_* and Δ_* has significant practical implications. These measures are trivial to compute, relative to other distance measures, and they have substantial intuitive appeal.

(viii) Distance measure R was proposed as a refinement of V_1, where the latter is used in the first stage of linguistic approximation and the former is used in the second stage (Bonissone [2]). However, the empirical results show no systematic evidence of R being superior in the "similar" word setting (Figure 4) or of V_1 being superior in the "dissimilar" word setting (Figure 5).

Recommendations.

If one wants to select a distance measure that performs well in the long-run on a broad spectrum of subjects, then the aggregated data of our study may be used as a guide. Measures S_4, q_*, Δ_∞, and Δ_* consistently distinguished themselves for good performance. If, on the other hand, the objective is to accurately model the behavior of a specific individual (e.g. in the linguistic approximation phase of an expert system program), then the following problem must be acknowledged: For each

distance measure, there existed some subject for whom that distance measure performed quite poorly (note the "minimum" values on Figures 4 and 5). Therefore, in these applications, it would be ideal to determine the best distance measure for the individual of interest. This could be accomplished by carrying out an experiment analogous to ours, but on the specific individual, and in the relevant context. It is possible that the relative performances of the distance measures could vary from one context to another, even for a fixed individual.

In many cases, the fuzzy concepts are unambiguously defined over a one-dimensional space (e.g. in our study of probability words). When this is not the case, then, in using those distance measures that do account for the ordering on the base-variable axes, it is imperative that the fuzzy concepts be correctly located in a space of the appropriate dimensionality.

References

1. F. Attneave, Dimensions of similarity, American J. of Psychol. 63:516-556 (1950).

2. P. P. Bonissone, A pattern recognition approach to the problem of linguistic approximation in system analysis, Proc. IEEE Int. Conf. Cybern. and Society, Denver, CO, 1979, pp. 793-798.

3. R. R. Bush and F. Mosteller, A model for stimulus generalization and discrimination, Psychol. Review 58:413-423 (1951).

4. A. De Luca and S. Termini, A definition of a non-probabilistic entropy in the setting of fuzzy sets theory, Inform. Control 20:301-312 (1972).

5. D. Dubois and H. Prade, Fuzzy Sets and Systems: Theory and Applications, Academic Press, New York, 1980.

6. D. Dubois and H. Prade, Fuzzy cardinality and the modeling of imprecise quantification, Fuzzy Sets and Systems 16:199-230 (1985).

7. H. Eisler and G. A. Ekman, A mechanism of subjective similarity, Acta Psychologica 16:1-10 (1959).

8. Y. Enta, Fuzzy decision theory, in Applied Systems and Cybernetics, Proceedings of the International Congress on Applied Systems Research and Cybernetics (G.E. Lasker, Ed.), Acapulco, Mexico, 1980, pp. 2980-2990.

9. F. Eshragh and E. H. Mamdani, A general approach to linguistic approximation, Int. J. Man-Machine Studies 11:501-519 (1979).

10. R. Giles, The practical interpretation of fuzzy concepts, Human Systems Management 3:263-264 (1983).

11. R. Goetschel and W. Voxman, Topological properties of fuzzy numbers, Fuzzy Sets and Systems 10:87-99 (1983).

12. R.M. Gregson, Psychometrics of similarity, Academic Press, New York, 1975.

13. H. M. Hersh and A. Caramazza, A fuzzy set approach to modifiers and vagueness in natural language, J. Exp. Psychol.: General 105:254-276 (1976).

14. H. M. Hersh, A. Caramazza and H. H. Brownell, Effects of context on fuzzy membership functions, in Advances in Fuzzy Set Theory and Applications (M. M. Gupta, R. K. Ragade and R. R. Yager, Eds.), North Holland, Amsterdam, 1979, pp. 389-408.

