Measures of Similarity Among Fuzzy Concepts: A Comparative Analysis*

Rami Zwick
Carnegie-Mellon University

Edward Carlstein
University of North Carolina at Chapel Hill

David V. Budescu
University of Haifa

ABSTRACT

Many measures of similarity among fuzzy sets have been proposed in the literature, and some have been incorporated into linguistic approximation procedures. The motivations behind these measures are both geometric and set-theoretic. We briefly review 19 such measures and compare their performance in a behavioral experiment. For crudely categorizing pairs of fuzzy concepts as either "similar" or "dissimilar," all measures performed well. For distinguishing between degrees of similarity or dissimilarity, certain measures were clearly superior and others were clearly inferior; for a few subjects, however, none of the distance measures adequately modeled their similarity judgments. Measures that account for ordering on the base variable proved to be more highly correlated with subjects' actual similarity judgments. And, surprisingly, the best measures were ones that focus on only one "slice" of the membership function. Such measures are easiest to compute and may provide insight into the way humans judge similarity among fuzzy concepts.

KEYWORDS: *similarity measures, fuzzy concepts*

INTRODUCTION

Giles [12] 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 (p. 263).

Fuzzy set theory would benefit from becoming a behavioral science, having its assumptions validated and its results verified by empirical findings (Kochen [22]). In particular, there has been virtually no experimental work comparing the many measures of distance (between fuzzy sets) that have been proposed in the literature. The major empirical works that have appeared in the fuzzy set literature focus on measuring the membership function and evaluating the appropriateness of operations on fuzzy sets. (See, for example, Hersh and colleagues [15, 16]; Kochen [22]; Norwich and Turksen [26]; Oden [28, 29, 30, 31]; Rapoport and colleagues [33]; Thole and Zimmermann [35]; Wallsten and colleagues [37]; Zimmer [42]; Zysno [44].) This article investigates experimentally the question of selecting an appropriate distance index for measuring similarity among fuzzy sets.

Several methods have been suggested for the process of linguistic approximation (Bonissone [3]; Eshragh and Mamdani [11]; Wenstøp [40]). Each of them suggests a different measure of similarity. However, there is no serious attempt to validate the techniques through behavioral experiments. Some authors have mentioned that their techniques work very well but do not provide the appropriate data to support their claim. For example, Bonissone [3] in his pattern recognition approach to linguistic approximation writes that ''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''; however, no results are reported, and it is not clear what criteria are used to make such a statement. Similarly, no serious attempts have been made by Wenstøp [40] to validate details of his semantic model. Neither do Eshragh and Mamdani [11] 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,'' once again no supporting data are 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 the crucial role (translation) that this index plays in any implementation of fuzzy reasoning theory, and 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 among fuzzy concepts, based on his or her separate membership functions of each concept.

The notion of similarity plays a fundamental role in theories of knowledge and behavior and has been dealt with broadly in the psychology literature (Gregson [14]). Overall, the theoretical analysis of similarity relations has been dominated by geometric models. These models represent objects as points in some coordinate space such that the observed dissimilarity among objects corresponds to the metric distance between the respective points.

The similarity indexes used in the linguistic approximation techniques adopt this approach. Bonissone [3] locates each concept initially in four-dimensional space, where the dimensions are power, entropy, first moment, and skewness of the membership function. He defines the distance between two concepts as the regular weighted Euclidean distance between the points representing these concepts. Wenstøp [40] locates 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 has been adopted by Eshragh and Mamdani [11] and by Kacprzyk [18].

Most conclusions regarding the appropriate distance metric have been based on studies using judgment of similarity among stimuli that can be located *a priori* along (objectively) distinguishable dimensions (such as color, tones, etc.). The question of integral versus 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 relations 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 [36], 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 demonstrate problems with the metric assumption. Tversky [36] shows that similarity may not be a symmetric relation (violating the symmetry axiom of a metric) and also suggests that all stimuli may not be equally similar to themselves (violating the minimality axiom.)

Therefore, similarity may be better modeled by a function that is not conceptually a geometric distance (such as a set-theoretic function instead).

