A Pattern Recognition Account of Decision Making

Dominic W. Massaro
Program in Experimental Psychology
University of California
Santa Cruz, CA 95064

Send Correspondence to
Dominic W. Massaro
Program in Experimental Psychology
University of California
Santa Cruz, CA 95064

Submitted to *Memory & Cognition*. An oral version of this paper was originally presented at the Psychonomic Society, Atlanta, Georgia, 1985 and the XXIV International Congress of Psychology, Sydney, Australia, 1988. The research reported in this paper and the writing of the paper were supported, in part, by grants from the Public Health Service (PHS R01 NS 20314) and the graduate division of the University of California, Santa Cruz. The author would like to thank B. Mehra Hosseini for his involvement in the the experiments and, of course, Michael Cohen for eclectic assistance. Michael Cohen, Dan Friedman, Jim Greeno, Danny Kahneman, Ranald Macdonald, Amos Tversky, and George Wolford provided valuable comments on this work. The data from these experiments is available by anonymous FTP from fuzzy.ucsc.edu in file pub/linda.tar.Z. For those without internet FTP please contact the author.

Abstract

In the domain of pattern recognition, experiments have shown perceivers integrate multiple sources of information in an optimal manner. In contrast, other research has been interpreted to mean that decision making is nonoptimal. As an example, Tversky and Kahneman (1983) have shown that subjects commit a conjunction fallacy because they judge it more likely that a fictitious person named Linda is a bank teller and a feminist than just a bank teller. This judgment supposedly violates probability theory because the probability of two events can never be greater than the probability of either event alone. The present research tests the hypothesis that subjects interpret this judgment task as a pattern recognition task. If this hypothesis is correct, subjects' judgments should be accurately by the fuzzy logical model of perception (FLMP)—a successful model of pattern recognition. In the first experiemnt, the Linda task was extended to an expanded factorial design with five vocations and five avocations. The probability ratings were well-described by the FLMP, and poorly described by a simple probability model. In the next study, the task was changed to include two fictitious people, Linda and Joan, as response alternatives. The results from the categorization judgments were better described by the FLMP than by an averaging of the sources of information. The ratings were accurately described by both models. The results reveal important similarities in recognizing patterns and in decision making. Given that the FLMP is an optimal method for combining multiple sources of information, the probability judgments appear to be optimal in the same manner as pattern-recognition judgments.

A Pattern Recognition Account of Decision Making

Dominic W. Massaro

The present research tests a viable model of pattern recognition as a description of cognitive probability judgments. The model, called a fuzzy logical model of perception (FLMP), formalizes developments in fuzzy logic (Zadeh, 1965), pattern recognition (Selfridge, 1959), and choice theory (Luce, 1959) to provide a systematic account of perceptual judgments. The FLMP has been tested extensively in a wide variety of domains and provides an informative account of the fundamental processes involved in situations in which there are multiple sources of information (Massaro, 1987, 1990, 1992; Massaro & Friedman, 1990). The FLMP incorporates three stages shown in Figure 1: evaluation, integration and decision.

Insert Figure 1 about Here

Perceivers are able to evaluate multiple sources of information independently of one another, integrate these sources with respect to alternative prototypes in memory, and make a decision on the basis of the relative goodness of match among the viable alternatives. These same operations have been observed in speech perception, written letter and word recognition, object recognition, categorization, memory, and sentence interpretation (Massaro, 1992). Similar processes might occur in decision making. The present paper extends the scientific framework, methodological technique, and theoretical approach to judgments about the probability of events. The research tests the hypothesis that subjects often interpret judgment tasks as a pattern recognition. If this hypothesis is correct, subjects' judgments should be well-described by the FLMP. To begin, a probability judgment task is described.

Probability Judgments about Events

In a provocative series of experiments, Tversky and Kahneman (1983) evaluated whether judgments about the conjunction of events agree with the prescription given by probability theory. As an example, subjects are given a description of a (now familiar) hypothetical person named Linda.

> Linda is 31 years old, single, outspoken, and very bright. She had a double major in philosophy and music. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

Subjects are then asked to indicate how probable is the case that 1) Linda is a bank teller, 2) Linda is a feminist, and 3) Linda is a bank teller and a feminist. A majority of subjects claim that the third case is more likely than the first (Fantino, Kulik, & Stolarz-Fantino, submitted; Stolarz-Fantino & Fantino, 1990; Tversky & Kahneman, 1983; Wolford, Taylor, & Beck, 1990; Wyer, 1976). Kahneman and Tversky (1983) call this type of behavior a conjunction fallacy. For reasons developed in the present research, this result will be called a *conjunction effect*. The conjunction effect in the

Linda task can be interpreted as a violation of probability theory and set theory because the probability of two events can never be greater than the probability of either event alone. According to simple probability theory, the probability of judging that Linda is a bank teller and a feminist must necessarily be less than or equal to both the probability that she is a bank teller and the probability she is a feminist. Consider the set of all bank tellers and the set of all feminists. The sets might or might not overlap. However, there can never be an element in the overlap between the sets that is not in both of the separate sets.

Tversky and Kahneman (1983) have claimed that subjects do not use probabilities but instead a representativeness heuristic, which usually involves some type of similarity matching. In the Linda task, representativeness is an assessment of the similarity of the description of Linda and some avocation, vocation, or avocation-vocation conjunction. When viewed in terms of representativeness, a conjunction effect is expected because adding feminism to the vocation of bank teller makes a better match with the description of Linda. Given the conjunction effect, subjects appear to use representativeness and not probability theory in their probability judgments. According to this explanation, subjects do *not* follow the instructions of the experimenter in this task. Subjects are instructed to indicate the probability of an alternative given the description of Linda, but they putatively judge the alternative in terms of the degree to which Linda resembles the typical member of that class (bank teller) or (bank teller and feminist).

Decision Making as Pattern Recognition

The present research follows in the spirit of Tversky and Kahneman in the sense that subjects are not doing what they are instructed to do. In fact, previous research appears to show that instructions have a relatively small effect on performance. For example, in the Linda task. Tversky and Kahneman found identical results with the Linda problem when subjects were asked to make similarity judgments and when they were instructed to make probability judgments. Macdonald and Gilhooly (1990) found no change in the size of the conjunction effect when believability was substituted for probability in the instructions. The present hypothesis is that subjects are carrying out pattern recognition (Massaro, 1985) or pattern-based reasoning (Lopes & Oden, 1991). The present inquiry is more optimistic than Tversky and Kahneman who state, "... the judged probability (or representativeness) of a conjunction cannot be computed as a function (e.g., product, sum, minimum, weighted average) of the scale values of its constituents." (1983, p. 305). Although this conjunction effect has been interpreted primarily as a violation of probability theory, it can also be interpreted as consistent with the FLMP (Massaro, 1985, 1987, pp. 272-277, 1988). These two interpretations create somewhat of a paradox because the currency and processes of the FLMP are isomorphic to a normative probability account (Massaro & Friedman, 1990).

Central to the FLMP is that two sources of information can be more informative than just one. Consider a situation in speech perception that is analogous to the Linda task. Perceivers listen to an audiotape of a male speaking in a high-pitched voice (falsetto) voice. They are then asked to rate the probability of a) a male speaking or b) a male speaking with a falsetto voice. They rate the statement b as more probable than a. This outcome is not unexpected, but did the subject make a conjunction fallacy? If simple probability theory is taken as the normative model, the conjunction of two properties (male and falsetto voice) should not be judged as more probable than either one alone. In terms of the FLMP, however, it is assumed that subjects are answering the question of the probability of the message on the audiotape given a) a male or b) a male speaking with a falsetto voice. In this case, it makes good sense to judge the second alternative as more probable as the first. The difference between the FLMP description and the simple probability description centers around the assumption of what subjects are judging (see Bar-Hillel, 1991, Wolford, 1991). According to the FLMP, individuals are likely to judge P(L|F), P(L|B), and P(L|F and B), where L, F, and B correspond to Linda, feminist, and bank teller, respectively. According to the simple probability theory, individuals should be judging P(F|L), P(B|L), and P(B and F|L).

One test between these two possible conclusions is to provide a quantitative test between these two models. Of course, given the previous research, we can expect the simple probability model to give a poor description of the Linda task. However, quantitative tests of this model have not been provided nor has its prediction been contrasted with other models. If the FLMP gives a good description of the results, we have some evidence for pattern recognition behavior in the Linda task. It would be valuable to contrast this model with a quantitative model of the representativeness heuristic, but one is not available. Tests of other quantitative models are possible, however, and these serve as alternatives to the optimal model.

Experiment 1: Test of Simple Probability Model and FLMP

Subjects were given the description of Linda and were asked to judge "how likely it was that various statements about Linda applied." Five avocations and five vocations were tested in an expanded factorial design, shown in Figure 2. Each possible vocation was paired with each possible avocation giving 25 unique pairings. In addition, the 10 individual characteristics were presented alone. This design gives 5 x 5 + 5 + 5 = 35 experimental questions.

