*Int. J. Man–Machine Studies* (1990) **33**, 657–676

# Integration of linguistic probabilities†

DAVID V. BUDESCU‡
*Department of Psychology, University of Haifa, Haifa 31999, Israel*

RAMI ZWICK
*Department of Marketing, 701-N Business Administration Building, The Pennsylvania State University, University Park, PA 16802, USA§*

THOMAS S. WALLSTEN AND IDO EREV
*Department of Psychology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA*

*(Received 5 April 1989 and accepted in revised form 5 October 1989)*

In a previous study, Zwick, Budescu and Wallsten (1988) found that the membership functions representing the subjective combinations of two independent linguistic probabilistic judgements could not be predicted by applying any dual *t*- and co-*t*-norm to the functions of the underlying terms. Their results showed further that judgements involving the "and" connective were best modelled as the fuzzy mean of the two separate components. The present experiment extended those results by manipulating the instructions regarding the "and" connective and also including an additional task in which subjects selected a third phrase to represent the integration of the two independent judgements. Again, no *t*-norm rule predicted subjects' responses, which were now best modelled by the point-wise arithmetic or geometric means of the functions. In addition, most subjects selected phrases and provided membership functions in response to two identical forecasts that were more extreme and more precise than the individual forecast, a result inconsistent with any *t*-norm or averaging model. A minority of subjects responded with the same phrase contained in the forecasts. The entire pattern of results in the Zwick *et al.* (1988) and the present study is used to argue against the indiscriminate application of mathematically prescribed, but empirically unsupported operations in computerized expert systems intended to represent and combine linguistic information.

## 1. Introduction

### 1.1. BACKGROUND

In many situations our own opinions or decisions depend on the probabilistic judgements and forecasts of external agents. For example, we depend on expert forecasts of interest rates and market fluctuations when selecting among investments. Thus our own decisions and choices are affected by the capability of others to properly express and communicate their beliefs, and by our ability to understand and use them. A well-known phenomenon in the forecasting literature (e.g. Beyth-Marom, 1982) is that many experts, like laypeople, prefer to communicate

† This research was supported by a grant from the USA–Israel Binational Science Foundation, and by a grant BNS86-08692 from the US National Science Foundation.

‡ This paper was completed while the first author was on sabbatical leave at the Department of Social and Decision Sciences at the Carnegie Mellon University.

§ All correspondence should be made to this address.



0020-7373/90/060657 + 20$03.00/0 

their opions by means of verbal probabilistic phrases rather than by means of numbers. Typically, this preference for words over numbers is linked to the vagueness of opinion, claiming that it would be misleading, improper, or even unethical to use a precise expression of a fuzzy opinion.

The preference for words over numbers has led many researchers to investigate the relationships between linguistic and numerical expressions of probabilistic opinions (see Budescu & Wallsten, 1985, 1987 for partial reviews). These studies show that most probabilistic phrases are understood to cover large ranges of probabilities and to overlap considerably. In addition, there exist consistent individual differences in the use and interpretation of these words (e.g. Beyth-Marom, 1982; Nakao & Axelrod, 1983; Budescu & Wallsten, 1985). Recently, Wallsten, Budescu, Rapoport, Zwick and Forsyth (1986) and Rapoport, Wallsten and Cox (1987) have shown that the vague meanings of probability phrases can be described reliably by means of membership functions over the [0, 1] interval.

The present investigation is a follow-up to the one by Zwick *et al.* (1988), and employs the theoretical framework of fuzzy set theory to evaluate various aggregation connectives (e.g. Dubois & Prade, 1985). Specifically, we are interested in how subjects combine the vague probability judgements of two observers or forecasters with regard to the likelihood of an event. These opinions are expressed linguistically and are independent (in the sense that the two observers do not communicate with each other prior to their expression of opinion). Further, there is no reason for the receiver to believe that one observer is more reliable than another. This situation was studied extensively from a prescriptive point of view (e.g. see Clemen, 1988 for a review) but little is known about the descriptive power of the fuzzy aggregation models and even less about the case when opinions themselves are expressed in an inexact linguistic manner.

### 1.2. FUZZY SET AGGREGATION CONNECTIVES

Given a generalized concept of membership, operations on sets are no longer restricted to the boolean binary algebra, and a much richer class of operations can be defined. Numerous definitions for the logical operators "**and**" and "**or**" have been suggested (Czogala & Zimmermann, 1984; Smithson, 1984, 1987; Dubois & Prade, 1985) on the basis of various normative, empirical, or purely mathematical considerations. Various definitions can be considered proper, in that they all yield the regular results for classical (crisp) sets as a special case, and this fact has led to confusion regarding which definitions are "correct". However, it is logically and mathematically impossible to determine the appropriate "general" rule on the basis of a special case, central as it may be. Indeed, since fuzzy set theory was offered as a more flexible model that is better suited for systems affected by human judgement, perception and emotion (Zadeh, 1975), it is a mistake to search for a single best set of definitions of the "**and**" and "**or**" operations. Zadeh's (1976) position, for example, has been that the choice of these definitions should reflect the unique characteristics of the particular situation at hand. However, it is important for both applications and basic theory to identify individual and situational factors that make more appropriate any given class of operators.

*1.2.1. Aggregation connectives based on triangular norms and their duals*

Dubois and Prade (1985) suggest that a consensus has formed in the literature that the concept of triangular norms and conorms proposed by Menger (1942) is appropriate for representing the fuzzy logical operators for "**and**" and "**or**" (see also Weber, 1983). Examples of $t$-norms and their duals that have been proposed in the literature to represent the pointwise fuzzy set–theoretic intersection or union are:

$$T^{(1)} = \min(a, b), \qquad S^{(1)} = \max(a, b) \quad \text{(Zadeh, 1965)} \tag{1}$$

$$T^{(2)} = ab, \qquad S^{(2)} = a + b - ab \quad \text{(Bellman \& Zadeh, 1970)} \tag{2}$$

$$T^{(3)} = \max(0, a + b - 1), \qquad S^{(3)} = \min(1, a + b) \quad \text{(Bellman \& Zadeh, 1977)} \tag{3}$$

where $T^{(i)}$ is a particular $t$-norm and $S^{(i)}$ is its corresponding conorms.

Smithson (1984; 1985) pointed out that these three classes can be thought of representing various degrees of extremity since for all $(a, b)$ in $[0, 1] \times [0, 1]$

$$T^{(1)} \geq T^{(2)} \geq T^{(3)}$$

and

$$S^{(1)} \leq S^{(2)} \leq S^{(3)}.$$

It was therefore suggested that the three classes can be incorporated into one general family of connectives with one or more free parameters. For an excellent review of the class of fuzzy set aggregation connectives based on triangular norms see Dubois and Prade (1982, 1985) and for specific examples of general rules see Yager (1980); Czogala and Zimmerman (1984) and Smithson (1984, 1985).