In this article we first review the various distance indexes suggested in the literature, in the general case and as adapted to fuzzy sets. Next, we present our experimental design. Finally, we discuss the results and implications of the results for 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 (Beckenbach and Bellman [2]). This metric is a one-parameter class of distance functions defined as follows:

$$d_r(x, y) = \left[\sum_{i=1}^{n} |x_1 - y_i|^r \right]^{1/r}, \qquad r \geq 1 \qquad (1)$$

where x and y are two points in an n-dimensional space with components (x_i, y_i) $i = 1, 2, \ldots, n$. 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 $r = \infty$. The case of $r = 1$ is known as the "city-block" model. As r approaches ∞, equation (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,

$$d_\infty(x, y) = \max_i |x_i - y_i| \qquad (1.1)$$

Each of the three distance functions, $r = 1, 2,$ and ∞, are used in psychological theory (Hull [17], Restle [34], Lashley [23]).

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:

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

or, if $E = R$,

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

and

$$d_\infty(A, B) = \sup_x |\mu_A(x) - \mu_B(x)| \qquad (1.4)$$

The cases $r = 1$ and 2 were studied by Kaufman [20]. Kacprzyk [18] proposed the distance measure $(d_2)^2$, and d_∞ was proposed by Nowakowska [27]. Our empirical evaluation will consider d_1, d_2, $(d_2)^2$, and d_∞.

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

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

where d_2 is as defined in equation (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 $F(R)$ be the set of all fuzzy numbers and fuzzy intervals of the real line (Dubois and Prade [6]). There is no unique metric in $F(R)$ that extends the Hausdorff distance. Ralescu and Ralescu [32] propose the following generalizations:

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

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

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 [13] suggest 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 [13]). Let supp $A = [a_A, b_A]$ and supp $B = [a_B, b_B]$, and let $a = \min \{a_A, a_B\}$ and $b = \max \{b_A, b_B\}$, and set

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

and

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

Then their distance is

$$Q(A, B) = q(A^*, B^*) \tag{2.5}$$

DISSEMBLANCE INDEX Kaufman and Gupta [21] 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

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

GENERALIZING THE DISSEMBLANCE INDEX TO FUZZY SUBSETS Now let A and B be two fuzzy numbers in 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 that surround A_α and B_α for all $\alpha \in [0, 1]$. Kaufman and Gupta [21] now define

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

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

$$\Delta_\infty(A, B) = \sup_\alpha \Delta(A_\alpha, B_\alpha) \tag{3.3}$$

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

Set-Theoretic Approach

In his well-known paper entitled "Features of Similarity," Tversky [36] describes similarity as a feature-matching process. Similarity among objects is expressed as a linear combination of the measure of their common and distinct features. Let $D = \{a, b, c, \ldots,\}$ be the domain of objects under study. Assume that each object in D is represented by a set of features or attributes, and let A, B, and C denote the set of features associated with objects a, b, and c, respectively. In this setting Tversky derives 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 [34].

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 $A \cap B = 0$), then the foregoing 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 [14]). If $\alpha = \beta = \frac{1}{2}$, then $s(a, b) = 2f(A \cap B)/(f(A) + f(B))$ (Eisler and Ekman [9]). If $\alpha = 1$ and $\beta = 0$, $s(a, b) = f(A \cap B)/f(A)$ Bush and Mosteller [4]). 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 that can be viewed as generalizations of the classical set-theoretic similarity functions (Dubois and Prade [7]). These generalizations rely heavily on 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 [8]. For generalizing the set-theoretic approach to a similarity index among 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 (DeLuca and Termini [5])

$$|A| = \sum_{u \in U} \mu_A(u)$$

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

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

Defining the following operations between fuzzy subsets,

$\forall x \in U, \ \mu_{A \cap B}(x) = \min [\mu_A(x), \mu_B(x)]$

$\forall x \in U, \ \mu_{A \cup B}(x) = \max [\mu_A(x), \mu_B(x)]$

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

$A \square B$ 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 [7]) 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 [34] proposal for classical sets, and

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

is analogous to Restle's proposal [34] for classical sets. Also,

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

and finally a disconsistency index ("degree of separation," Enta [10]);

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

A PATTERN RECOGNITION APPROACH In this approach (Bonissone [3]), 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 (it is hoped) 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 [3] chooses four features. The first is the power of the set (area under the curve), and the second is a measure of the fuzziness of the set (nonprobabilistic entropy) defined by De Luca and Termini [5] as

$$\text{entropy } (A) = \int_{-\infty}^{+\infty} S(\mu_A(x)) \, dx$$

where $S(y) = -y ln(y) - (1 - y) ln (1 - y)$

The third feature is the first moment (center of gravity of the membership function) and is defined by

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

And finally, skewness, the fourth feature, is defined as

$$\text{skew } (A) = \int_{-\infty}^{+\infty} (x - \text{FMO } (A))^3 \mu_A(x) \, dx$$

Bonissone [3] defines the distance between two fuzzy sets as the 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 $V_1(A, B)$. 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 (Kailath [19])

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

where the membership functions have been normalized, that is,

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

and similarly for μ_B.