Insert Figure 2 about Here

The five vocations ordered from most likely to least likely descriptions of Linda were:

1. social worker for the county
2. public elementary school teacher
3. sales clerk at a book store
4. teller at Bank of America
5. IBM executive

The five avocations were:

1. active in a national feminist organization
2. plays violin in an amateur chamber group
3. avid science fiction reader
4. crew member on a local sailing team
5. active in the right-to-life movement

These statements were chosen to span each continuum from being a very probable description of Linda to being very improbable.

METHOD

Subjects. Thirty subjects were recruited from psychology classes and satisfied part of a course requirement for participating for about one hour in the experiment.

Procedure. Two subjects were tested at a time in separate rooms. Each subject was seated at a IBM PC computer. Instructions were presented on the video display monitor followed by the description of Linda. This description of Linda remained in view throughout the experiment. Subjects were instructed to rate the probability of each statement being true of Linda on a scale from 1 (completely improbable) to 9 (completely probable). Subjects had as long as they wanted to make each decision and could change their decision before the next trial if they wished. The subjects entered their response by hitting one of the keys 1-9 and then the return key. The next trial was presented after an intertrial interval of one sec. The 35 statements were randomized within a block of 35 trials. Two successive trial blocks were presented. The ratings for each subject were linearly translated into values between 0 and 1.

RESULTS

The mean ratings across the 30 subjects are plotted as points as a function of the vocation (left plot), avocation (right plot), and combined vocation-avocation conditions (middle plot) in Figure 3. As can be seen in the figure, the continuum of vocations was more influential than avocations. Even so, both variables had statistically significant effects in both the single and combined conditions (all p values $< .001$). The interaction between the vocation and avocation factors was not quite statistically significant, $F(16, 464) = 1.637$, $p = .056$.

Insert Figure 3 about Here

The three circled points in Figure 3 depict the conjunction effect. Bank teller (BT) and feminist (FE) received a higher average rating than just bank teller. We now formalize quantitative models to test against these results.

Simple Probability Model (SPM)

Given the expanded factorial design, all of the results can also be used to test a simple probability model (SPM). The rating that Linda has an avocation can be defined as simply R(avocation). Similarly, R(vocation) would be the rating given a particular vocation statement. According to this model, the rating given the statement that Linda has an avocation and a vocation is predicted to be equal to the multiplicative combination of these two ratings.

$$R(vocation-avocation)=R(vocation) \times R(avocation) \tag{1}$$

In the test of this model, a different scale value is assumed for each level of vocation and for each level of avocation. Given five unique avocations and five unique vocations, 10 free parameters are necessary to predict the 35 data points.

It can be justly argued that this model is unrealistic because no correlation is assumed between the avocation and vocation in Equation 1. The conjunction effect appears to occur when there is a negative correlation between the two attributes (a feminist is unlikely to be a bank teller). Assuming some correlation between each avocation-vocation pair, however, precludes any test of the simple probability model. Twenty-five free parameters would be needed for the correlation, as well as the 10 parameters required for the scale values for the 5 avocations and 5 vocations. Given the independence assumption in Equation 1 and the conjunction effect, we would expect this probability model to give a poor description of the results. Even so, it is important to test this model because it provides a formal alternative and because independence between the vocation and avocation is assumed in alternative models that are tested.

This SPM and the other models was fit to the results of each subject and to the average results of the 30 subjects, using the program STEPIT (Chandler, 1969). The model is represented to the program in terms of a set of prediction equations and a set of unknown parameters. By iteratively adjusting the parameters of the model, the program minimizes the squared deviations between the observed and predicted points. The outcome of the program STEPIT is a set of parameter values which, when put into the model, come closest to predicting the observed results. Thus, STEPIT maximizes the accuracy of the description of each model. The goodness-of-fit of the model is given by the root mean square deviation (RMSD)—the square root of the average squared deviation between the predicted and observed values.

The lines in Figure 3 give the average predictions of the SPM. The model gave an average RMSD of .1205 in the fits to the individual subjects and an RMSD of .0881 to the average results.

Fuzzy Logical Model of Perception (FLMP)

Testing the FLMP requires knowledge of the candidate set of alternatives that subjects use in the task. Although only Linda is mentioned in the task, it is assumed that subjects generate a contrasting alternative that is the antithesis of Linda. We call this hypothetical person not(Linda). Applying the FLMP to the rating judgments, both the avocation and the vocation are assumed to provide continuous and independent evidence for the Linda and not(Linda). We assume that each subject generates a prototype for Linda and one for not(Linda). Given that we are manipulating the avocation and vocation in the test statements, it is sufficient to assume that these two prototypes differ on these two dimensions. A reasonable prototype for Linda would be

Linda : Intellectual avocation & Altruistic vocation

The prototype for not(Linda) is assumed to be

(not)Linda : (not)Intellectual avocation & (not)Altruistic vocation

It should be noted that these descriptors only approximate the subject's actual representation of both Linda and (not)Linda. It is only necessary that that the two prototypes differ from one another on the two relevant dimensions.

As indicated in Figure 1, the FLMP assumes three operations. The evaluation process determines the degree to which each dimension or source of information in the test statement matches each prototype. For each source of information, a fuzzy truth value between 0 and 1 is assigned for each prototype. This fuzzy truth value represents the degree to which the source of information matches the corresponding entry in the prototype. In fuzzy logic with two alternative prototypes, 0 is no support, .5 is ambiguous support, 1 is complete support, and other values are intermediate degrees of support. If aL_i represents the degree to which the avocation A_i from test statement i supports the alternative "Linda", then the outcome of prototype matching for "Linda" would be:

Linda : aL_i

where the subscript i indexes the five levels of the avocation dimension.

In fuzzy logic, the negation of a proposition is one minus its truth value. In this case, it can be assumed that the support from one source of information for not(Linda) is one minus its support for Linda. Thus, the outcome of prototype matching for "Not(Linda)" would be:

(Not)Linda : $(1 - aLi)$

The integration operation has no material consequences when there is just a single source of information. The decision operation determines the relative merit of Linda and not(Linda) alternatives, leading to the prediction that

$$R(Linda \mid A_i) = \frac{aL_i}{\sum} \qquad (2)$$

where R(Linda| A_i) is the rating of the statement being true of Linda, and $\sum$ is equal to the sum of the merit of the Linda and Not(Linda) alternatives. Given that the denominator of Equation 2 is one, the predicted rating is simply aL_i.

$$R(Linda \mid A_i) = \frac{aL_i}{aL_i + (1 - aL_i)} = aL_i \qquad (3)$$

Using a similar logic, the predicted rating for the single vocation conditions can be shown to be vL_j, where the subscript j indexes the vocation dimension.

The same evaluation operation occurs for the test statements containing both sources of information (an avocation and vocation). An important assumption is that the two sources are evaluated independently of one another. Given a prototype's independent specifications for the two sources, the value of one source cannot change the value of the other source. The integration of the two sources is determined by the product of their feature values from the evaluation stage. The outcome of the integration of the two sources of support for "Linda" would be:

Linda : $aL_i \; vL_j$

where the subscripts i and j index the levels of the avocations and vocations, respectively. Similarly, the outcome of prototype matching for "Not(Linda)" would be:

(Not)Linda : $(1 - aL_i)\,(1 - vL_j)$

The decision operation would determine their relative merit leading to the prediction that

$$R(Linda \mid A_i V_j) = \frac{aL_i \; vL_j}{\sum} \tag{4}$$

where $\sum$ is equal to the sum of the merit of the Linda and Not(Linda) alternatives. Ten free parameters, corresponding to the 5 avocations and 5 vocations, are necessary to predict the results.

These predictions of the FLMP were tested against both the results from individual subjects and the average results. The lines in Figure 4 give the average predictions. The RMSD was .0954 for the average fit of the individual subjects and .0465 for the fit of the average results. The latter RMSD is about twice as good as the fit of the SPM. An analysis of variance of the individual RMSDs revealed that the FLMP gave a better fit than the SMP, $F(1,29)=25.596$, $p<.001$.

Insert Figure 4 about Here

Figure 5 gives the parameter values from the FLMP. The scale on the right gives the scale values corresponding to area of the circle. The parameter values for the vocation changed in an orderly manner across the five levels. For the avocation, however, only feminist (FE) gave strong support for Linda and the other four levels were roughly neutral. The avocations were modified in the next experiment to give a better spread across this continuum.

Insert Figure 5 about Here

Although the FLMP gave a reasonable description of the results, it should be clear to the reader that the present interpretation of the conjunction effect rests on an important assumption about the prototype alternatives the subject generates in the task. We assumed that the subject generates a prototype for not(Linda), and the evidence supports this assumption. A better state of affairs would be to give the subject explicit alternatives so that choice alternatives do not have to be generated by the subject or assumed by the theorist. In this case, a direct test of the FLMP would be possible with the assumption that the prototype alternatives being used are equivalent to those given as response alternatives in the task.