15. C. L. Hull, Principles of Behavior, Appleton, New York, 1943.

16. J. Kacprzyk, Fuzzy set-theoretic approach to the optimal assignment of work places, in Large Scale Systems Theory and Applications (G. Guardabassi and A. Locatelli, Eds.), Udine, Italy, 1976, pp. 123-131.

17. A. Kaufman, Introduction to the Theory of Fuzzy Subsets. Vol 1: Fundamental Theoretical Elements, Academic Press, New York, 1973.

18. A. Kaufman and M. M. Gupta, Introduction to Fuzzy Arithmetic, Van Nostrand Reinhold Co., New York, 1985.

19. M. Kochen, Applications of fuzzy sets in psychology, in Fuzzy sets and their applications to cognitive and decision processes (L. Zadeh, K. S. Fu, K. Tanaka and M. Shimura, Eds.), Academic Press, New York, 1975, pp. 395-408.

20. K. S. Lashley, An examination of the "continuity theory" as applied to discriminative learning, J. General Psychol. 26:241-265 (1942).

21. C. A. Murthy, S. K. Pal and D. Dutta Majumder, Correlation between two fuzzy membership functions, Fuzzy Sets and Systems 17:23-38 (1985).

22. M. Nowakowska, Methodological problems of measurement of fuzzy concepts in the social sciences, Behavioral Sci. 22:107-115 (1977).

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

24. G. C. Oden, Integration of fuzzy logical information, J. Exp. Psychol.: Hum. Percept. Perform. 3:565-575 (1977).

25. G. C. Oden, A fuzzy logical model of letter identification, J. Exp. Psychol.: Hum. Percept. Perform. 5:336-352 (1979).

26. G. C. Oden, Integration of fuzzy linguistic information in language comprehension, Fuzzy Sets and Systems 14:29-41 (1984).

27. A. L. Ralescu and D. A. Ralescu, Probability and fuzziness, Inform. Sci. 34:85-92 (1984).

28. A. Rapoport, T. S. Wallsten and J. A. Cox, Direct and indirect scaling of membership functions of probability phrases, J. Math. Modeling, to appear.

29. F. Restle, A metric and an ordering on sets, Psychometrica 24:207-220 (1959).

30. A. Tversky, Features of similarity, Psychol. Review 84:327-352 (1977).

31. T. S. Wallsten, D. V. Budescu, A. Rapoport, R. Zwick, and B. Forsyth, Measuring the vague meaning of probability terms, Tech. Report 173, L. L. Thurstone Psychometric Laboratory, University of North Carolina, Chapel Hill, 1985.

32. T. S. Wallsten, S. Fillenbaum and J. A. Cox, Base rate effects on the interpretations of probability and frequency expressions, J. Memory and Language, to appear.

33. T. S. Wallsten, Zwick R. and D. V. Budescu, Integrating vague information, Paper presented at the 18th Annual Mathematical Psychology Meeting, La Jolla, California, August 29, 1985.

34. F. Wenstop, Quantitative analysis with linguistic values, Fuzzy Sets and Systems 4:99-115 (1980).

35. M. Zeleny, On the (ir)relevancy of fuzzy sets theories, Human Systems Management 4:301-306 (1984).

36. A. C. Zimmer, Verbal vs. numerical processing of subjective probability, in Decision making under uncertainty (R. W. Scholz, Ed.), North Holland, Amsterdam, 1973, pp. 159-182.

37. P. Zysno, Modeling membership functions, in Empirical Semantics (B. B. Rieger, Ed.), Brockmeyer, Bochum, 1981, pp.350-375.