Little psychological understanding of the parameters can be claimed. It is not clear whether one or more of them might be descriptive of subjective union or interaction, and if so how they depend on individual or situational factors, how stable or reliable they are, and what other factors may affect them. In fact, estimation procedures for most of the parameters have not been developed. However the $t$-norm conceptualization is useful, since it allows us to test simultaneously the appropriateness of a large class of possible dual connective operators.

*1.2.2. Aggregation connectives based on compensatory operators*

Another class of operators is based on the notion that in everday life people rarely use "**and**" and "**or**" in their respective strict noncompensatory and fully compensatory senses. Rather, people's judgements are based on a partially compensatory interpretation, represented by a free parameter, $g$, which can be called "grade of compensation". The first to introduce this idea were Zimmermann and Zysno (1980) who suggested a weighted geometric mean:

$$\mu_{A\&B}(X) = [\mu_{A\cap B}(X)]^{1-g}[\mu_{A\cup B}(X)]^{g} \tag{4}$$

Here $(A \,\&\, B)$ denotes the generalized compensating connective, which varies between the regular "**or**" when $g = 1$ and the regular "**and**" when $g = 0$. Note, that this approach does not eliminate the need to decide on a definition of the strict "**or**" and "**and**". Thus, any of the operations described earlier can be used. In a similar fashion Luchandjula (1982), and Smithson (1984) proposed different forms of generalized connectives that include the pure union and intersection formulas as special cases. In a recent development Dyckhoff and Pedrycz (1984) suggest the use of generalized means to model the compensatory connective. This model yields a variety of simple well know rules as special cases. Among them are the minimum, maximum, as well as the weighted arithmetic, geometric and harmonic means of the membership functions.

The importance of this approach is that it offers a definition of a compensatory connective without explicitly adopting a set of definitions for the strict "**and**" and "**or**". Although this approach simplifies the situation conceptually and computationally, it is questionable whether the Dyckhoff and Pedrycz model is comparable with the others. It appears that this model should be considered a general model of combination and aggregation of levels of membership.

### *1.2.3. Aggregation connective as a fuzzy mean*

Other forms of averaging the two individual membership functions can also be considered as models of aggregation. One deserving particular mention is the "fuzzy mean", i.e. the mean of two fuzzy numbers. Dubois and Prade (1980) define a real fuzzy number as a fuzzy subset with a continuous membership function satisfying some mild regularity conditions, and the mean of two fuzzy numbers, $A$ and $B$, as:

$$\mu_{\text{Mean}(A,B)}(x) = \max_{x=(y+z)/2} [\min(\mu_A(y), \mu_B(z))]. \tag{5}$$

## 1.3. PREVIOUS EMPIRICAL RESULTS

Norwich and Turksen (1982, 1984), Wallsten, *et al.* (1986) and Zwick (1987) have pointed out the relationships between the axiomatic formulation of the algebraic difference (ratio) structure (e.g. Krantz, Luce, Suppes and Tversky, 1971) and the measurement of memberships functions. This approach was recently refined and successfully tested by Wallsten, *et al.* (1986), who developed a graded pair-comparisons procedure (Oden 1977*b*), that allows simultaneous testing of the necessary axioms, scaling of the responses in order to obtain memberships, and tests of goodness of fit. In a subsequent study Rapoport *et al.* (1987) demonstrated a high level of similarity between membership values determined through graded pair-comparison and direct magnitude estimation. Thus, a sound theoretical and empirical justification was established for the quantification of the vague meanings of linguistic terms by means of direct scaling.

Several studies have tested empirically the various combination rules. Hersch and Caramazza (1976) tested the max rule (Equation 1) for the union operator. The shape of the "**or**" function resembled the max rule, but was consistently lower. Oden (1977*a*, 1979) contrasted the min–max and sum-product operators for the two connectives. Group and individual analyses indicated that both classes of operators fit the data quite well, but the sum-product connectives clearly outperformed the min–max rules. Thöle, Zimmerman and Zysno (1979) compared intersection operators. Their group analyses slightly favored the min rule over the product operator.

Zimmermann and Zysno's study (1980) has the distinction of being the most realistic. Subjects were presented with exemplars of fire resistant tiles whose quality was to be judged according to solidity and dovetailing. Group membership functions were obtained for "solidity", "dovetailing", and "ideal tile". The subjects' judgements of "ideal" were best represented as a combination of the two membership functions that did not fit any of the rules proposed for conjunctions and disjunction. The authors ultimately advocated the generalized operator described earlier (Equation 4).

To summarize, the evidence regarding the "**and**" and the "**or**" operations is inconclusive. Oden (1977*a*; 1979) advocates the sum-product rule; Thöle *et al.* (1979) favors the min rule for "**and**", and Hersh and Caramazza report (at least weak) support for the max rule for "**or**". Furthermore, given the various methods used to quantify the degrees of membership and the variety of experimental procedures used to elicit the judgements, it is impossible to determine whether the different results are really contradictory, or can be linked to the use of different methodologies.

Finally, Zimmermann and Zysno's (1980) work clearly illustrates that information is sometimes combined according to an averaging rule (e.g. Anderson, 1981; 1982) that allows for compensation between the extreme alternatives. Their results are intriguing. They suggest that although the duality between the union and intersection operations, and the corresponding "**or**" and "**and**", may reflect a nice mathematical normative structure (imposed by DeMorgan's theorem), they do not necessarily reflect a psychological reality.

### *1.3.1. The experiment of Zwich et al. (1988)*

Zwick *et al.* (1988) performed an experiment in order to investigate this possibility. The experiment tested a series of requirements which must hold if the dual rules for intersection and union are based on *t*-norms and their corresponding co-*t*-norms. Let $T(a, b)$ stand for the *t*-norm of two values, and $S(a, b)$ be its dual co-*t*-norm. Then under any of the dual rules (e.g. 1, 2 and 3) the following three conditions must hold:

(i) $T(a, b) \leq \min(a, b) \leq \max(a, b) \leq S(a, b)$
(ii) $T(0, 0) = 0$; $T(a, 1) = S(a, 0) = a$ and $S(1, 1) = 1$ (boundary conditions)
(iii) $T$ and $S$ are monotonic in both arguments

In addition, under any strict dual rule,

(iv) $T$ and $S$ are strictly increasing in both arguments.

Finally if $T$ belongs to the Frank family (Frank, 1979) then:

(v) $T(a, b) + S(a, b) = a + b$.

In the study of Zwick *et al.* (1988) 16 graduate students performed three judgemental tasks in two experimental sessions. In one task, they estimated membership functions of individual probability phrases, in another they estimated membership functions of pairs of phrases connected by "and", and in the third they estimated membership functions of pairs of phrases connected by "or". Three low probability phrases (*doubtful, slight chance,* and *improbable*) and three high probability ones (*likely, good chance,* and *fairly certain*) were used separately and in all possible pairs. Individual phrase membership functions were assessed for 11 probabilities equally spaced across the 0·20–0·98 range. Membership functions for the pairs connected by "**and**" and "**or**" were obtained for five probabilities each.