Experiment 2: Modification of the Linda Task

A simple extension of the Linda problem makes the choice alternatives explicit. In this task, subjects are given descriptions of two hypothetical people (Linda and Joan) and have to rate or categorize which of the two people is more likely to have a vocation, an avocation, or both a vocation and an avocation. According to the present framework, this task better represents a pattern-recognition scenario in which the choice alternatives are explicitly given or known in advance. For example, deciding between Linda versus Joan is exactly analogous to choosing between /ba/ and /da/ in a speech recognition task given auditory and visual sources of information (Massaro & Cohen, 1990). To the extent that subjects behave similarly in the task with two explicit choice alternatives and the task with just one, the conjunction effect in the standard task can be described as the engagement of pattern recognition processes.

METHOD

In addition to the Linda description, subjects were given a description of Joan.

> Joan is 29 years old, married, athletic, and intelligent. She majored in economics, and graduated with honors. She was a writer for the conservative campus newspaper, and participated in intramural sports.

As in the first Linda experiment, five avocations and five vocations were tested in an expanded factorial design. The levels were the same as in the first experiment, except for a few modifications to improve the spread across each continuum. The five vocations ordered from most like Linda to most like Joan were:

1. social worker for the county
2. public elementary school teacher
3. sales clerk at a book store
4. loan manager at Bank of America
5. IBM executive in charge of computer programs

The five avocations were:

1. active in a national feminist organization
2. plays violin in an amateur chamber group
3. avid science fiction reader
4. competitive frisby player
5. active in the Republican party

Subjects. Two groups of eleven subjects each were recruited from psychology classes and satisfied part of a course requirement for participating for about one hour in the experiment.

Procedure. Two subjects were tested at a time in separate rooms. Each subject was seated at a IBM PC computer. Instructions were presented on the video display monitor followed by the descriptions of Linda and Joan. These descriptions of Linda and Joan remained in view throughout the experiment. Subjects were instructed to indicate whether Linda or Joan was more likely to have some avocation, vocation, or avocation and vocation. On each trial, an avocation, a vocation, or an avocation-vocation pair was presented. In the rating group, subjects were instructed "to type a number between 1 and 9 indicating "whether Linda or Joan is more likely." The number 1 would correspond to "definitely Linda" and 9 as "definitely Joan". The number 5 would correspond to "Linda and Joan equally likely", and so on for the intermediate numbers. In the categorization group, subjects were instructed "to type a letter "L" or "J" corresponding to whether you think Linda or Joan is more likely." These subjects entered their decisions by hitting one of two keys labeled "L" and "J" corresponding to Linda and Joan. Subjects in both groups had as long as they wanted to make each decision and could change their decision before the next trial if they wished. The subjects entered their response by hitting the return key and the next trial was presented after an intertrial interval of one sec.

Given an expanded factorial design with two independent variables with five levels of each variable, there were 35 unique trials. The 35 conditions were sampled randomly with replacement in each block of 35 trials. A given test session had six blocks of 35 trials in each block, and subjects were tested in two sessions. The dependent variables are the average ratings and the proportion of "Linda" judgments for each subject at each of the 35 experimental conditions. Based on research in other domains, the 12 observations per condition should be sufficient to obtain reliable data for individual-subject analyses and model tests (Massaro & Cohen, 1993).

RESULTS

The average ratings of "Linda" as a function of avocation and vocation are shown as the points in Figure 6. As can be seen in the left and center plots, the rating of "Linda" significantly decreased across the vocation continuum, for both the unidimensional, $F(4,40) = 39.48$, $p < .001$, and bidimensional, $F(4,40) = 51.14$, $p < .001$, conditions. Similarly, the right and center plots show that the rating of "Linda" significantly decreased across the avocation continuum, both for the unidimensional, $F(4,40) = 19.99$, $p < .001$, and bidimensional, $F(4,40) = 20.39$, $p < .001$, conditions. There was also a significant avocation by vocation interaction, $F(16,160) = 4.15$, $p < .001$, in the bidimensional condition, because each stimulus dimension had a larger effect to the extent that the other was ambiguous.

Insert Figure 6 about Here

The mean observed proportion of "Linda" identifications averaged across subjects is shown as the points in Figure 6. As can be seen in the left and center plots, the proportion of "Linda" responses significantly decreased across the avocation continuum, both for the unidimensional, $F(4,40) = 34.70$, $p < .001$, and bidimensional, $F(4,40) = 21.31$, $p < .001$, conditions. Similarly, the right and center plots show that the proportion of "Linda" responses significantly decreased across the avocation continuum, both for the unidimensional, $F(4,40) = 27.33$, $p < .001$, and bidimensional, $F(4,40) = 17.54$, $p < .001$, conditions. There was also a significant avocation by vocation interaction, $F(16,160) = 3.53$, $p < .001$, in the bidimensional condition, because each stimulus dimension had a larger effect to the extent that the other was ambiguous.

Insert Figure 7 about Here

A necessary first question is whether the conjunction effect was replicated. The three circled points in Figures 6 and 7, respectively, show a conjunction effect. Subjects were more likely to respond that Linda was a social worker (SW) and a frisby player (FR) than simply a social worker. Of course, subjects were more likely to respond that Linda was only a feminist than a loan manager and feminist.

Simple Probability Model (SPM)

The fit of the SPM was implemented in the same manner as in the Linda task (see Equation 1). The lines in Figures 6 and 7 give the average predictions of the SPM for the ratings and categorization tasks, respectively. The model gives a very poor description of the identifications of both rating and categorization judgments, with average RMSDs of .142 and .256 across the individual subject fits.

Fuzzy Logical Model of Perception (FLMP)

Applying the FLMP to the probability judgments, both the avocation and the vocation are assumed to provide continuous and independent evidence for the alternatives Linda and Joan. Defining the avocation and vocation as the important sources of information, the prototype for Linda would be the same as in the first Linda task.

Linda : Intellectual avocation & Altruistic vocation

The prototype for Joan would be defined in an analogous fashion,

Joan : Yuppie avocation & Ambitious vocation

At the evaluation stage, the avocation supports each alternative to some degree and analogously for the vocation. These degrees of support are integrated following the multiplicative rule given by the FLMP. With just two choice alternatives, the predictions do not change when it is also assumed that the degree of support for one alternative is the additive complement of the degree of support for the other (Massaro, 1989b). In this case, the prototype for Joan can be defined equivalently to the prototype for not(Linda) in the first experiment. Thus, the avocation support for Joan is given by $aJ_i = 1 - aL_i$ and the value $vJi = 1 - vLi$ gives the vocation support for Joan. The predictions are given by Equations 2-4 and require 10 parameters.

The FLMP was fit to the individual results of each of the 11 subjects in both the rating and the categorization tasks. The lines in Figures 8 and 9 give the average predictions of FLMP. Figures 10 and 11 give the average best fitting parameters of the FLMP. The parameter values change in a systematic fashion across the five levels of the avocation and vocation dimensions. The model provides a good description with an average individual RMSD of .048 for the rating judgments and .040 for the categorization judgments across the individual subject fits. Thus the fit of the FLMP is about three to six times better than the fit of the SPM. Analyses of variance were carried out on the RMSD values given by the fits of the SPM and FLMP. The FLMP gave a significantly better fit than the SPM for both the rating, $F(1, 10)=104.35$, $p<.001$, and categorization judgments, $F(1,10)=105.23$, $p<.001$.

Insert Figures 8-11 about Here

Weighted Averaging Model

In pattern recognition studies, the FLMP is usually compared to several other models. An important contender is a weighted averaging model (WAM). Fantino et al. (submitted) found some evidence for an averaging model in the Linda task. They claimed that subjects judge conjunctions by averaging the likelihood of their component parts. The present results and tests allow a stronger quantitative test of the averaging hypothesis. In addition, it allows for a weighted averaging, rather than a simple averaging, of the two components (Anderson, 1981). One WAM can be expressed as

$$R(Linda \mid A_i V_j) = (p)\, aL_i + (1-p) vL_j \qquad (5)$$

The WAM predicts that the rating of "Linda" given two sources of information is a simple weighted average of the rating of "Linda" given each of the separate sources. In this case, the weight corresponds to relative influence of the avocation and vocation sources of information.

A more appealing averaging model would be to assumed that the weight given a particular avocation (vocation) would vary with its actual scale value. A more informative avocation or vocation should receive more weight. Mathematically, this model can be expressed as

$$R(Linda \mid A_i V_j) = (p_i)\, aL_i + (p_j)\, vL_j \qquad (6)$$

In this case, each unique scale value carries a unique weight. The predictions of this model, however, cannot improve on the predictions given by the WAM in Equation 5. The reason is that the weights and scale values change together for each unique avocation or vocation. Thus, the scale value and the weight can be reduced to a single value without changing the predictions. Thus, Equation 6 reduces to

$$R(Linda \mid A_i V_j) = aL_i + vL_j \qquad (7)$$

Equation 5 is a more general version of Equation 7 because it allows a weight that is independent of the scale value.