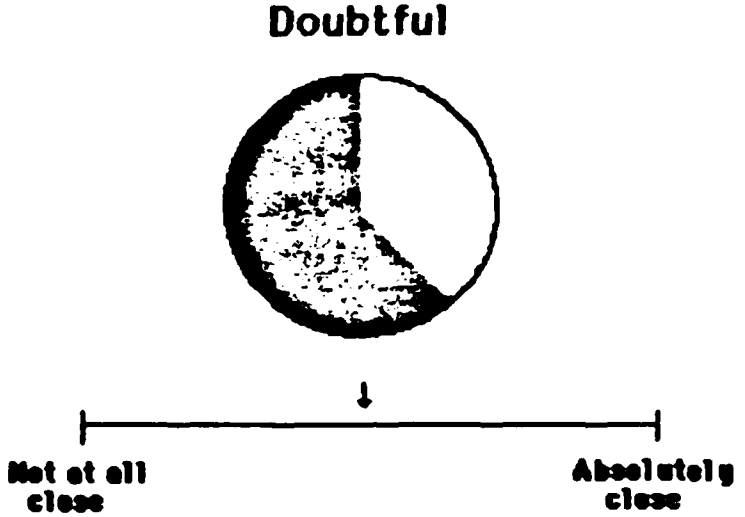


Figure 1: Direct Estimation Trial

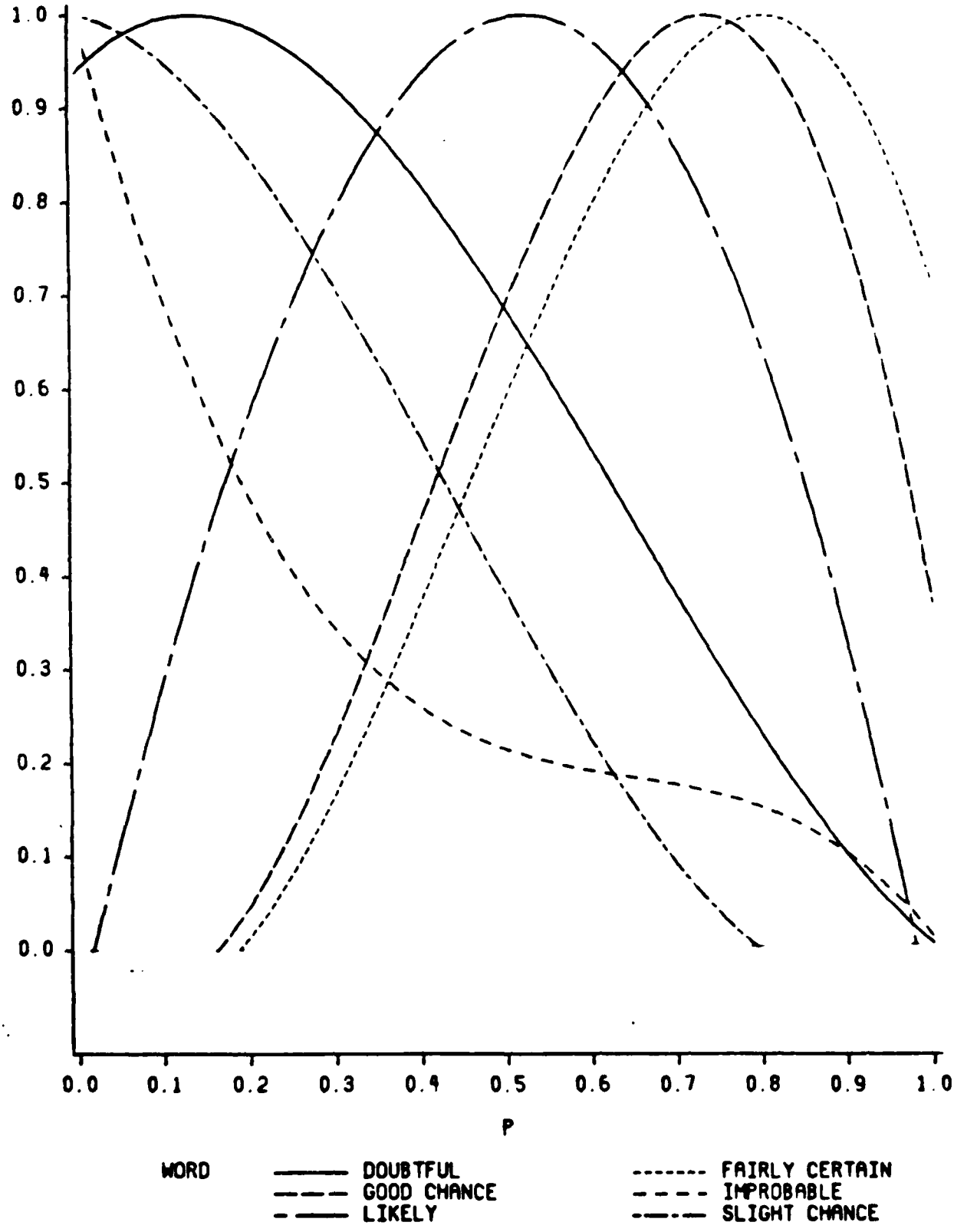


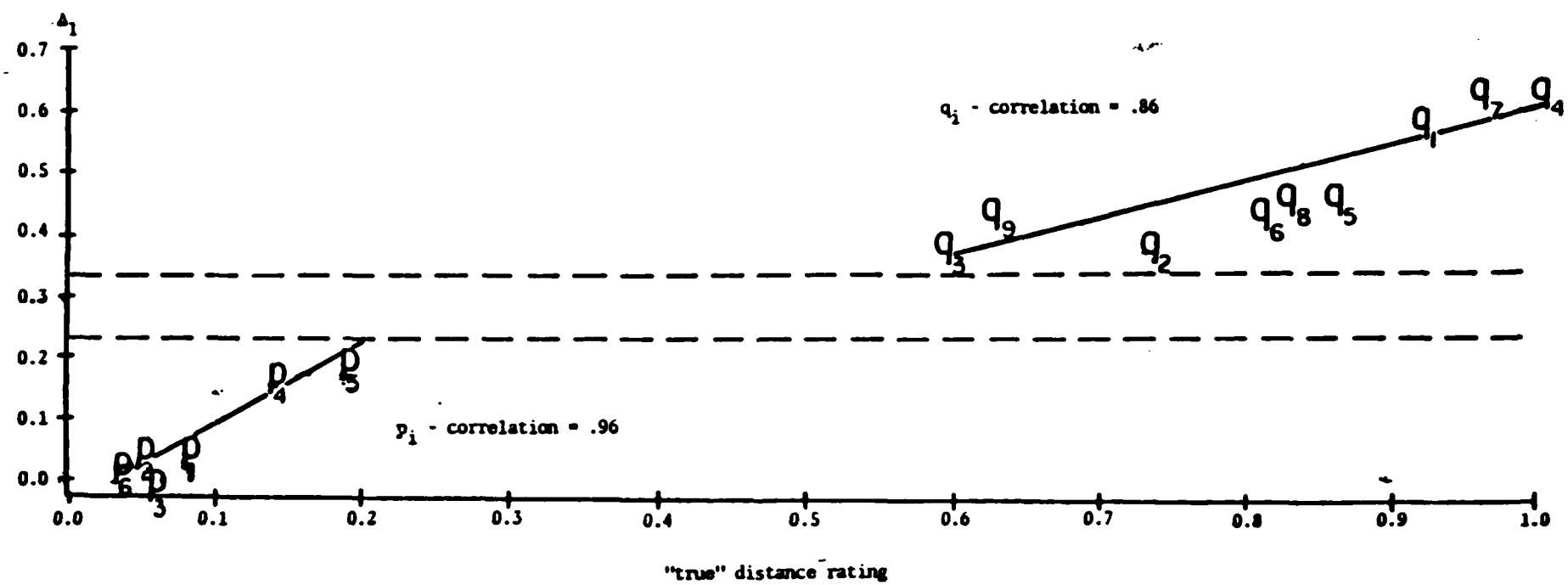
Figure 2:
Membership functions from a single subject

FIGURE 3

Plot of distance measure Δ_1 vs "true" distance rating, for subject #5.

p_i = "similar" words in pair.

q_i = "dissimilar" words in pair.