Wenstøp [40] adopts a similar approach. He represents each fuzzy set as the two-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 [24] define 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

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

and define

$$\text{CORR }(A, B) = 1 - \left(\frac{4}{X_A + X_B}\right)(d_2)^2 \tag{6}$$

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

METHOD

Subjects

Fifteen 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 nonnumerical 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 were made using a joystick. During the data session, subjects worked through four types of trials: linguistic probability scaling trials, similarity judgment trials, and two types of trials involved integrating two probability terms connected by *"and"* and *"or"*. (These two types of trials are discussed in Wallsten and co-workers [39] and will not be commented on here.)

LINGUISTIC PROBABILITY SCALING TRIALS The objective of these trials was to establish the subject's membership function for various linguistic probability phrases. A linguistic probability phrase is a value of the linguistic variable "probability" (Zadeh [41]). In this study we adopted the direct magnitude estimation technique (for instance, Norwich and Turksen [25, 26], Rapoport and colleagues [33]).

In these trials, probabilities were represented as relative areas on a radially divided two-colored spinner (see Figure 1). On each trial a spinner and a linguistic probability word (such as "doubtful") appeared on the screen. The subject was asked to indicate how "close" the probability word is to the actual probability represented by the dark area of the spinner. The subject's response was given by placing the cursor along the horizontal axis (see Figure 1).

Six probability 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.
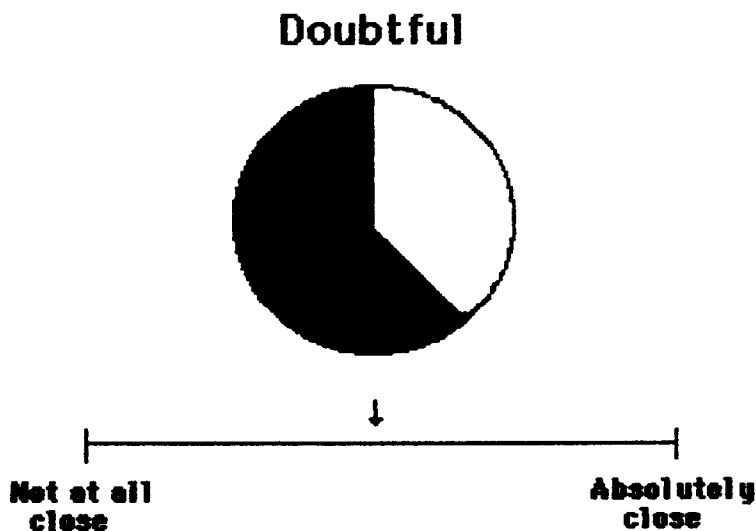

Figure 1. Direct Estimation Trial

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

SIMILARITY JUDGMENT TRIALS In these trials two probability phrases were printed on the screen. The subject then moved the cursor on a horizontal axis to indicate how similar the phrases were to each other. 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 We adopted the view that an individual's membership function for a given fuzzy concept is not purely deterministic. Rather, the value of the membership function at a point is itself a random variable (Norwich and Turksen [25], Zwick [43]). Hence, in the linguistic probability scaling trials, the subject's placement of the cursor yielded a realization of this random variable. On the basis of previous research (Wallsten and colleagues [37]; Rapoport and colleagues [33]), we concluded that a cubic polynomial can accurately represent the expected value of the membership function for a probability phrase. Note that a cubic polynomial resembles the ''S'' and ''Π'' functions that have been proposed in the literature in this context (Eshragh and Mamdani [11]). A cubic polynomial was fit to the 22 points representing each phrase within a subject, using the least squares technique. Each equation 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 Figure 2. All membership functions for all subjects were either nondecreasing, nonincreasing, or single peaked.