The data obtained were used to test the five properties of the dual rules, and to identify the best models for the judgements. The monoticity and boundary conditions were confirmed, but the others were not. Specifically, for all subjects, $T(a, b) + S(a, b) > a + b$, in violation of condition (v). This result was mainly due to the fact that in a majority of cases $T(a, b) > \min(a, b)$, in clear violation of requirement (i).

Thus, the results show that no dual *t*- and co-*t*-norm could fit the empirical data, and that the violations are due mainly to the "**and**" operator. This conclusion was further confirmed in the metric comparison of the models, where the max rule best described the "**or**" judgements for 15 of the 16 subjects, but "averaging" models yielded the best description of the "**and**" judgements. Specifically, the judgements of a majority of the subjects (9 out of 16) were best captured by the fuzzy mean model and the remaining were better described by simple pointwise arithmetic or geometric means.

In previous research (Hersh & Caramazza, 1976; Hersh, Caramazza & Brownell, 1979; Thöle *et al.*, 1979; Oden, 1977*a*, *b*, 1979; Zimmermann & Zysno, 1980; Kulka & Novak, 1984) it was found that the logical (fully noncompensatory) "**and**" operator does not describe the linguistic "and" as applied to two linguistic categories referring to different scales (e.g. a *high brewing speed* and *low capacity coffeemaker*). In these cases "not only are the idempotent intersection and union not a good model, but one may drift out of the domain of triangular norms and conorms, and get operations such as the arithmetic or the geometric means as proper models" (Dubois & Prade, 1985, p. 97). Zwick *et al.* (1988) confirmed this phenomenon, with two linguistic categories referring to the same scale (i.e. probability).

Zwick *et al.* (1988) suggested that when subjects judge whether a probability value is well described by *at least one* expert (the "or" task), they effectively decide which expert had come closest to the value and then ignore the input from the other person. In the more common situation, where the individual integrates verbal probabilities inputs to form a single overall judgement, he or she uses an averaging process. Consequently, the final judgement may be very different from either of the two inputs, just as in many other situations when integrating precise values (e.g. Anderson, 1981, 1982).

A possible explanation of this somewhat surprising result is that it was an artifact induced by the instructions to the subjects. Perhaps the description of the task was understood as a call for a "compromise" between the two forecasts, instead of an integration.

#### *1.3.2. The new experiment*

The primary goal of this study is to replicate the results of Zwick *et al.* (1988) and to test the possibility that they were caused by the particular instructions. To this end the "And" task was run under two different sets of instructions, and the number of judgements was increased to obtain more stable data. In addition, a task was included in which the subjects were explicitly required to select a single phrase (possibly, but not necessarily, one of the two presented to them) that best described the integrated meaning of the two forecasts.

The primary analysis attempts to identify the model that fits best the subjects' responses in the various integration tasks, given their membership functions for the individual forecasts. The six models we examine are the three operators suggested in the fuzzy sets literature (min, product and bounded sum), two pointwise means (arithmetic and geometric) and the fuzzy mean. To keep the comparisons simple and meaningful, no model involving free parameters is used.

An interesting special situation occurs when the subject is presented with two identical forecasts by two independent agents judging the same information (e.g. two radiologists who look at the same X-Ray and both tell you it is "likely" that you have a tumor). With numerical forecasts the subject can either adopt the unanimity principle (Morris, 1983; Winkler, 1986) stating that if the two forecasts coincide the combined value should be equal to this common probability, or combine the two values in a Bayesian analysis yielding a probability higher than each of the forecasts (Lindley, 1986; Winkler, 1986). In the verbal case the three types of means (arithmetic, geometric and fuzzy) and the min rule all predict the unanimity principle, while the product and bounded sum rules both predict uniformly flatter functions for the integrated forecast than for each of the individual forecasts. A third possibility, which is more consistent with the Bayesian analysis, but which is not predicted by any of the models considered here is that the integrated forecast is more precise and extreme than the original one. A secondary goal of the present study is to characterize how people integrate two independent identical forecasts.

## 2. Method

### 2.1. SUBJECTS

Twenty graduate students in social sciences and business at the University of North Carolina at Chapel Hill were recruited by placing notices advertising the experiment in the departmental mailboxes. They received $25 for participation in all three sessions of the experiment.

### 2.2. GENERAL PROCEDURE

All subjects participated in one practice session and two data collection sessions in which they performed three tasks: (i) estimation of membership functions of individual probability phrases; (ii) estimation of membership functions for pairs of phrases connected by "and", which we call the "And" task; and (iii) selection of a single phrase, which we call the "Selection" task. The experiment was controlled by an IBM/PC computer.

#### *2.2.1. Membership estimation*

The instructions for this task read in part:

> You are to imagine that you are to predict whether a spinner you cannot see 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 will use a non-numerical probability phrase to give you his or her best opinion about the chances of the spinner landing on white. This gives you some basis for judging the probability of that event. Thus a phrase (from your friend), a spinner, and a response line will come on the screen for a single trial. You are to indicate how close the displayed probability of landing on white is to the judgement you would form upon hearing the particular phrase. If the

> displayed probability is not all close to your judgement, move the arrow (on the response line) all the way to the left. If it is as close as possible to your judgement, move the arrow all the way to the right. If the displayed probability matches your judgement to some degree, then place the arrow accordingly.

Subjects responded by pushing the arrow buttons of the keyboard. The precision of the response was determined by the resolution of the monitor, and allowed 201 locations anchored by two verbal labels: "not at all" on the left, and "absolutely" on the right.

### 2.2.2. *"And" task*

We used two sets of instructions for this task. Both started with the following description:

> Two friends have viewed the rotating spinner. They have not spoken to each other, so neither knows what the other thinks, but each uses a non-numerical phrase to tell you his or her opinion about the chances of the spinner landing on white. They may use the same or different phrases.
>
> Now we are interested in how well the displayed probability of landing on white was described by both of your friends. Thus, two phrases (one from each friend), a spinner (perhaps the one they saw, perhaps not), and a response line will appear on the screen for a single trial.

In one set of instructions (Group A), identical to those used by Zwick *et al.* (1988), the subject's task was described as the following:

> The question is, to what degree is the displayed probability simultaneously consistent with the judgements of friend 1 and friend 2? As before, move the cursor all the way to the left for probabilities that are not at all consistent with both judgement 1 and judgement 2, and all the way to the right for probabilities that are as consistent as possible with both judgement 1 and judgement 2. If the displayed probability matches the two judgements to some intermediate degree, place the arrow accordingly.

In the second set of instructions (Group B) the subjects were told:

> On the basis of the two phrases, you can form some judgement about the probability of landing on white for the spinner your friends saw. The question is, how close does the displayed probability of landing on white come to the judgement you formed based on the two phrases? As before, . .

Thus, these instructions explicitly called on the subject to create his or her own estimate on the basis of the two forecasts.