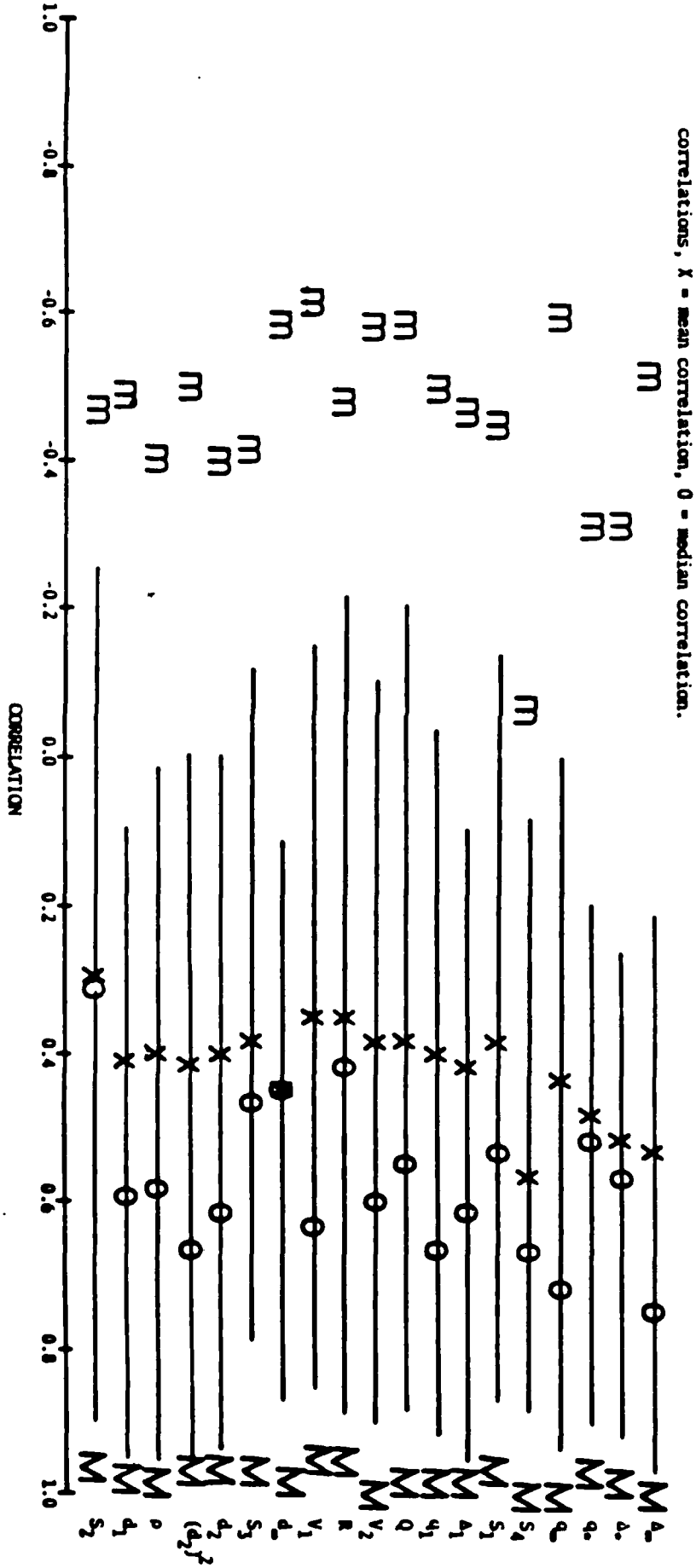


FIGURE 4

Line-plots of p_i - correlations for each distance measure. For a particular subject and a particular distance measure, the correlation is calculated between the "true" distance rating and the distance measure, over all p_i (pairs containing "similar" words). Data in a single line is aggregated over all 15 subjects: m = minimum correlation, M = maximum correlation, —— = interquartile range of correlations, X = mean correlation, O = median correlation.