RESULTS AND DISCUSSION

For each subject and each pair of words, all 19 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 compared its computed values to the ''true'' distance ratings as given directly by the subject in the similarity judgment trials. This evaluation was done on two levels. First, we asked if the distance measure correctly categorized a ''similar'' pair of words by returning a ''small'' distance, and if it correctly categorized a ''dissimilar'' pair of words by returning a ''large'' distance. This crude evaluation was in practice independent of the subject-specific ''true'' distance rating, because the subjects generally
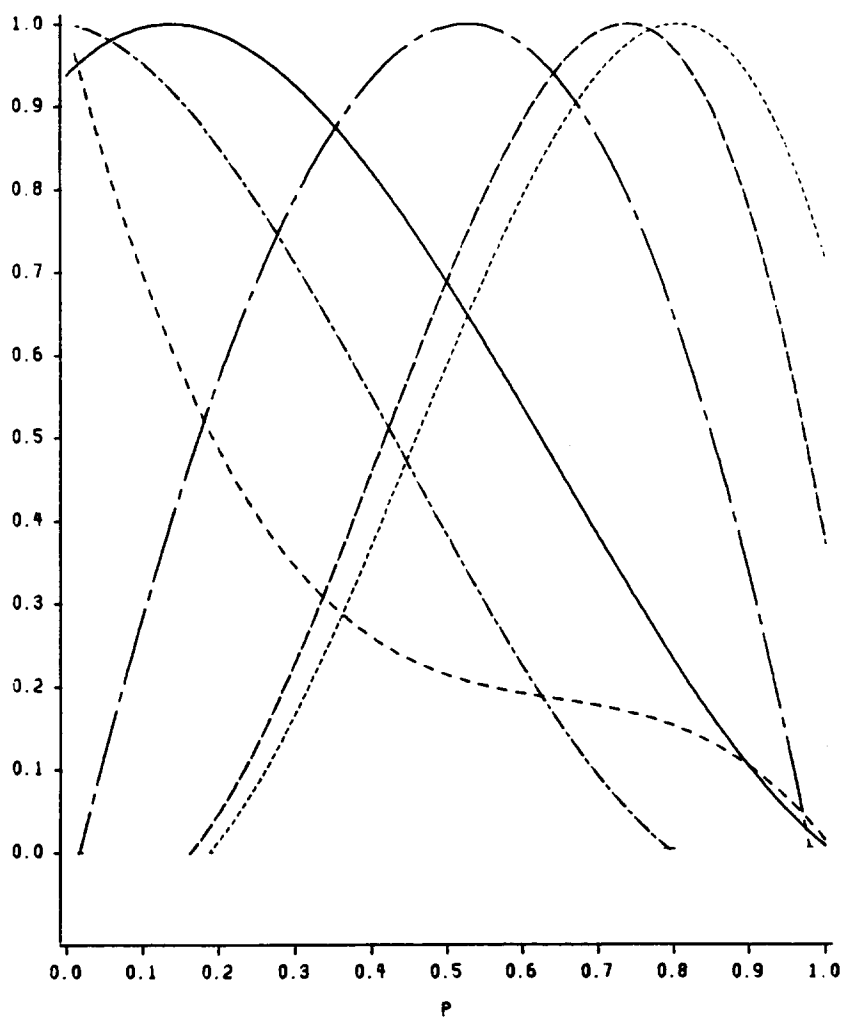
Figure 2. Membership Functions from a Single Subject

——— Doubtful
— — Good chance
— — Likely
-------- Fairly certain
----- Improbable
— - — Slight chance

agreed 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), and p_6 = (likely, good chance) are each composed of two "similar" words. Likewise, the subjects generally agreed 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), and 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.