Analogous to the other models, Equation 5 is also used to predict the probability of a Linda response in the categorization judgments.

To fit the WAM (given by Equation 5) to the results, each unique level of the avocation requires a unique parameter aL_i, and analogously for vL_j. The modeling of "Linda" responses thus requires 5 avocation parameters plus 5 vocation parameters.

The additional p value would be fixed across all conditions for a total of 11 parameters. Thus, the WAM requires one more parameter than the 10 required by the FLMP.

Figures 12 and 13 give the average observed results and the average predicted results of the WAM. As can be seen in the figures, the WAM gave a reasonable description of the rating judgments but a poor description of the categorization results. The average RMSD from the fit of the individual subjects' ratings was .061, only slightly poorer than the fit given by the FLMP, F(1,10)=1.302, ns. The average RMSD for the categorization judgments was .172, about four times poorer than the fit of the FLMP. An analysis of variance of the individual RMSDs revealed that the FLMP gave a better fit than the WAM, F(1,10)=43.66, p<.001.

Insert Figures 12 and 13 about Here

Yates and Carlson (1986) have proposed a modification of the WAM, called a "signed summation" model. In this model, individuals are assumed to categorize events as likely or unlikely, rather than evaluating them on some scale of certainty. This model can predict a conjunction effect, but can also supposedly predict that the conjunction of two likely events is more likely than either of the events alone. The authors claim that this model is identical to Anderson's weighted averaging model with an initial impression. To test this model against the present results, 14 free parameters were estimated: the initial impression, the 10 scale values, and 3 weights for the initial impression, avocation, and vocation, respectively. This model gave only a slightly better fit (RMSD = .061 and .170 for the rating and categorizations) than the fit of the WAM. It appears that the model can only predict the results only if a different initial impression is assumed for each unique avocation-vocation pair. This assumption would give the model as many free parameters as predicted data points, an unacceptable state of affairs because this is equivalent to assuming as much as is being predicted.

DISCUSSION

Previous Research

Fantino et al. and Stolarz-Fantino (1990) report several challenging results within the conjunction-fallacy paradigm. In a series of studies, they found a significant number of conjunction effects even when the personality description was eliminated. In some cases, the proportion of subjects showing a conjunction effect was not any larger with the description than without it. In other case, more subjects showed the conjunction effect with the description. The authors hypothesized that subjects simply average the likelihoods of the component parts of the conjunction. By independently manipulating the hypothetical person's hobby and job, they found evidence that favored the averaging explanation over that given by the representativeness heuristic. However, their results are based on averages across subjects rather than individual performance. Massaro and Cohen (1993) have shown that average results can inappropriately favor a linear model (averaging) relative to a nonlinear model (the FLMP). In a speech perception task, for example, quantitative tests of models against the results of individual subjects show that the multiplicative integration assumed by the FLMP gives a better description than averaging. Given the present results, we would expect that the rating results would be equally well-described by the multiplicative integration given by the FLMP.

Smith and Osherson (1989) offer a similarity explanation of the conjunction effect. The description of Linda leads to certain settings of her attributes. Subjects rate how likely Linda is a bank teller by comparing the similarity of a prototypical bank teller to this representation of Linda. When asked how likely Linda is a feminist bank teller, however, the attribute feminist is assumed to modify the component attributes of bank teller. This similarity explanation is a form of meaning change or nonindependence because the representation of bank teller differs in the two situations. A nonindependence explanation is difficult to test quantitatively because an inordinate number of free parameters is usually necessary to predict the results (Massaro, 1987). However, the nonindependence explanation implies that a model assuming independence would give a poor description of the results. Given the good description given by the FLMP (an independence model), we are able to reject this similarity explanation and nonindependence models more generally.

The present contribution is a normative account of the conjunction effect—that is, subjects are doing exactly what there should be doing when they claim that it is more probable that Linda is a bank teller and a feminist than Linda is a bank teller. Wolford et al (1990) arrived at a similar interpretation. They claimed that a Bayesian decision strategy makes sense in some contexts of the conjunction effect paradigm. Specifically, if the subjects interpret the possible outcomes as known, then a Bayesian analysis becomes appropriate. If the possible outcomes are not yet known, then probability theory remains appropriate. By varying the context to stress either known or unknown outcomes, the proportion of subjects showing the conjunction effect was modified in the predicted direction. However, over 50% of the subjects still gave a conjunction effect in the unknown context, leading the authors to conclude that subjects may still use the Bayesian model even when it is inappropriate. Wolford et al's distinction between known and unknown outcomes does not appear to account for when a conjunction effect will be found. Within the context of the present perspective, it is claimed that subjects typically follow the FLMP and a conjunction effect should usually be found.

The present analysis differs from other explanations that have stressed the ambiguity of the scenario established by the Linda problem. Margolis suggests two senses or meanings of the word probable: 1) probability in the gambling sense (and most appropriately equated with probability theory), and 2) probability synonymous with believable or plausible. The latter would seem to be best equated with making decisions about the world around us based on ambiguous information. To instantiate the gambling sense of probability, Margolis suggests modifying the Linda problem by adding the following warning along with a rephrasing of the question.

> A personnel survey showed that of clerical workers in banks (including tellers) fewer than 1% have personality profiles than sound similar to Linda's. If you stood to win 10 dollars if the statement you choose turns out to be true (whether or not the other statement is also true), which choice is more likely to win you the 10 dollars?

Phrased in this manner, this question should not produce a conjunction effect.

Margolis claims that the interpretation of probability in Tversky and Kahneman's experiments is in the second sense of believability or plausibility. The conjunction effect in Linda task is rational when this interpretation is made. The modification suggested by Margolis supposedly enforces the first interpretation in terms of gambling appropriate for probability theory. In this case, a conjunction fallacy should be avoided. In contrast to his claim, the perspective of the FLMP leads us to expect a conjunction effect even given his revised wording. Supporting the FLMP prediction, Macdonald and Gilhooly (1990) found little or no effect of formulating the Linda problem in terms of probability rather than belief.

Fiedler (1988) presented evidence that the conjunction effect can be modulated by the wording of the task. Subjects are usually asked to indicate how probable (or likely) a given outcome is. In a modification of the question, Fiedler's subjects were told that there was a sample of 100 people. They were asked to indicate how many persons out of this sample does the outcome apply. This modification reduced the conjunction effect considerably. Focusing the individual's attention on a large sample apparently attenuates use of the pattern recognition algorithm. Thus, it would be wrong to conclude that individuals will always follow the FLMP (or its equivalent Bayes' theorem). Birnbaum and Mellers (1983) presented strong evidence against the FLMP (equivalent to their subjective form of Bayes' theorem). Their task involved the use of base rate information and the opinion of an information source. Performance given only base rate information was, in some cases, inappropriately compromised by the opinion source. In this task, subjects appear to judge the average strength of evidence rather than accumulating several weak sources to imply a strong conclusion. This result might be specific to the use of base rate information. Specifying the conditions that engage the pattern recognition algorithm given by the FLMP should be an important issue in future research.

Fine-Grained Analyses

The present experiment improves on the methodology of previous studies of the Linda task. In previous studies, many subjects are asked just one or a few questions. For example, Fiedler (1988) analyzed 7 conjunction problems in terms of the overall proportion of correct answers as a function of the independent variables. No information was given about individual differences. In general, the results are necessarily presented in terms of the proportion of subjects that show a conjunction effect or the group's overall proportion.. However, there is necessarily some ambiguity what a given result of this kind means. When 80% of the subjects show a conjunction effect, we do not know that these subjects will always give the same response with repeated presentations of the same question under the same conditions. Similarly, the 20% that did not give a conjunction effect may have done so if the question were given again at another time.

It seems improbable to this writer that individuals can be easily partitioned into two groups—fallacy and non-fallacy. A fact of human performance is that choices are probabilistic rather than deterministic. In the prose of William James (1890/1950, p. 236), "A permanently existing "idea" or "Vorstellung" which makes its appearance before the footlights of consciousness at periodic intervals, is as mythological an entity as the Jack of Spades". From the simplest psychophysical task to judgments of category membership, individuals respond inconsistently given the same stimulus environment (Massaro, 1989a, chapters 10 and 18; McCloskey & Glucksberg, 1978). Thus, it is possible that there is less variation across individuals than possibly implied by previous studies of the Linda problem. Obtaining repeated measures and carrying out single-subject analyses permits one to address this question. The model tests will reveal whether all subjects behave qualitatively the same, or whether some subjects conform to one model and other subjects to another model.