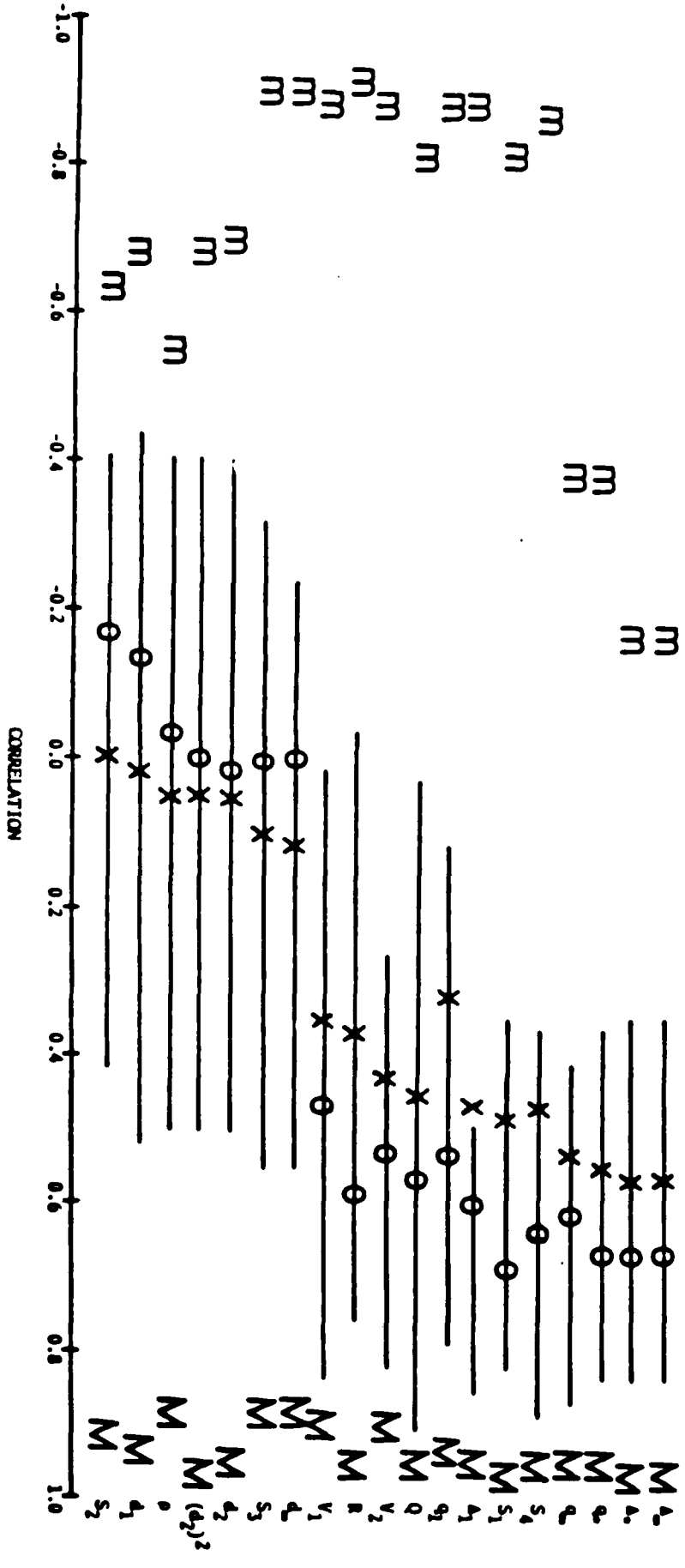


FIGURE 5

Line-plots of q_1 - correlations for each distance measure. For a particular subject and a particular distance measure, the correlation is calculated between the "true" distance rating and the distance measure, over all q_1 (pairs containing "dissimilar" words). Data in a single line is aggregated over all 15 subjects: m = minimum correlation, M = maximum correlation, —— = interquartile range of correlations, I = mean correlation, 0 = median correlation.