The second level of our evaluation asked 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 has his or her 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 from the similarity between the subject's membership functions for the words in pair p_j. The degree of similarity within each pair is given, for that subject, by his or her "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 computed 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 was linear association. The same considerations applied to the "dissimilar" pairs. Here we computed 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 allowed for the possibility that a particular distance measure may be quite accurate in modeling fine variations in similarity (i.e., small distances) but quite inaccurate in modeling fine variation among pairs that are each composed of two "dissimilar" words. Furthermore, in practical applications one may need to find only a distance measure that is sensitive to the degree of similarity in pairs of "similar" words (as 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).
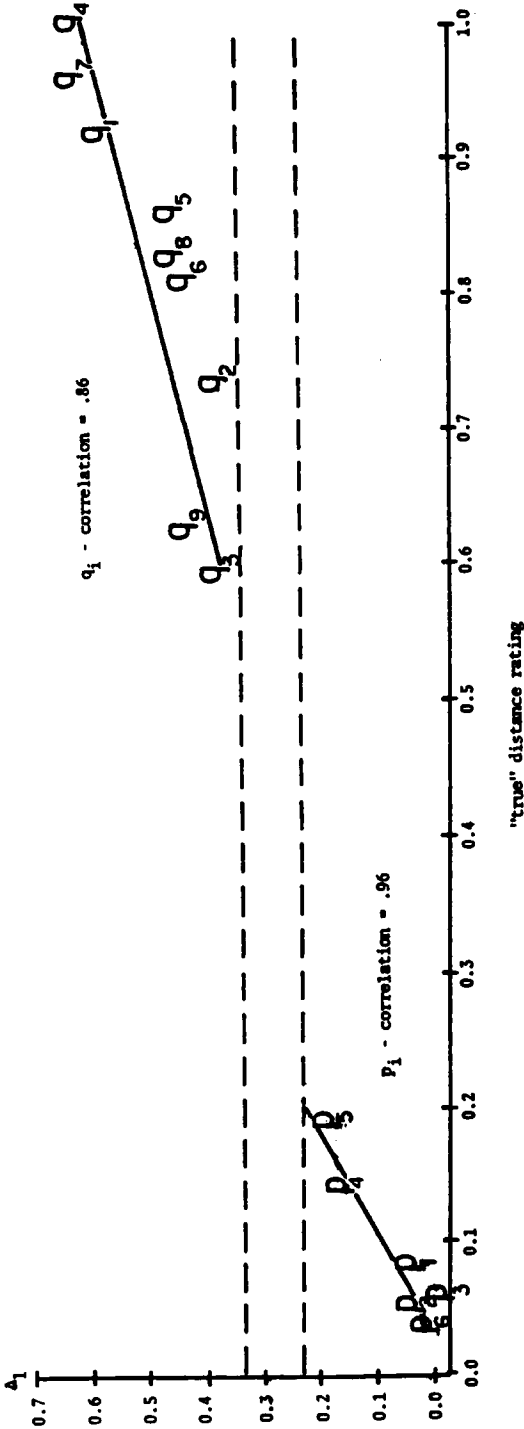
Figure 3. Plot of Distance Measure Δ_1 Versus "True" Distance Rating for Subject 5

p_i = "similar" words in pair

q_i = "dissimilar" words in pair

For each distance measure, its p_i-correlations for the 15 subjects were 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 corelations (i.e., interquartile range), and to be free of extremely low (i.e., negative) correlations.

Several trends are clear from these displays.

1. There is a great deal more variability between the performances of the various distance measures on "dissimilar" pairs (Figure 5) than on "similar" pairs (Figure 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 nine data points, while those for "similar" pairs are calculated from six data points.) This immediately suggests that more caution must be exercised when selecting a distance measure to distinguish between varying degrees of *dissimilarity*.

2. On the "dissimilar" pairs (Figure 5), those measures which perform the worst (d_2, $(d_2)^2$, d_1, d_∞, S_2, S_3, ρ) are measures that ignore the ordering on the x-axis (the base variable axis). Conversely, those measures which perform the best (q_∞, q_*, Δ_∞, Δ_*) are measures that 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 (unordered) spaces, but such measures have the disadvantage of being insensitive to varying degrees of dissimilarity (for instance, as in pairs q_i). In the "similar" pairs (Figure 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 which look at α-level sets (notably q_*, Δ_∞, Δ_*), and those which ignore the ordering on x (notably S_4).

3. Among those measures accounting for x-ordering (q_1, q_∞, q_*, Δ_1, Δ_∞, Δ_*, 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 six such measures, possibly accounting for the difference in performance.