### 2.2.3. *Selection Task*

The instruction for this task read in part:

> We will not present any spinners at all in this type of trial. Rather, two non-numerical probability phrases will be printed at the top of the screen on each trial, each phrase representing the chances that the spinner will land on white as was told to you by two friends. Now imagine that on the basis of this information you have to convey the probability of the spinner landing on white to a fourth person, using a word rather than a probability number.
>
> At the bottom of the screen, there is a list of probability words on the right, and a list of modifiers on the left. You must respond with a probability word, and you may also add one or two modifiers before the word. To indicate your response, simply type in the appropriate combination.

The list of possible responses included seven phrases (*good chance, likely, probable, doubtful, unlikely, improbable,* and *slight chance*) and nine modifiers (*very, not, quite, rather, fairly, highly, somewhat, extremely,* and *pretty*).

#### 2.2.4. *Stimuli*

Six phrases were used, three (*fairly certain, likely,* and *good chance*) representing probabilities higher than 0·5(H), and three (*doubtful, improbable,* and *slight chance*) representing probabilities lower than 0·5(L). Nine probabilities were employed, with values randomly selected on a given trial from nine uniform distributions of values over the ranges $p_i \pm 0{\cdot}03$, where $p_i = 0{\cdot}1, 0{\cdot}2, \ldots, 0{\cdot}9$. This particular technique was employed in order to minimize the effects of replication of the same nine probabilities across all trials. Membership judgements were obtained for each of the six phrases alone at all nine probabilites (a total of $6 \times 9 = 54$ judgements), and for each of the ($6 \times 7/2 =$) 21 combinations of phrases at all nine probabilites (a total of $21 \times 9 = 189$ judgements). The selection task was performed for each of the ($6 \times 6 =$) 36 pairs of phrases. Finally, membership judgements were also obtained at the nine probabilities for all new words selected by the subject. The number of these judgements varied across subjects according to their selections.

### 2.3. PROCEDURE

Subjects were randomly allocated to two groups, differing only in terms of the instructions for the "And" task. During the first session all three types of judgements were explained and practised, but no data were collected. All judgements were performed in the second session and then replicated in the third. The order of the tasks and the order of stimuli presentation within each session and task was randomly determined. Also, in the "And" and "Selection" tasks the left/right location of the two words was randomly counterbalanced for each subject. All subjects completed the three sessions within two weeks and were paid at the conclusions of the last one. Because of a computer malfunction the data from the last session of one subject were lost.

## 3. Results

All the results are reported at the individual subject level, because there is no logical or theoretical reason to expect the same integration rule to apply universally. In most cases we report separately data in response to pairs of similar (LL and HH) and mixed (LH) terms. In the former case the subjects must integrate two congruent pieces of information, while in the latter the two forecasts are inconsistent. Thus, we must consider the possibility that the subjects employ different integration rules, and that our models have differential predictive validity in the two cases. Prior to the analysis all judgements were re-expressed on a scale ranging from 0–1.

In summarizing the results we used the Mean Absolute Deviation (MAD) as the primary criterion of goodness of fit. It is simple, straight forward and easy to interpret, especially when dealing with results bounded, by the response scale, in the [0, 1] interval. This measure was selected over more commonly used statistics, such as the mean squared deviation or the product moment correlation, because its properties fit our goal: MAD measures the absolute closeness between the observed and the predicated values while disregarding the direction and pattern of the deviations. Also, many of the membership functions to be compared are relatively flat, and under such circumstances the regular correlation would be misleading.

### 3.1. STABILITY

Table 1 presents MADs between repeated judgements as a measure of the responses' stability across the two sessions. In this, and all the other tables, a value of 0 indicates a perfect result (stability or fit), and a value of 1 indicates the worst possible results. Since stability measures are typically reported as correlations, and in order to allow the reader to form an impression regarding the relationship between MAD and the usual measure of reliability, we also report in this table the regular correlations for each case in parentheses.

The first column is based on the judgements of the membership functions of the six phrases and it indicates a good level of stability for most subjects. The next two columns summarize the results of the "And" task, for the two types of pairs separately. Note that the stability of the responses is about equal for the two groups,

TABLE 1
*Mean absolute deviations (and correlations) between repeated judgements in the various tasks (Decimal points omitted)*

| Group | Subject | Membership judgements | "And" (LL + HH) | "And" LH | Selection (LL + HH) | Selection LH |
|---|---|---|---|---|---|---|
| A | 1 | 16(73) | 16(75) | 25(24) | 15(78) | 21(32) |
| | 2 | 07(95) | 09(93) | 13(61) | 09(90) | 14(70) |
| | 3 | 13(87) | 14(84) | 16(09) | 11(88) | 20(40) |
| | 4 | 18(73) | 11(89) | 11(37) | 12(83) | 18(38) |
| | 5 | 08(94) | 09(92) | 20(39) | 09(89) | 20(36) |
| | 6 | 09(90) | 14(82) | 11(00) | 10(88) | 21(34) |
| | 7 | 07(96) | 08(95) | 10(87) | 08(91) | 15(27) |
| | 8 | 08(94) | 07(95) | 08(63) | 08(91) | 36(25) |
| | 9 | 11(87) | 16(75) | 20(14) | 12(84) | 15(35) |
| | 10 | 11(90) | 08(42) | 12(−11) | 11(84) | 25(−13) |
| | Mean | 11(88) | 11(82) | 15(32) | 10(86) | 20(32) |
| B | 1 | 08(94) | 14(78) | 13(31) | 14(80) | 18(31) |
| | 2 | 11(94) | 14(88) | 17(40) | 14(81) | 25(53) |
| | 3 | 15(70) | 19(59) | 13(12) | 18(26) | 24(−05) |
| | 4 | 12(85) | 08(92) | 18(51) | 12(80) | 26(32) |
| | 5 | 12(90) | 11(92) | 20(44) | 12(85) | 27(45) |
| | 6 | 10(90) | 11(87) | 13(59) | 10(86) | 25(02) |
| | 7 | 14(77) | 15(71) | 10(53) | 15(67) | 15(30) |
| | 8 | 08(95) | 09(92) | 19(69) | 07(96) | 25(37) |
| | 10 | 09(93) | 11(90) | 15(41) | 12(84) | 20(12) |
| | Mean | 11(87) | 12(83) | 15(44) | 13(76) | 23(26) |
| Mean | | 11(88) | 12(82) | 15(38) | 11(82) | 21(30) |

Note; Data from second session of Subject 9 in group B are missing.

and that most subjects are slightly more consistent in judging similar (i.e. LL and HH) than dissimilar (LH) pairs. Generally, the stability is only slightly under the level recorded for the simple membership functions judgements.

The last two columns display the results from the selection task. It is important to keep in mind that in most cases these data reflect similarity between potentially different phrases. For example, for a given pair of terms ($w_1$ and $w_2$) a subject might have selected $w_3$ in Session 1 and $w_4$ in Session 2. (In fact, the proportion of cases in which the same phrase was selected for a given pair is a relatively low 0·25). A quick glance at this part of the table shows that despite this factor the judgements for the similar words are stable and at a level of consistency close to that achieved for the other tasks. However, the judgements for the incongruent pairs are less consistent than for the congruent ones (on the average twice as different). This result indicates that this task was much more difficult and all subsequent analyses based on these data must be interpreted with great caution. It is also interesting to note that the stability of judgements in the "And" and the "Selection" tasks is very similar for the congruent terms ($r = 0{\cdot}82, p < 0{\cdot}05$), but not for the incongruent ones ($r = 0{\cdot}003, p > 0{\cdot}05$).