Single-subject analyses are also necessary in tasks of this kind because not all subjects will order the properties in the same manner. For example, some subjects might decide that it is more probable that Linda plays the violin than Linda is an avid science fiction reader. Other subjects might make the opposite decision. Averaging these ratings would dilute the results and would make them inappropriate for tests of mathematical models of integration,. When contrasting linear and nonlinear models, averaging results across individuals also tends to favor an additive outcome even though the underlying process might be nonadditive (Massaro & Cohen, 1993).

Modularity and Input versus Central Systems

The present results show a striking parallel between pattern recognition given several sources of information and cognitive decision making. Fodor (1983) distinguished between input systems and central systems. The modular structure of input systems allows them to be understood and their functions described. Central systems, on the other hand, are nonmodular and cannot be understood or described. The good description by the FLMP in both domains weakens this distinction between input and central systems.

Averaging versus Multiplicative Integration

There is a fundamental difference in averaging integration and multiplicative integration within the FLMP (Massaro, 1987, chapter 7). An averaging integration of two sources of information produces an outcome that is less extreme than either source presented alone. In the FLMP, the operations of integration and decision can lead to a probability or rating judgment of a conjunction that is more true than the judgment of either or both of the single propositions. The current experiments were not completely successful in distinguishing an averaging from a multiplicative integration process.

These two forms of integration gave roughly similar descriptions of the rating judgments in both Experiments 1 and 2. However, averaging integration given by the WAM could not accurately describe the categorization judgments. The fit of the WAM was about four times poorer than the fit of the multiplicative integration in the FLMP. It should be noted that it will not be a simple matter to modify some aspect of a WAM to predict categorization judgments. One might modify the decision stage by allowing some type of nonlinear transformation of the output of integration. However, a scaling transformation on the output of integration cannot override a central feature of averaging integration. Although an averaging integration can explain the conjunction effect in the Linda task, it cannot explain another common finding that two sources of information can be more informative than either source. A loan manager at Bank of America and a competitive frisby player is less like Linda (or more like Joan) than either one of these characteristics taken alone.

Optimality of Decision Making

A perennial issue in the psychology of judgment and decision making is whether human decision and choice follow directly from an optimal prescriptive theory. That is, can our decision making be described by normative logic and mathematics? Almost four decades ago, Simon demonstrated (1955) that our behavior fell somewhat short of optimality. Research within probability learning (e.g., Estes, 1959) also appeared to indicate that human judgment was not perfectly rational. These findings and others widened the gulf between experimental psychology and other disciplines such as economics—which take optimality and rationality as first principles. Lopes (1987) and Larrick, Morgan, and Nisbett (1990) have shown that, with training, people can learn to behave more optimally. As in the previous studies, however, naive subjects in everyday life and in the laboratory appear to behave nonoptimally when compared to normative theory.

Given the mathematical equivalence between the FLMP and Bayes' theorem (Massaro, pp. 196-198, Massaro, 1989b), the good FLMP description of the Linda-Joan task demonstrates that decision making can be optimal. The appropriate normative theory for the two-alternative Linda-Joan task is Bayes' theorem—which assumes probability theory. That is, multiplying probabilities is also assumed within Bayes' theorem. When subjects are given a task requiring pattern-based reasoning paralleling pattern recognition, they behave appropriately. When subjects are given a task involving conjunctive events, however, they do not appear to follow the appropriate normative model. Our interpretation, like the original Tversky and Kahneman (1983) interpretation, is that subjects are not doing what they were instructed to do. Macdonald (1986), for example, views performance in the Linda task as uncertain because of ambiguities in the concept of probability, the sample space, and the intention of the experimenter's communication. However, the present research goes beyond the representativeness heuristic because the present interpretation claims that subjects are interpreting the task as pattern-based reasoning and performing this type of reasoning optimally. Subjects are carrying out pattern recognition in the standard Linda task even though they are instructed otherwise.

It should be noted previous rejections of Bayes' theorem have several limitations, see Massaro, 1987, pp. 197-198. Most importantly, one can distinguish between information and information-processing components of models. Traditionally, Bayes' theorem has been tested by asking whether both of these are optimal in human performance. The FLMP, on the other hand, allows for the information to be less than optimal, but requires that the processing of this information be optimal. Bayes' theorem and the FLMP make identical predictions in the Linda task because there is no objective measure of information. In other tasks, however, the models can make different predictions. If base rates (prior probability) is manipulated, for example, we can ask if subjects have a veridical representation of this source of information. The FLMP, on the other hand, allows for a representation of base rates that is not equivalent to the objective value. However, the integration of base rate is predicted to be integrated with other sources in an optimal manner (multiplicative integration as in Bayes' theorem).

Finally, one might also question the optimality of the relative goodness rule (RGR) for categorizations generated by the decision stage in the FLMP. According to this rule, subjects do not always choose the strongest candidate, but essentially probability match. This decision rule might be considered nonoptimal because it does not maximize the accuracy of performance. As pointed out by Massaro and Friedman (1990), however, a subject's goal might be to communicate the relative goodness of match (or posteriori probabilities in Bayesian terms), as they can easily do in a continuous rating task. There is also another possibility that allows subjects to be maximizing their performance. This possibility acknowledges the close relationship between the RGR and a deterministic criterion with the context of Thurstone's Law of Comparative Judgment and in signal detection theory (Massaro & Friedman, 1990; Yellott, 1977). The deterministic criterion can be considered an optimal decision rule, but performance is nonoptimal because of noise on the input. The deterministic criterion and the RGR make identical or very similar predictions. Our current results cannot distinguish between these possibilities. For our purposes, it is important to observe that information processing can be optimal even though the information transmitted by the subject is not veridical.

References

Bar-Hillel, M. (1991). Commentary on Wolford, Taylor, and Beck: The conjunction fallacy? *Memory & Cognition, 19,* 412-414.

Birnbaum, M. H., & Mellers, B. A. (1983). Bayesian inference: Combining base rates with opinions of sources who vary in credibility. *Journal of Personality and Social Psychology, 45,* 792-804.

Chandler, J. P. (1969). Subroutine STEPIT—finds local minima of a smooth function of several parameters. *Behavioral Science*, *14*, 81-82.

Estes, W. K. (1959). Component and pattern models with Markovian interpretation. In R.R. Bush & W.K. Estes (Eds.), *Studies in mathematical learning theory*, 9-52. Stanford, CA: Stanford University Press.

Fantino, E., Kulik, J., & Stolarz-Fantino, S. (submitted). The conjunction fallacy: A case of representativeness or averaging? *Psychological Science*.

Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research, 50*, 123-129.

Fodor, J. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT press.

James, W. (1950). *The principles of psychology*. New York: Dover. (Original work published 1890)

Larrick, R. P., Morgan, J. N., & Nisbett, R. E. (1990). Teaching the use of cost-benefit reasoning in everyday life. *Psychological Science, 1*,. 362-370.

Lopes, L. L. (1987). Procedural debiasing. *Acta Psychologica*, *64*, 167-185.

Lopes, L. L., & Oden, G. C. (1991). The rationality of intelligence. In E. Eells & T. Maruszewski (Eds.) *Probability and rationality: Studies on L. Jonathan Cohen's philosophy of science* (pp.199-223). Rodopi.

Luce, R. D. (1959). *Individual choice behavior*. New York: Wiley.

Macdonald, R. R. (1986). Credible conceptions and implausible probabilities. *British Journal of Mathematical and statistical psychology*, *39*, 15-27.

Macdonald, R. R., & Gilhooly, K. J. (1990). More about Linda *or* conjunctions in context. *European Journal of Cognitive Psychology, 2,* 57-70.

Margolis, H. (1987). *Patterns, thinking, and cognition: A theory of judgment.* Chicago: University of Chicago Press.

Massaro, D. W. (1985). *Conjoining multiple sources of information in pattern perception*. Paper read at the 26th Annual Meeting of the Psychonomic Society, Boston, MA, November.

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Massaro, D. W. (1988). *A pattern recognition account of decision making*. Paper read at the XXIV International Congress of Psychology, Sydney, Australia, August.

Massaro, D. W. (1989a). *Experimental psychology: An information processing approach.* San Diego, CA: Harcourt Brace Jovanovich.

Massaro, D. W. (1989b). Multiple Book Review of *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry*. *Behavioral and Brain Sciences, 12,* 741-794.

Massaro, D. W. (1990). A Fuzzy logical Model of Speech Perception. *Proceedings of the XXIV International Congress of Psychology*. In D. Vickers & P. L. Smith (Eds.) *Human Information Processing: Measures, Mechanisms, and Models* (pp. 367-379). Amsterdam: North Holland.

Massaro, D. W. (1992). Broadening the domain of the fuzzy logical model of perception. In H. L. Jr. Pick, P. Van den Broek, and D. C. Knill (Eds.), *Cognition: Conceptual and Methodological Issues* (pp. 51-84). Washington, D. C.: American Psychological Association.

Massaro, D. W., & Cohen, M. M. (1990). Perception of synthesized audible and visible speech. *Psychological Science, 1*, 55-63.