4. Measure S_2 is arguably the worst both for "similar" pairs and for "dissimilar" pairs.

5. 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 *is* relevant in
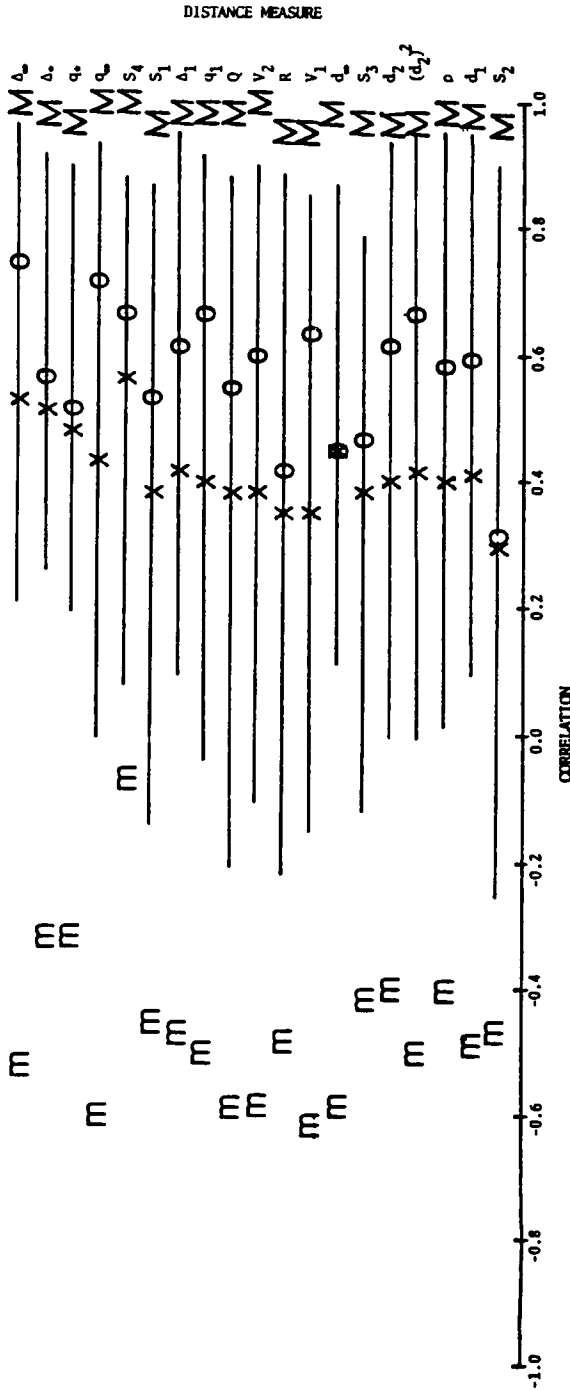
DISTANCE MEASURE

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 are aggregated over all 15 subjects: m = minimum correlation, M = maximum correlation, ——— = interquartile range of correlations, X = mean correlation, 0 = median correlation.
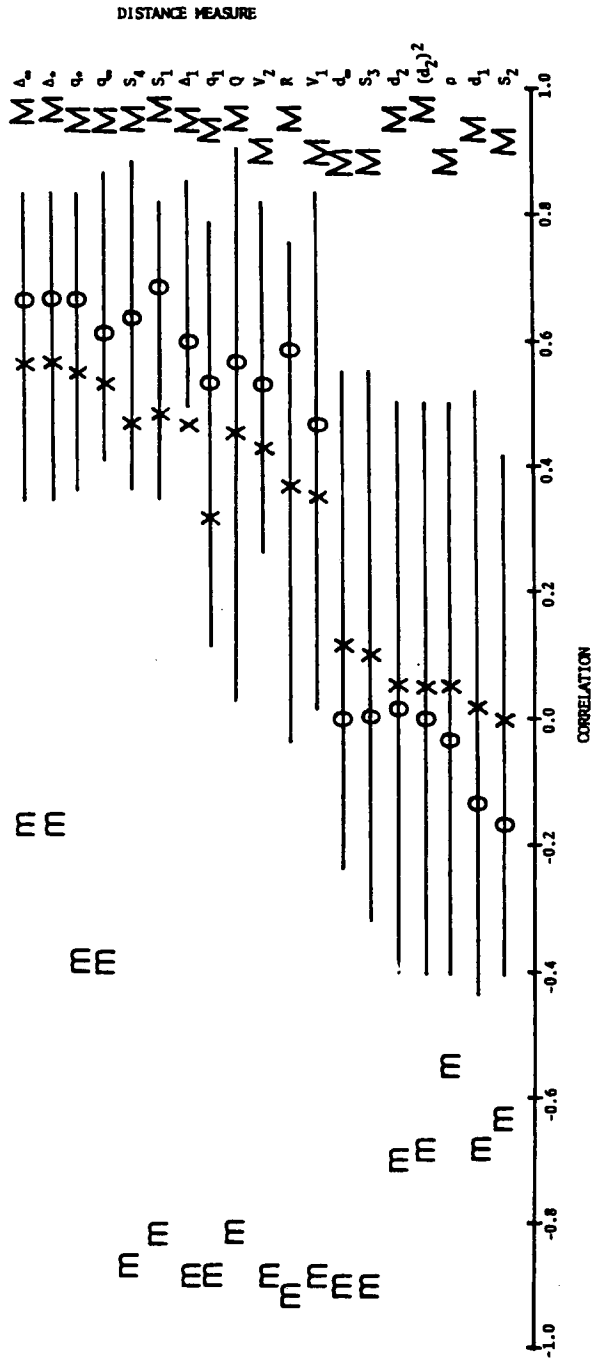
DISTANCE MEASURE