### 3.2. FIT OF THE MODELS

All analyses in this section are based on the average of the two judgement per stimulus over the two sessions for each person. These analyses use the individual membership functions to predict the functions representing the integrated judgements. We compare the min, product and bounded sum rules (Equations 1–3, respectively), two pointwise means (the arithmetic and the geometric) and the fuzzy mean (Equation 5) for combining the separate functions. The fit of the various models can be easily compared, as they are all parameter-free and simply different formulas for combining the membership functions of the individual terms. A preliminary inspection of the data shows that the geometric and arithmetic mean are almost always the best two models. Therefore, we present in the following tables a complete summary of their fit. In those cases where one of the other four models is the best, or the second best, we also present its fit.

### 3.3. THE "AND" TASK

Tables 2 and 3 summarize the two best fitting models for the similar (LL and HH) and mixed (LH) pairs of terms, respectively, and indicate when other models are equally good or better for a particular subject. Not surprisingly, the MADs are much better for the similar pairs (Table 2). If Subject 10 from Group A whose data are not well described by any model is excluded, the average MAD of the geometric and arithmetic mean is in this case, an impressive 0·07, which is as good as one can expect given the stability of the judgements. The fit is worse, yet satisfactory, in the case of incongruent pairs (Table 3). However, even in this case the average MAD of the best model for each subject is 0·12. For the HH and LL pairs the two types of means are equally good predictors as reflected in their average MAD and in the number of best fitting models (eight subjects best fitted by the arithmetic mean and nine by the geometric one). For the HL pairs the geometric mean seems to be doing a slightly better job in terms of its overall fit and the

TABLE 2
*Mean absolute deviation from the best two models for the "and" task (LL + HH Pairs) (Decimal points omitted)*

| Group | Subject | Arithmetic mean | Geometric mean | Other model |
|---|---|---|---|---|
| A | 1 | 10 | 10 | 09† (Fuzzy) |
| | 2 | 05† | 05 | — |
| | 3 | 07 | 07† | — |
| | 4 | 09 | 08 | — |
| | 5 | 06 | 06 | 06† (Min) |
| | 6 | 05† | 06 | 05 (Fuzzy) |
| | 7 | 08 | 07† | 07 (Min) |
| | 8 | 06 | 05† | — |
| | 9 | 10† | 10 | — |
| | 10 | 47† | 47 | — |
| | Mean | 10† | 11 | |
| B | 1 | 08† | 08 | — |
| | 2 | 08 | 08† | — |
| | 3 | 15† | 15 | — |
| | 4 | 07 | 07† | — |
| | 5 | 08 | 07† | — |
| | 6 | 08 | 07† | 07† (Fuzzy) |
| | 7 | 08 | 08 | 08† (Fuzzy) |
| | 8 | 05 | 06 | — |
| | 9 | 08† | 08 | 08 (Fuzzy) |
| | 10 | 08 | 08† | |
| | Mean | 07† | 08 | |
| Mean | | 09† | 10 | |

† Best fitting model.

number of cases in which it is best (12 *vs* 5 for the arithmetic mean). Finally, note that only for nine subjects (45%) is the same model best for both types of pairs (in six cases the geometric mean, in three the arithmetic one), and that the more reliable subjects seem to have better levels of fit (the correlation between the stability and fit of the best model $r = 0{\cdot}6$, $p < 0{\cdot}05$).

### 3.4. THE SELECTION TASK

Tables 4 and 5 display the two best fitting models as well as others that approach them for each subject when presented with two congruent and incongruent terms, respectively. As in the previous task the MADs are excellent, and equally good for both types of means when the words are similar (Table 4). On the other hand, neither model fits well when the words are different (Table 5). In this case arithmetic mean is the best model for 16 subjects (80%) but its fit is poor. Only seven subjects (35%) are best described by the same model for the two types of pairs (five by arithmetic and two by geometric means). Finally, there is no indication that the more reliable subjects can be better modelled (the two statistics correlate only $r = 0{\cdot}12$, $p > 0{\cdot}05$).

TABLE 3
*Mean absolute deviation from the best two models for the "**and**" task (LH Pairs) (Decimal points omitted)*

| Group | Subject | Arithmetic mean | Geometric mean | Other model |
|---|---|---|---|---|
| A | 1 | 16 | 14† | — |
| | 2 | 13† | 20 | — |
| | 3 | 13 | 10† | — |
| | 4 | 14† | 14 | — |
| | 5 | 17† | 20 | — |
| | 6 | 30 | 06† | 06 (Min) |
| | 7 | 21 | 13 | 12† (Min) |
| | 8 | 38 | 06† | 07 (Min) |
| | 9 | 16 | 11 | 11† (Min) |
| | 10 | 32 | 19 | 08† (Product) |
| | Mean | 21 | 13† | |
| B | 1 | 09† | 14 | — |
| | 2 | 13† | 24 | 22 (Fuzzy) |
| | 3 | 10 | 10† | — |
| | 4 | 27 | 21† | 21 (Fuzzy) |
| | 5 | 22 | 15† | 21 (Fuzzy) |
| | 6 | 11 | 11† | — |
| | 7 | 09 | 08† | — |
| | 8 | 28 | 29 | 15† (Fuzzy) |
| | 9 | 11† | 13 | — |
| | 10 | 13 | 12† | — |
| | Mean | 15† | 16 | — |
| Mean | | 18 | 15† | |

† Best fitting model.

### 3.5. COMPARISON BETWEEN GROUPS AND TASKS

In order to evaluate the effect of the instruction on the results reported by Zwick *et al.* (1988), two groups of subjects were run in this study with different sets of instructions. It is apparent in Tables 2 and 3 that the groups do not differ in either the overall goodness of fit of the various models or the proportion of cases in which a certain model is best.

An additional issue that merits some attention is the degree of correspondence between the "And" and the "Selection" tasks. Several analyses point to dissimilatities between them. First, as should be evident from Tables 2–5, different models are best for the two tasks. The geometric mean is, on the average, better for the "And" task and the arithmetic mean is better for the "Selection" one. In fact, it is impossible to predict reliably the best fitting model for one task on the basis of the results of the other. Second, there is no significant correlation between the fit of the best models in the two tasks ($r = 0{\cdot}33$, $p > 0{\cdot}05$ for the mixed pairs, and $r = 0{\cdot}12$, $p > 0{\cdot}05$ for the congruent ones).