Massaro, D. W., & Cohen, M. M. (1993). The paradigm and the fuzzy logical model of perception are alive and well. *Journal of Experimental Psychology: General, 122*, 115-124.

Massaro, D. W., & Friedman, D. (1990). Models of integration given multiple sources of information. *Psychological Review, 97*, 225-252.

McCloskey, M., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets? *Memory & Cognition, 6,* 462-472.

Selfridge, O. G. (1959). Pandemonium: A paradigm for learning. In *Mechanization of thought processes* (pp. 511-526). London: Her Majesty’s Stationery Office.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69,* 174-183.

Smith, E. E., & Osherson, D. N. (1989). Similarity and decision making. In S. Vosniadou & A. Ortony (Eds.) *Similarity and analogical reasoning* (pp. 60-75). New York: Cambridge University Press.

Stolarz-Fantino, S., & Fantino, E. (1990). Cognition and behavior analysis: A review of Rachlin’s *Judgment, Decision, and choice. Journal of the Experimental Analysis of Behavior, 54,* 317-322.

Tversky,, A., & Kahneman, D. (1983). Extension versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*, 293-315.

Wolford, G. (1991). The conjunction fallacy? A reply to Bar-Hillel. *Memory & Cognition, 19,* 415-417.

Wolford, G., Taylor, H. A., & Beck, J. R. (1990). The conjunction fallacy? *Memory & Cognition, 18,* 47-53.

Wyer, R. S. (1976). An investigation of the relations among probability estimates. 1-18. *Organizational Behavior and Human Performance, 15,*

Yates, J. F., & Carlson, B. W. (1986). Conjunction errors: Evidence for multiple judgment procedures, including "signed summation". *Organizational Behavior and Human Decision Processes, 37,* 230-253.

Yellott, J. I. Jr. (1977). The relationship between Luce’s choice axiom, Thurstone’s theory of comparative judgment, and the double exponential distribution. *Journal of Mathematical Psychology, 15,* 109-144.

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

Figure Captions

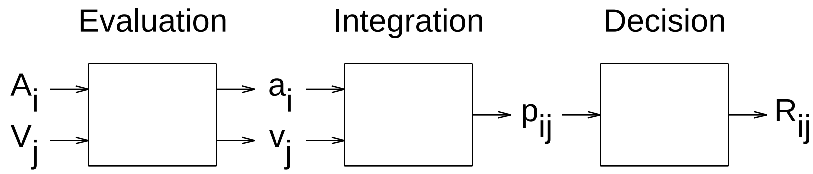

Figure 1. Schematic representation of the three operations involved in perceptual recognition and decision making. The evaluation of a source of information A_i produces a truth value a_i, indicating the degree of support for alternative R. An analogous evaluation occurs for another source V_j. Integration of the truth values gives an overall goodness of match p_{ij}. The response R_{ij} is equal to the value p_{ij} relative to the goodness of match of all response alternatives.

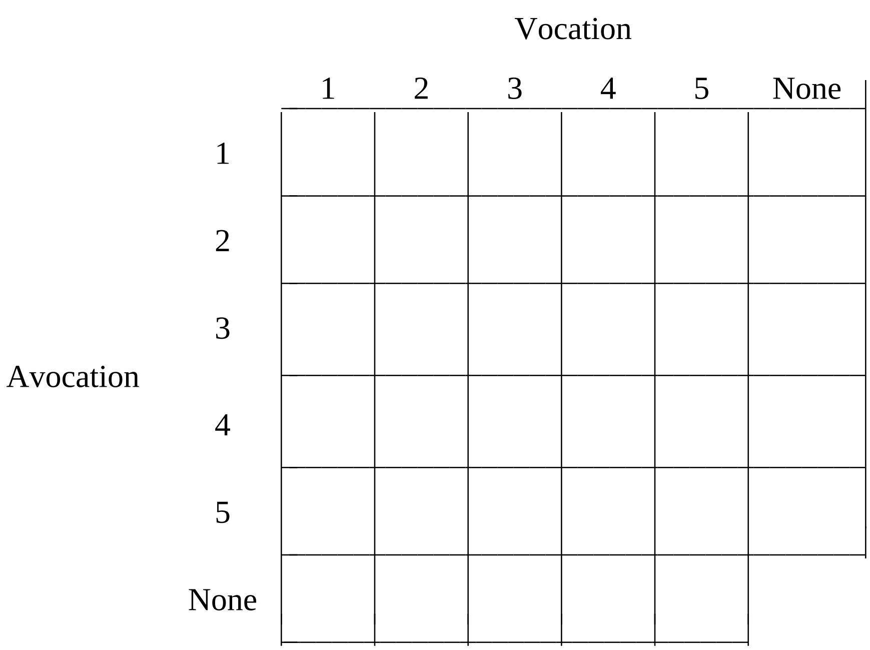

Figure 2. Expansion of a typical factorial design to include both the factorial combination of all conditions and the conditions presented alone. The five levels along the avocation and vocation continua represent statements varying in the degree to which they are true of Linda.

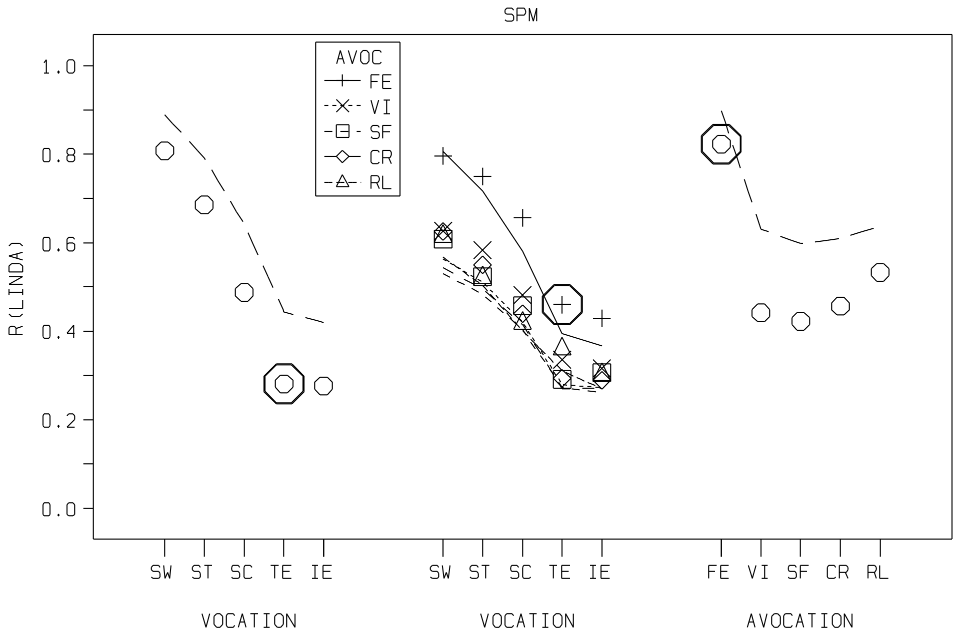

Figure 3. Observed (points) and predicted (lines) rating responses in Experiment 1 as a function of vocation (left graph), vocation and avocation (center graphs), and avocation (right graph). The three circled points FE, TE, and FE-TE replicate the conjunction effect. The predictions are for the simple probability model (SPM).

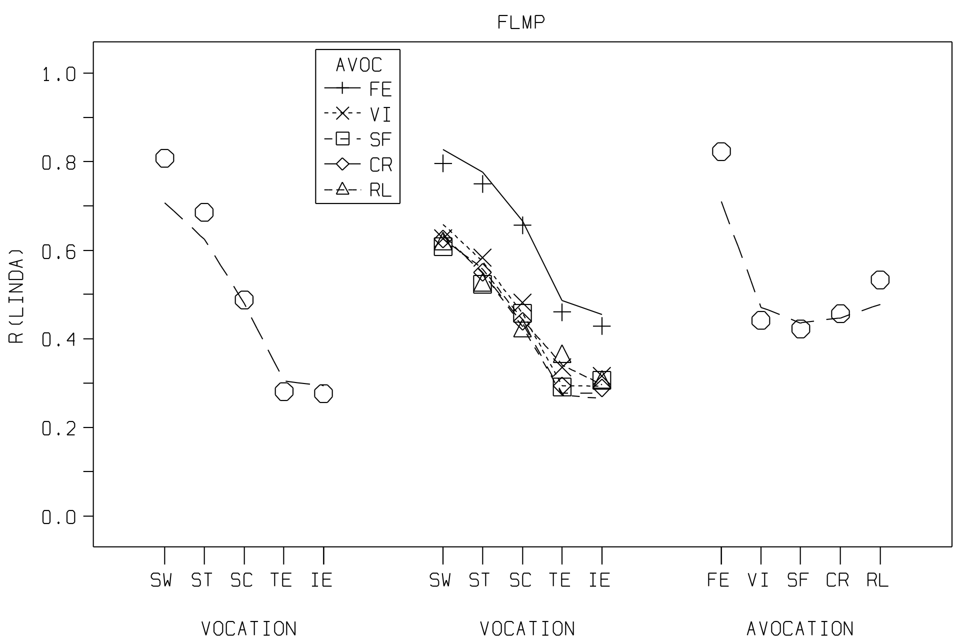

Figure 4. Observed (points) and predicted (lines) rating responses in Experiment 1 as a function of vocation (left graph), vocation and avocation (center graphs), and avocation (right graph). The predictions are for the fuzzy logical model (FLMP).