Figure 5.   Line Plots of q_i-correlation for Each Distance Measure.

this setting. Furthermore, measure S_4 performs reasonably well (among all measures) in the "similar" setting also.

6. Quite surprisingly, all the measures with consistently good performance (S_4, q_∞, q_*, Δ_∞, Δ_*) 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_∞ and Δ_∞, 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 (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.

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

8. 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 [3]). 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).

9. For a given subject, the relative rank of his or her correlation (within a line plot) tends to be consistent over all distance measures. (This fact is revealed by examining the individual subjects' correlations within each line plot.) For example, the q_i-correlation of subject 6 is the highest or second highest correlation within each of 18 line plots in Figure 5. At the low end, subjects 15 and 2 are responsible for 17 of the "minimum" q_i-correlations in Figure 5. An analogous situation exists in Figure 4, but interestingly, the particular subjects whose q_i-correlations are consistently the lowest (say) are *not* the same particular subjects whose p_i-correlations are consistently the lowest.

RECOMMENDATIONS

If one wants to select a distance measure that performs well in the long run on a broad spectrum of subjects, the aggregated data of our study may be used as a guide. Measures S_4, q_*, Δ_∞, and Δ_* consistently distinguished themselves with good performance.

If, however, the objective is to accurately model the behavior of a specific individual (for instance, 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). Moreover, there were some subjects with consistently low (negative) correlations, indicating that for them, none of the distance measures adequately models their thought processes in judging *degrees* of similarity (or dissimilarity). (This in no way detracts from the ability of all distance measures to successfully make gross categorizations of word pairs as "similar" or "dissimilar" for all subjects.) In contrast, for those subjects having consistently high (near 1.0) correlations, there is evidently a certain robustness with respect to the choice of a distance measure. In practice, then, it would be ideal to evaluate the performances of the various distance measures on 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.)

Having done this, one can determine which distance measure is the best for the situation at hand. If the individual attains consistently high correlations, then it does not matter which distance measure is used (perhaps computational convenience should then indicate the choice). If the individual shows much variability in his or her correlations, then of course the distance measure with the highest correlation should be chosen. If the individual produces consistently negative correlations, then this itself is an extremely important finding: it may be quite difficult to quantify and mathematically model the mental process of similarity judgment for such an individual.

In many cases the fuzzy concepts are unambiguously defined over a one-dimensional space (such as in our study of probability words). When this is not the case, then, in using those distance measures which 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. Attneave, F., Dimensions of similarity, *American J. Psychol.* 63, 516–556, 1950.

2. Beckenbach, E., and Bellman, R., *An Introduction to Inequalities*, Random House, New York, 1961.

3. Bonissone, P. P., A pattern recognition approach to the problem of linguistic approximation in system analysis, *Proceedings of the IEEE International Conference on Cybernetics and Society*, Denver, Col., 793–798, 1979.

4. Bush, R. R., and Mosteller, F., A model for stimulus generalization and discrimination, *Psychol. Rev.* 58, 413–423, 1951.

5. De Luca, A., and Termini, S., A definition of a non-probability entropy in the setting of fuzzy sets theory, *Info. Control* 20, 301-312, 1972.

6. Dubois, D., and Prade, H., Operations on fuzzy numbers, *Int. J. Syst. Sci.* 9(6), 613-626, 1978.

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

8. Dubois, D., and Prade, H., Fuzzy cardinality and the modeling of imprecise quantification, *Fuzzy Sets & Syst.*, 16, 199-230, 1985.

9. Eisler, H., and Ekman, G. A., A mechanism of subjective similarity, *Acta Psychologica* 16, 1-10, 1959.

10. Enta, Y., 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, 2980-2990, 1980.

11. Eshragh F., and Mamdani, E. H., A general approach to linguistic approximation, *Int. J. Man-Machine Stud.* 11, 501-519, 1979.