Finally, it was impossible to predict the phrase selected by the subject on the basis of his or her membership judgements in the "And" task. Recall that phrases were

TABLE 4
*Mean absolute deviation from best two models for the section task (HH + LL pairs) (Decimal points omitted)*

| Group | Subject | Arithmetic mean | Geometric mean | Other model |
|---|---|---|---|---|
| A | 1 | 11† | 11 | 11 (Fuzzy) |
| | 2 | 06 | 06† | — |
| | 3 | 10 | 10 | 10† (Min) |
| | 4 | 13 | 12 | 12 (Min) |
| | 5 | 08 | 08 | 07† (Fuzzy) |
| | 6 | 04 | 04 | 03† (Fuzzy) |
| | 7 | 05† | 06 | 05 (Fuzzy) |
| | 8 | 10 | 09 | 07† (Fuzzy) |
| | 9 | 04 | 04† | — |
| | 10 | 08 | 08 | 07† (Fuzzy) |
| | Mean | 08 | 08 | |
| B | 1 | 08† | 08 | — |
| | 2 | 05† | 05 | — |
| | 3 | 13 | 13† | — |
| | 4 | 08† | 08 | — |
| | 5 | 10† | 10 | — |
| | 6 | 04 | 04† | — |
| | 7 | 09 | 09† | 09 (Fuzzy) |
| | 8 | 02 | 02† | — |
| | 9 | 03† | 03 | — |
| | 10 | 07 | 07 | 07† (Fuzzy) |
| | Mean | 07 | 07 | |
| Mean | | 07 | 07 | |

† Best fitting model.

selected for 36 pairs of terms. On the average subjects selected 18 different phrases in each session, for which the membership functions were subsequently elicited. One would expect the membership function of the selection phrase to be very similar to the function obtained in the "And" task for the same pair. However, the MAD between the two functions was no greater on average than the MAD between function for the "And" task and the functions for the phrases selected in response to other pairs.

### 3.6. INTEGRATION OF TWO IDENTICAL TERMS

This section focuses on the case where the two forecasts presented to the subjects were identical. Although six phrases were employed in the study the analysis is based on only five of them (the phrase *fairly certain* was excluded because the selection procedure did not allow its generation as a response so the unanimity principle could not be tested for this phrase). Table 6 summarizes some of the results of the two tasks for this special case. The first column in the table presents the proportion of cases (out of 10) in which the unanimity principle was supported in the selection task, i.e. the proportion of cases in which the subjects used the

TABLE 5
*Mean absolute deviation from best two models for the selection task (LH pairs) (Decimal points omitted)*

| Group | Subject | Arithmetic mean | Geometric mean | Other model |
|---|---|---|---|---|
| A | 1 | 17† | 19 | — |
| | 2 | 24† | 28 | — |
| | 3 | 14† | 20 | — |
| | 4 | 23† | 35 | 25 (Fuzzy) |
| | 5 | 21† | 26 | — |
| | 6 | 26† | 40 | — |
| | 7 | 22† | 28 | — |
| | 8 | 27† | 39 | 37 (Fuzzy) |
| | 9 | 14 | 13† | — |
| | 10 | 13† | 18 | — |
| | Mean | 21† | 28 | — |
| B | 1 | 14† | 18 | — |
| | 2 | 22 | 23 | 20† (Fuzzy) |
| | 3 | 11† | 13 | — |
| | 4 | 30 | 34 | 25† (Fuzzy) |
| | 5 | 26† | 26 | — |
| | 6 | 13† | 16 | — |
| | 7 | 12 | 11† | — |
| | 8 | 26† | 34 | 26† (Fuzzy) |
| | 9 | 16† | 16 | — |
| | 10 | 17† | 21 | — |
| | Mean | 19† | 21 | — |
| Mean | | 20† | 24 | |

† Best fitting model.

stimulus phrase to describe the integrated meaning of the two forecasts. It is clear that the subjects fall into two distinct groups. Seven people (Subjects 6, 7, 9 from Group A and Subjects 2, 6, 8, 9 from Group B) operated according to the unanimity principle in over 90% of the cases. The remaining subjects did not follow this rule: seven of them *never* used the same word, and the other six used the same phrase only in very few cases (less than 30%). Most, (53%) of the selections of this group of subjects involved a modified version of the original forecast, and in only 44% of the cases did the subjects select totally "unrelated" responses.

We compared the membership functions of the stimulus phrases and the respective responses in terms of their location and fuzzyness as measured by appropriate indices. Yager (1981) defined the location of the (discrete) function, by analogy to the expected value of a random variable, as:

$$W = \sum_{i=1}^{n} \mu(p_i) \cdot p_i \Big/ \sum_{i=1}^{n} \mu(p_i).$$

By analogy to the variance of a random variable, the spread (or fuzzyness) of the

TABLE 6
*Analysis of responses to two indentical forecasts*

| Group | Subject | Percentage identical responses | "Selection" Percentage more extreme | Percentage less fuzzy | "And" Percentage more extreme | Percentage less fuzzy |
|---|---|---|---|---|---|---|
| A | 1 | 10 | 67 | 11 | 90 | 70 |
| | 2 | 30 | 86 | 71 | 80 | 10 |
| | 3 | 0 | 70 | 70 | 60 | 60 |
| | 4 | 0 | 100 | 100 | 40 | 30 |
| | 5 | 20 | 75 | 25 | 60 | 30 |
| | 6 | 100 | — | — | 90 | 90 |
| | 7 | 100 | — | — | 100 | 100 |
| | 8 | 0 | 90 | 80 | 40 | 40 |
| | 9 | 90 | 100 | 100 | 60 | 70 |
| | 10 | 10 | 67 | 33 | 0 | 0 |
| | Mean | 36 | 83 | 61 | 62 | 59 |
| B | 1 | 0 | 70 | 80 | 70 | 60 |
| | 2 | 100 | — | — | 60 | 60 |
| | 3 | 10 | 44 | 53 | 60 | 70 |
| | 4 | 20 | 50 | 75 | 60 | 70 |
| | 5 | 0 | 60 | 50 | 50 | 50 |
| | 6 | 100 | — | — | 50 | 50 |
| | 7 | 0 | 70 | 40 | 70 | 50 |
| | 8 | 100 | — | — | 50 | 60 |
| | 9 | 100 | — | — | 80 | 40 |
| | 10 | 0 | 70 | 70 | 70 | 60 |
| | Mean | 44 | 61 | 58 | 62 | 57 |
| Mean | | 40 | 73 | 60 | 62 | 58 |

function was defined by Fillenbaum, Wallsten, Cohen and Cox (1987), as:

$$V^2 = \sum_{i=1}^{n} \mu(p_i) \cdot (p_i - W)^2 \Big/ \sum_{i=1}^{n} \mu(p_i).$$

Both summations were performed across all nine probabilities. The second and third columns in Table 6 summarize the proportion of cases (out of 10) in which the phrases selected by the subject were more extreme (i.e. its $W$ was more distant from 0·5) and less fuzzy (i.e. its $V^2$ was smaller) than the original forecasts presented, given that the response was not identical to the original phrase.

Clearly, in the vast and significant majority of cases ($Z = 1{\cdot}72$, $p < 0{\cdot}05$) subjects selected more extreme phrases. Usually, these phrases were also less fuzzy, but this proportion is not significantly different from chance. The last two columns of the table indicate that in the "And" task responses are also more extreme and less fuzzy. This pattern holds for the 13 subjects, as well as for the group mean, but it is not statistically significant.