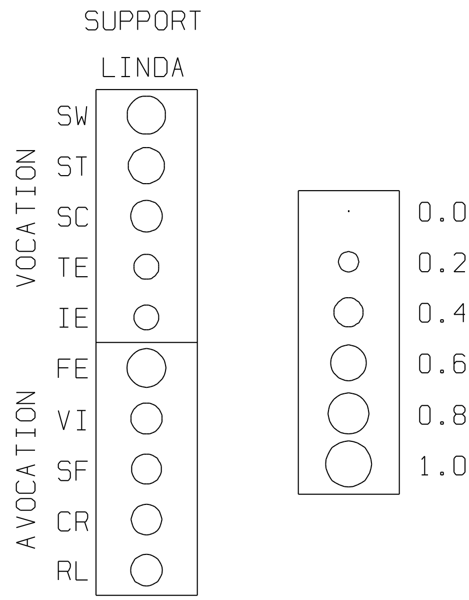

Figure 5. Plot of FLMP parameter values indicating the degree of support for Linda for the 5 vocations and 5 avocations. The scale on the right shows the relationship between area and scale value.

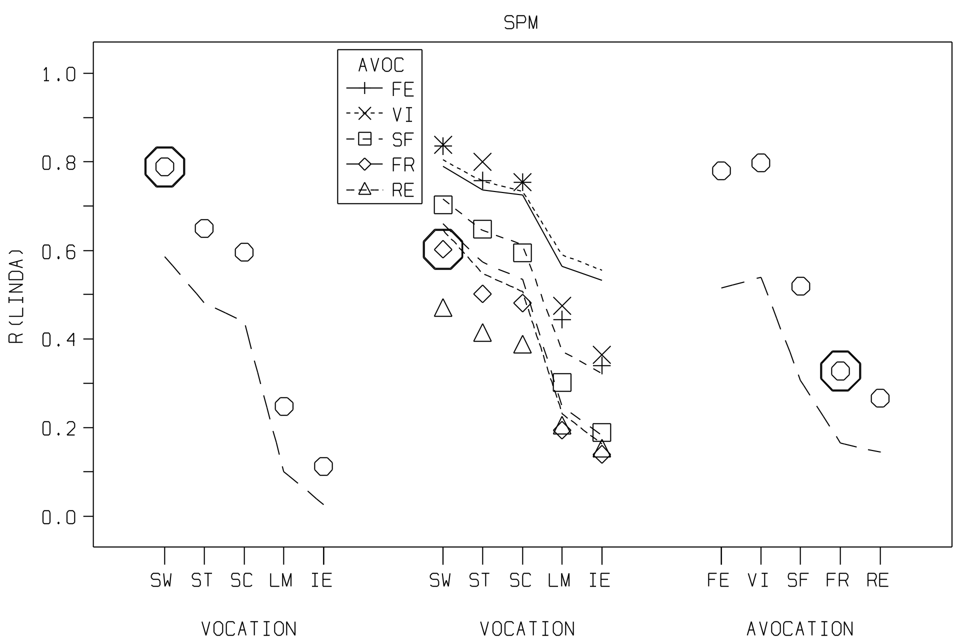

Figure 6. Observed (points) and predicted (lines) rating responses in Experiment 2 as a function of vocation (left graph), vocation and avocation (center graphs), and avocation (right graph). The three circled points SW, FR, and SW-FR replicate the conjunction effect. The predictions are for the simple probability model (SPM).

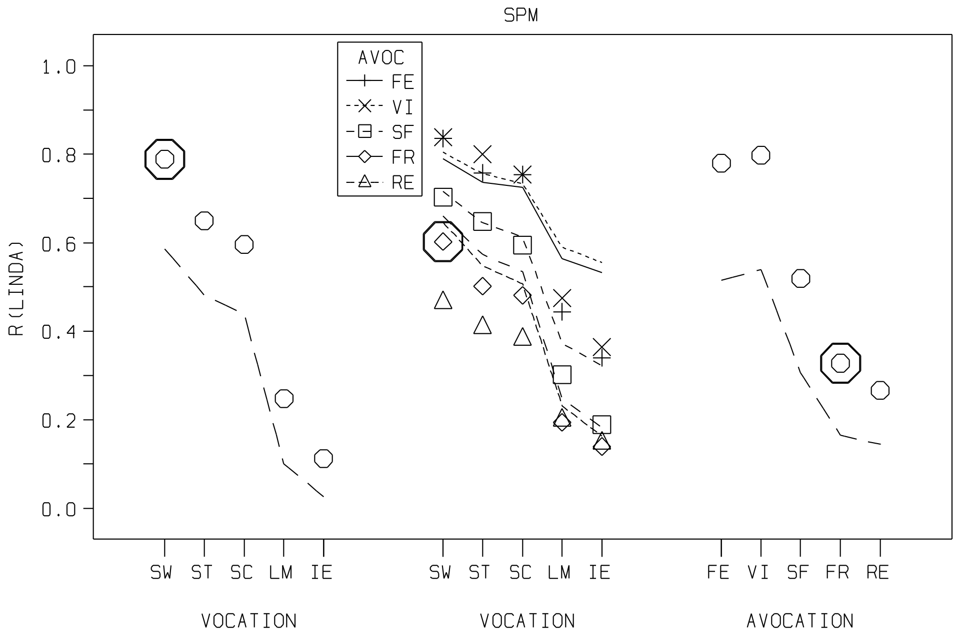

Figure 7. Observed (points) and predicted (lines) categorization responses in Experiment 2 as a function of vocation (left graph), vocation and avocation (center graphs), and avocation (right graph). The three circled points FE, BT, and FE-BT replicate the conjunction effect. The predictions are for the simple probability model (SPM).

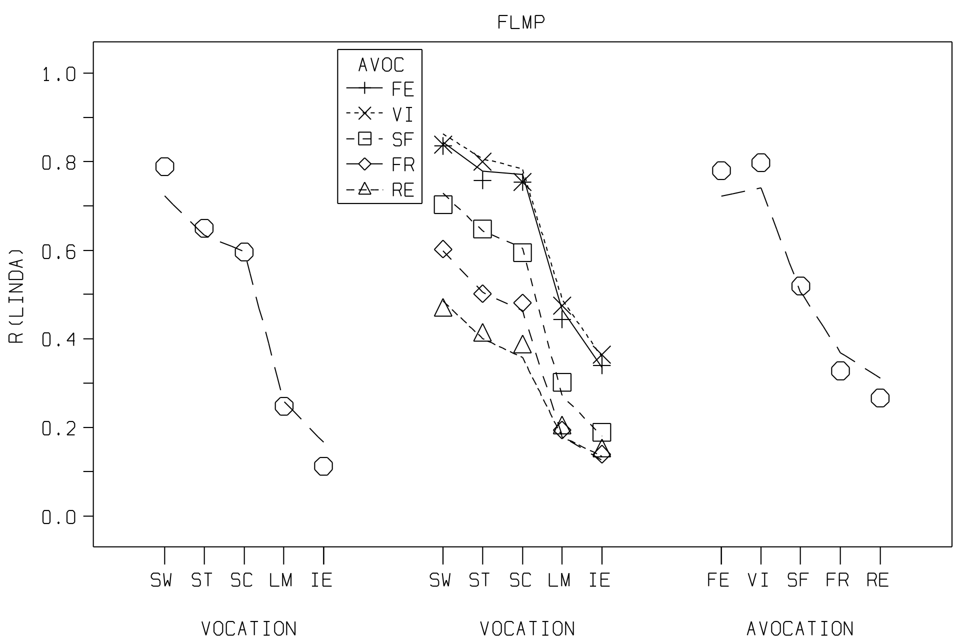

Figure 8. Observed (points) and predicted (lines) rating responses in Experiment 2 as a function of vocation (left graph), vocation and avocation (center graphs), and avocation (right graph). The predictions are for the fuzzy logical model (FLMP).