12. Giles, R., The practical interpretation of fuzzy concepts, *Human Syst. Mgt.* 3, 263-264, 1983.

13. Goetschel, R., and Voxman, W., Topological properties of fuzzy numbers, *Fuzzy Sets & Syst.* 10, 87-99, 1983.

14. Gregson, R. M., *Psychometrics of Similarity*, Academic Press, New York, 1975.

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

16. Hersh, H. M., Caramazza, A., and Brownell, H. H., 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, 389-408, 1979.

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

18. Kacprzyk, J., 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, 123-131, 1976.

19. Kailath, T., The divergence and Bhattacharyya distance measure in signal detection, *IEEE Transaction on Communication Technology* 15(1), 609-637, 1967.

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

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

22. Kochen, M., 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, 395–408, 1975.

23. Lashley, K. S., An examination of the "continuity theory" as applied to discriminative learning, *J. Gen. Psychol.* 26, 241–265, 1942.

24. Murthy, C. A., Pal, S. K., and Dutta Majumder, D., Correlation between two fuzzy membership functions, *Fuzzy Sets & Syst.* 17, 23–38, 1985.

25. Norwich, A. M., and Turksen, I. B., Stochastic fuzziness, in *Approximate Reasoning in Decision Analysis* (M. M. Gupta and E. E. Sanchez, Eds.), North Holland, Amsterdam, 13–22, 1982.

26. Norwich, A. M., and Turksen, I. B., A model for the measurement of membership and the consequences of its empirical implementation, *Fuzzy Sets & Syst.* 12, 1–25, 1986.

27. Nowakowska, M., Methodological problems of measurement of fuzzy concepts in the social sciences, *Behav. Sci.* 22, 107–115, 1977.

28. Oden, G. C., Fuzziness in semantic memory, Choosing exemplars of subjective categories, *Memory & Cognition* 5, 198–204, 1977.

29. Oden, G. C., Integration of fuzzy logical information, *J. Exp. Psychol., Human Percept. Perform.* 3, 565–575, 1977.

30. Oden, G. C., A fuzzy logical model of letter identification, *J. Exp. Psychol., Human Percept. Perform.* 5, 336–352, 1979.

31. Oden, G. C., Integration of fuzzy linguistic information in language comprehension, *Fuzzy Sets & Syst.* 14, 29–41, 1984.

32. Ralescu, A. L., and Ralescu, D. A., Probability and fuzziness, *Infor. Sci.* 34, 85–92, 1984.

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

34. Restle, F., A metric and an ordering on sets, *Psychometrica* 24, 207–220, 1959.

35. Thole, U., Zimmerman, H. J., and Zysno, P., On the suitability of minimum product operations for the intersection of fuzzy sets, *Fuzzy Sets & Syst.* 2, 167–180, 1979.

36. Tversky, A., Features of similarity, *Psychol. Rev.* 84, 327–352, 1977.

37. Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R., and Forsyth, B., Measuring the vague meaning of probability terms, *J. Exp. Psychol., General* 115, 348–365, 1986.

38. Wallsten, T. S., Fillenbaum, S., and Cox, J. A., Base rate effects on the interpretations of probability and frequency expressions, *J. Memory & Language* 25, 571–587, 1986.

39. Wallsten, T. S., Zwick, R., and Budescu, D. V., *Integrating Vague Information*, Paper presented at the 18th Annual Mathematical Psychology Meeting, La Jolla, Calif., 1985.

40. Wenstøp, F., Quantitative analysis with linguistic values, *Fuzzy Sets & Syst.* 4, 99–115, 1980.

41. Zadeh, L. A., The concept of a linguistic variable and its application to approximate reasoning, Parts 1, 2, 3, *Info. Sci.* 8, 199–249; 8, 301–357; 9, 43–98, 1975.

42. Zimmer, A. C., Verbal vs. numerical processing of subjective probability, in *Decision Making Under Uncertainty* (R. W. Scholz, Ed.), North Holland, Amsterdam, 159–182, 1983.

43. Zwick, R., A note on random sets and the Thurstonian scaling methods, *Fuzzy Sets & Syst.* 21, 351–356, 1987.

44. Zysno, P., Modeling membership functions, in *Empirical Semantics* vol. 12 (B. B. Rieger, Ed.), Brockmeyer, Bochum, West Germany, 350–375, 1981.