## 4. Discussion

Zwick *et al.* (1988) reported two intriguing results. First subjects' judgements of probability phrases in an "**and**" and an "**or**" context did not support any of the dual rules. Second, the "**and**" judgements were best fitted by some sort of an average of the two stimulus terms. The present experiment focused on the "And" task and confirmed these conclusions. Despite the differences in design, we found again overwhelming support for averaging rules and practically none for any rule that can be represented as a *t*-norm. It is particularly important to point out that by manipulating the nature of the instructions we eliminated the possibility that the previous results were due to the specific wording used in describing the task. The results for the two groups are indistinguishable, indicating that the observed superiority of averaging models reflects a real phenomenon.

Zwick *et al.* (1988) found that the fuzzy mean fitted the data better than the other simple pointwise means for about 50% of their subjects. This result was not replicated in the present study, where we found almost exclusive support for the arithmetic and geometric means in combining congruent and conflicting forecasts. It is not clear what caused this difference but the fuzzy mean deserves further considerations in the future. Further, there is no indication that either the arithmetic or geometric mean is the better one in this context. Their fits are almost equally good and they are highly related. The median absolute difference between the two sets of predictions is 0·045 for LL + LH pairs and 0·15 for LH pairs, so the observed differences are probably due to chance and other nonsystematic factors.

From our point of view the most important conclusion is that when integrating imprecise information from two independent sources subjects use some sort of compensatory averaging model. As we stressed in the introduction, this result is consistent with a large and heterogeneous body of psychological evidence accumulated in the context of integration of precise information (Anderson, 1981; 1982).

The notion of integration as an averaging process receives further support in our case because we found that these models can account in most cases for the results of the selection task as well. Admittedly, the support is weaker and is restricted to the case of two similar words (LL or HH). However, no model could describe well the process of selecting a single word for the LH pairs, simply because the subjects could not perform this task reliably.

One of the most interesting aspect of the present data concens the combination of two identical terms. There is a well documented debate in the Bayesian literature (e.g. Winkler, 1986) regarding the proper normative model and axioms that must be invoked in this situation. Our data clearly point to the underlying cause of this disagreement, namely that there is no single intuitive way to combine repetitive probabilistic information from two independent sources regarding the same event. About half of the participants in our experiment implicitly adopted the unanimity principle, while the other half behaved more in line with a "regular" Bayesian notion of updating probabilistic information. It must be noted that the behavior of this second group of subjects in this regard is inconsistent with their results in the other cases, since all pointwise averaging rules prescribe the unanimity principle. Obviously, they do not perceive the two cases to be identical and invoke different rules of behavior when faced with the two. These results confirm Winkler's conclusion (1986) that often the important question is not which particular rule is employed, but rather how one perceives, structures, and models the decision problem.

Recently there has been considerable interest in the use and manipulation of linguistically expressed probabilities in the context of various computerized expert systems (e.g. Bonissone & Decker, 1985; Degani & Bortolan, 1988; Godo, López de Mántaras, Sierra & Verdaguer, 1989). In these systems, the experts' probabilistic vocabulary is represented by triangular or trapezoidal fuzzy functions over the [0, 1] interval, and the terms are combined according to some version of the calculus of fuzzy sets. In principle, we are sympathetic to this approach. As we have argued elsewhere (Budescu & Wallsten, 1987; Zwick & Wallsten, 1989; Wallsten, 1990) in many cases numerical probabilities are not appropriate for modeling the vagueness and uncertainty experienced by the forecaster of the expert, and in such cases linguistic probabilities seem appropriate. We have also shown (Wallsten *et al.* 1986; Rapoport *et al.* 1987) that membership functions provide excellent representations of such expressions, and have demonstrated empirically that decisions based on these terms are in some cases as good as others based on standard numerical representations of uncertainty (Budescu, Weinberg & Wallsten, 1988; Wallsten, Budescu & Erev, 1988).

However, we object to the indiscriminate application of mathematically prescribed, but empirically unsupported, functions and operations to these forecasts. In many cases, the rules have little or no justification and lack any empirical support in the context of combination of *human* opinions and forecasts. Consequently their use might be inappropriate and lead to highly misleading inferences. This study and its predecessor (Zwick *et al.* 1988) clearly illustrate this point by demonstrating that human probabilistic judgements do not necessarily obey the mathematically compelling symmetry implied by De Morgan's theorem, and that simple averaging rules model the combination of vague opinions better than does any other operation. Thus, we urge that systems whose goal is to capture, replicate, and improve experts' diagnostic and inferential skills rely more heavily and make better use of results of empirical behavioral studies documenting these skills and their characteristics.

## References

ANDERSON, N. H. (1981). *Foundations of Information Integration Theory.* New York: Academic Press.

ANDERSON, N. H. (1982). *Methods of Information Integration Theory.* New York: Academic Press.

BELLMAN R. E. & ZADEH, L. A. (1970). Decision-making in fuzzy environment. *Management Science,* **17,** 141–164.

BELLMAN, R. E. & ZADEH, L. A. (1977). Local and fuzzy logics. In J. M. DUNN & E. EPSTEIN, Eds. *Modern Uses of Multiple Valued Logic,* pp. 103–165. Dordrecht: Reidel.

BEYTH-MAROM, R. (1982). How probable is probable? Numerical translation of verbal probability expressions. *Journal of Forecasting,* **1,** 267–269.

BONISSONE, P. P. & DECKER, K. S. (1985). *Selecting Uncertainty and Granularity: an Experiment in Trading Off Precision and Complexity.* KBS Working Paper, General Electric Corporate Research and Development Center, Schenectady, New York.

BUDESCU, D. V. & WALLSTEN, T. S. (1985). Consistency in interpretation of probabilistic phrases. *Organizational Behavior and Human Decision Processes,* **36,** 391–405.

BUDESCU, D. V. & WALLSTEN, T. S. (1987). Subjective estimation of precise and vague uncertainties. In G. Wright & P. Ayton Eds. *Judgmental Forecasting,* pp. 63–82. New York: John Wiley & Sons.

BUDESCU, D. V., WEINBERG, S. & WALLSTEN, T. S. (1988). Decisions based on numerically and verbally expressed uncertainties. *Journal of Experimental Psychology: Human Perception and Performance*, **14**, 281–294.

CLEMEN, R. T. (1988). *Combining Forecasts: A Review and Annotated Bibliography*. Working Paper, University of Oregon, Eugene, OR 97403.

CZOGALA, E. & ZIMMERMANN, H.-J. (1984). The aggregation operations for decision making in probabilistic fuzzy environment. *Fuzzy Sets and Systems*, **13**, 223–239.

DEGANI, R. & BORTOLAN, G. (1988). The problem of linguistic approximation in clinical decision making. *International Journal of Approximate Reasoning*, **2**, 143–162.