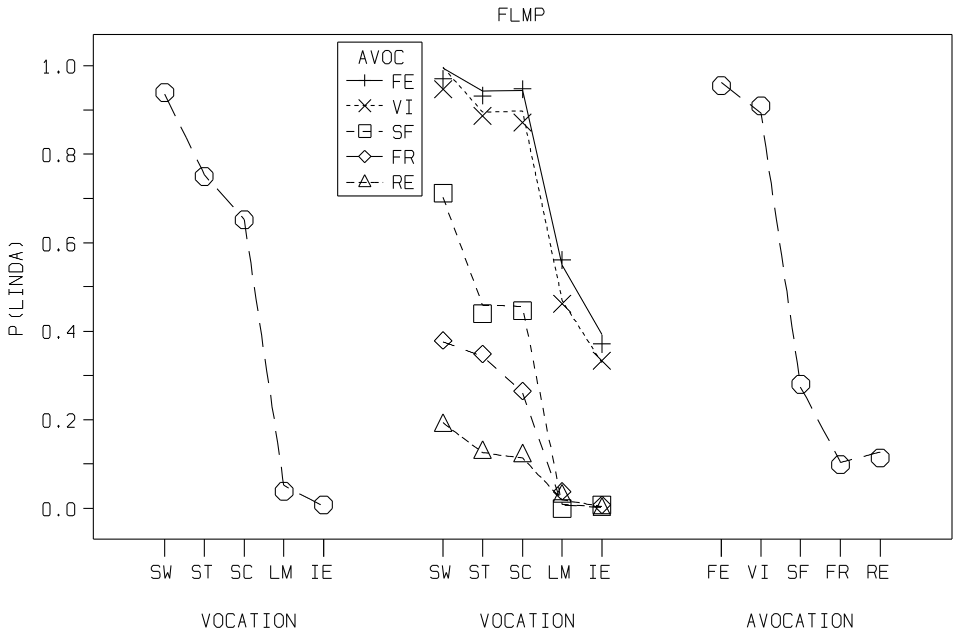

Figure 9. Observed (points) and predicted (lines) categorization responses in Experiment 2 as a function of vocation (left graph), vocation and avocation (center graphs), and avocation (right graph). The predictions are for the fuzzy logical model (FLMP).

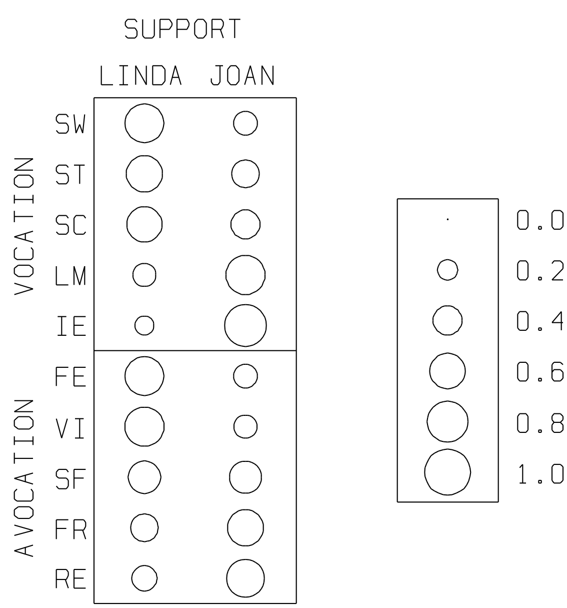

Figure 10. Parameter values of the FLMP indicating the degree of support for Linda for the 5 vocations and 5 avocations (Experiment 2, rating results).

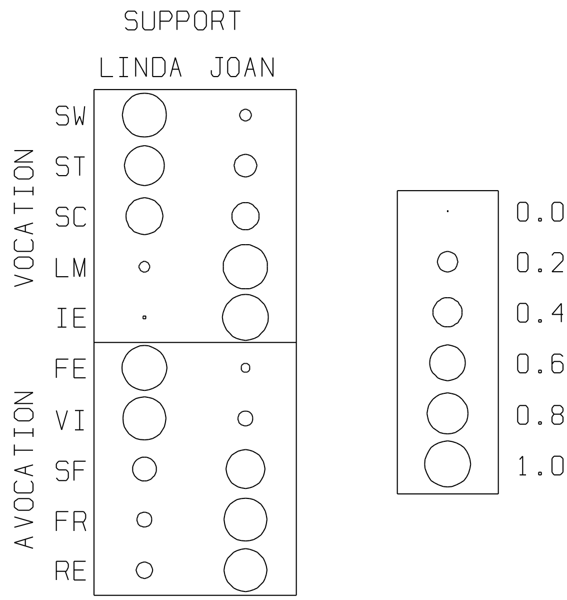

Figure 11. Parameter values of the FLMP indicating the degree of support for Linda for the 5 vocations and 5 avocations (Experiment 2, categorization results).

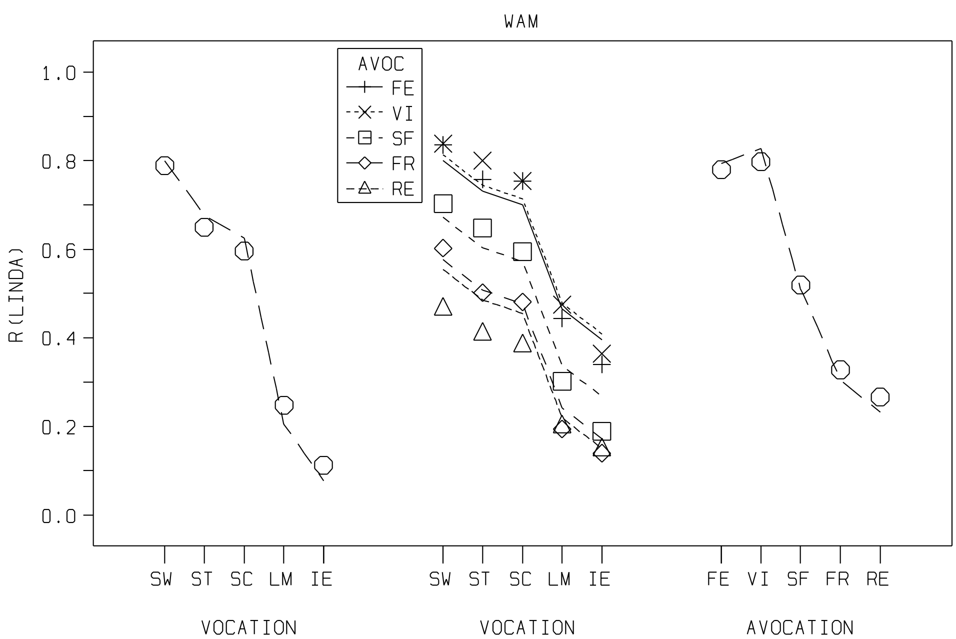

Figure 12. Observed (points) and predicted (lines) rating responses in Experiment 2 as a function of vocation (left graph), vocation and avocation (center graphs), and avocation (right graph). The predictions are for the weighted averaging model (WAM).

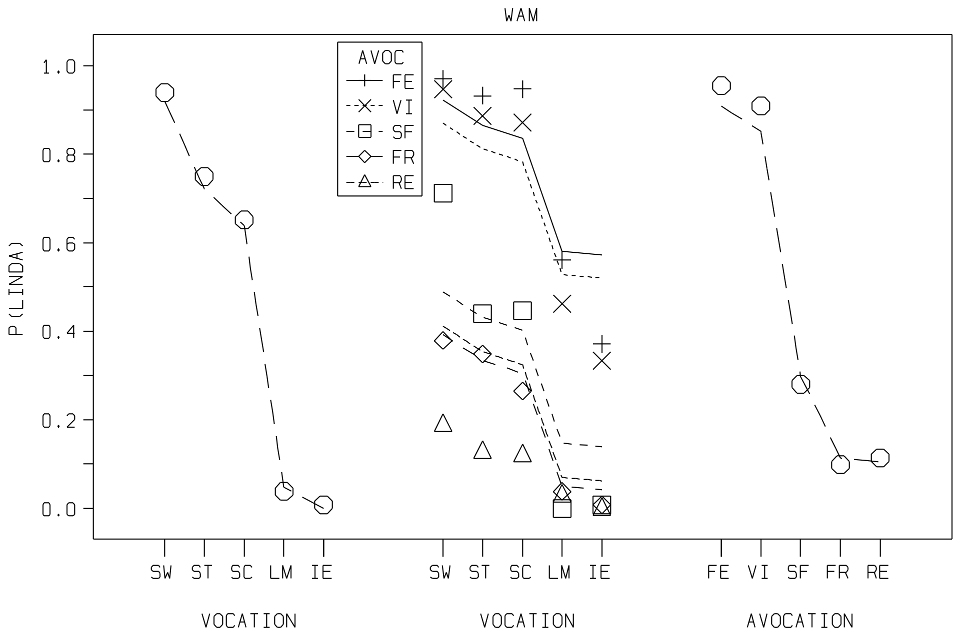

Figure 13. Observed (points) and predicted (lines) categorization responses in Experiment 2 as a function of vocation (left graph), vocation and avocation (center graphs), and avocation (right graph). The predictions are for the weighted averaging model (WAM).