DUBOIS, D. & PRADE, H. (1980). *Fuzzy Sets and Systems: Theory and Applications*. New York: Academic Press.

DUBOIS, D. & PRADE, H. (1982). A class of fuzzy measures based on triangular norms. *International Journal of General Systems*, **8**, 43–61.

DUBOIS, D. & PRADE, H. (1985). A review of fuzzy set aggregation connectives. *Information Sciences*, **36**, 85–121.

DYCKHOFF, H. & PEDRYCZ, W. (1984). Generalized means as model of compensative connectives. *Fuzzy Sets and Systems*, **14**, 143–154.

FILLENBAUM, S., WALLSTEN, T. S., COHEN, B. & COX, J. A. (1987). *Some Effects of Available Vocabulary and Communication Task on the Understanding and Use of Non-Numerical Probability Expressions*. (Report No. 177). Chapel Hill, NC: University of North Carolina, L. L. Thurstone Psychometric Laboratory.

FRANK, M. J. (1979). On the simultaneous associativity of $F(x, y)$ and $x+y-F(x, y)$. *Aequationes Mathematicae*, **19**, 194–226.

GODO, L., LÓPEZ DE MÁNTARAS, R., SIERRA, C. & VERDAGUER, A. (1989). MILORD: The architecture and the management of linguistically expressed uncertainty. *International Journal of Intelligent Systems*, **4**, 471–501.

HERSH, H. M. & CARAMAZZA, A. (1976). A fuzzy set approach to modifiers and vagueness in natural language. *Journal of Experimental Psychology: General*, **105**, 254–276.

HERSH, H. M., CARAMAZZA, A. & BROWNELL, H. H. (1979). Effects of context of fuzzy membership functions. In M. M. GUPTA, R. K. RAGDE & R. R. YAGER, Eds. *Advances in Fuzzy Set Theory and Applications*, pp. 389–408. Amsterdam: North-Holland.

KRANTZ, D. H., LUCE, R. D., SUPPES, P. & TVERSKY, A. (1971). *Foundations of Measurement*; *Vol 1*. New York: Academic Press.

KULKA, J. & NOVAK, V. (1984). Have fuzzy operations a psychological correspondence? *Studia Psychologica*, **26**, 131–140.

LINDLEY, D. V. (1986). Another look at an axiomatic approach to expert resolution. *Management Science*, **32**, 303–306.

LUCHANDJULA, M. K. (1982). Compensatory operators in fuzzy linear programming with multiple objective. *Fuzzy Sets and Systems*, **8**, 245–252.

MENGER, K. (1942). Statistical metrics. *Proceedings of the National Academy of Sciences U.S.A.*, **28**, 535–537.

MORRIS, P. A. (1983). An axiomatic approach to expert resolution. *Management Science*, **29**, 24–32.

NAKAO, M. A. & AXELROD, S. (1983). Numbers are better than words. Verbal specifications of frequency have no place in medicine. *The American Journal of Medicine*, **74**, 1061–1065.

NORWICH, A. M. & TURKSEN, I. B. (1982). The fundamental measurement of fuzziness. In R. R. Yager, Ed. *Fuzzy Set and Possibility Theory*, pp. 49–60. New York: Pergamon Press.

NORWICH, A. M. & TURKSEN, I. B. (1984). A model for the measurement of membership and the consequences of its empirical implementation. *Fuzzy Sets and Systems*, **12**, 1–25.

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

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

ODEN, G. C. (1979). Fuzzy propositional approach to psycholinguistic problems: an application of fuzzy set theory in cognitive science. In M. M. GUPTA, R. K. RAGDE & R. R. YAGER, Eds. *Advances in Fuzzy Set Theory and Applications*, pp. 409–419. Amsterdam: North-Holland.

RAPOPORT, A., WALLSTEN, T. S. & COX, J. A. (1987). Direct and indirect scaling of membership functions of probability phrases. *Mathematical Modelling*, **9**, 397–411.

SMITHSON, M. (1984). Multivariate analysis using "And" and "Or". *Mathematical Social Sciences*, **7**, 231–251.

SMITHSON, M. (1985). Fitting "And" and "Or" Models to Data. Paper presented at the first IFSA Congress, Palma de Mallorca.

SMITHSON, M. (1987). *Fuzzy Set Analysis for Behavioral and Social Sciences*. New York: Springer-Verlag.

THÖLE, U., ZIMMERMANN, H.-J. & ZYSNO, P. (1979). On the suitability of minimum and product operators for the intersection of fuzzy sets. *Fuzzy Sets and Systems*, **2**, 167–180.

WALLSTEN, T. S. (1990). The costs and benefits of vague information. In R. HOGARTH, Ed. *Insights in Decision Making: Theory and Applications; A Tribute to the Late Hillel Einhorn* pp. 28–43. Chicago: University of Chicago Press.

WALLSTEN, T. S., BUDESCU, D. V. & EREV, I. (1988). Understanding and using linguistic uncertainties, *Acta Psychologica*, **68**, 39–52.

WALLSTEN, T. S., BUDESCU, D. V., RAPOPORT, A., ZWICK, R. & FORSYTH, B. (1986). Measuring the vague meaning of probability terms. *Journal of Experimental Psychology: General*, **115**, 348–365.

WEBER, S. (1983). A general concept of fuzzy connectives, negations and implication based on *t*-norms and *t*-conorms. *Fuzzy Sets and Systems*, **11**, 115–134.

WINKLER, R. L. (1986). Expert resolution. *Management Science*, **32**, 298–303.

YAGER, R. R. (1980). On a general class of fuzzy connectives. *Fuzzy Sets and Systems*, **4**, 235–242.

YAGER, R. R. (1981). A procedure for ordering fuzzy subsets on the unit interval. *Information Sciences*, **24**, 143–161.

ZADEH, L. A. (1965). Fuzzy sets. *Information and Control*, **8**, 338–353.

ZADEH, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning, Part I, *Information Sciences*, **8**, 199–249; Part II, *Information Sciences*, **8**, 301–357; Part III, *Information Sciences*, **9**, 43–80.

ZADEH, L. A. (1976). A fuzzy algorithmic approach to the definition of complex or imprecise concepts. *International Journal of Man–Machine Studies*, **8**, 249–291.

ZIMMERMAN, H.-H. & Zysno, P. (1980). Latent connectives in human decision making. *Fuzzy Sets and Systems*, **4**, 37–51.

ZWICK, R. (1987). A note on random sets and Thurstonian scaling methods. *Fuzzy Sets and Systems*, **21**, 351–356.

ZWICK, R., BUDESCU, D. V. & WALLSTEN, T. S. (1988). An empirical study of the integration of linguistic probabilites. In T. ZÉTÉNYI, Ed., *Fuzzy Sets in Psychology*, pp. 91–125. New York: Elsevier Science Publishers.

ZWICK, R. & WALLSTEN, T. S. (1989). Combining stochastic uncertainty and linguistic inexactness: Theory and experimental evaluation of four fuzzy probability models. *International Journal of Man–Machine Studies*, **30**, 69–111.