COMMENTS

The Paradigm and the Fuzzy Logical Model of Perception Are Alive and Well

Dominic W. Massaro and Michael M. Cohen

Cutting, Bruno, Brady, and Moore (1992) criticized the paradigm for inquiry and the fuzzy logical model of perception (FLMP) presented in Massaro (1988a). In this reply to their remarks, it is shown that (a) the properties of the paradigm are ideal for inquiry; (b) models are best tested against the results of individual subjects and not average group data; (c) model fitting and analysis of variance do not give contradictory results; (d) the FLMP can be proven false and does not have a superpower to predict a plethora of functions or to absorb random variability; and (e) various extraneous characteristics of a model, such as equation length, cannot account for the success of the FLMP. On the other hand, the empirical findings of Cutting et al. give important new properties of pattern recognition. Finally, Cutting's theory of directed perception is compared with the FLMP.

Cutting, Bruno, Brady, and Moore (1992) questioned the value of the fuzzy logical model of perception (FLMP). Their claim that the good fit of the FLMP is not necessarily because of ability "in capturing psychological process" (p. 364). However, we believe that there are some fundamental errors in their analyses and arguments. Our reply shows that (a) the properties of our paradigm are ideal for inquiry; (b) models are best tested against the results of individual subjects and not group data; (c) model fitting and analysis of variance (ANOVA) do not give contradictory results; (d) the FLMP can be proven false and does not have a superpower to absorb random variability; and (e) various extraneous characteristics of a model, such as equation length, cannot account for the success of the model. We consider each of these points in separate sections of the article after we summarize our paradigm and the FLMP. We also discuss the importance of the new empirical findings of Cutting et al. and contrast Cutting's theory of directed perception with the FLMP. Readers familiar with our work can proceed to the section Individual Versus Group Analyses.

The Paradigm and the FLMP

Within the framework of the FLMP, perceptual events are processed in accordance with a general algorithm, regardless of the modality or particular nature of the patterns (Massaro, 1987; Oden, 1981, 1984). As shown in Figure 1, the model consists of three operations: feature evaluation, feature integration, and decision. The sensory systems trans-

duce the physical event and make available various sources of information called *features*. These continuously valued features are evaluated, integrated, and matched against prototype descriptions in memory, and an identification decision is made on the basis of the relative goodness of match of the stimulus information with the relevant prototype descriptions.

During feature evaluation, the features of the stimulus are evaluated in terms of prototypes that are generated for the task at hand. For each feature and for each prototype, feature evaluation provides information about the degree to which the feature in the signal matches the corresponding feature value of the prototype. During the second operation of the model, called *feature integration*, the features (actually the degrees of matches) corresponding to each prototype are combined (or *conjoined* in logical terms). The outcome of feature integration consists of the degree to which each prototype matches the stimulus. The third operation is *decision*. During this stage, the merit of each relevant prototype is evaluated relative to the sum of the merits of all relevant prototypes. This relative goodness of match gives the proportion of times the stimulus is identified as an instance of the prototype or a rating judgment indicating the degree to which the stimulus matches the category. A strong prediction of the FLMP is that the contribution of one source of information to performance increases with increases in the ambiguity of the other available sources of information.

Individual Versus Group Analyses

Replicating Massaro's (1988a) analyses, Cutting et al. (1992) found that model tests based on individual data did not discriminate between the FLMP and an additive model of perception (AMP). In contrast to the FLMP, the AMP predicts that the contribution of one source of information to performance is independent of the ambiguity of the other sources of information. However, considering the individual fits of all 44 subjects across three experiments, the FLMP
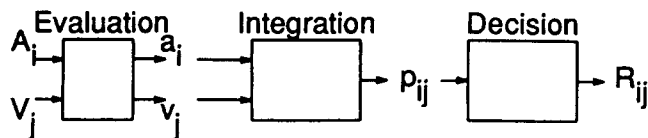
Figure 1. Schematic representation of the three operations involved in perceptual recognition. (The sources of information are represented by uppercase letters. Auditory information is represented by A_i, and visual information is represented by V_j. The evaluation process transforms these sources of information into psychological values [indicated by lowercase letters a_i and v_j]. These sources are then integrated to give an overall degree of support for a given alternative p_{ij}. The decision operation maps this value into some response, such as a discrete function or a rating, R_{ij}.)

did show a slight edge that approached statistical significance. When the models were fitted to the group data, determined by taking the average of the individual subjects, the AMP gave a better fit than the FLMP (see Cutting et al., p. 373, Table 4). Thus, there appears to be some discrepancy between the individual and group fits.

Cutting et al. (1992) also computed difference scores for each subject and carried out an ANOVA on these scores. The difference score is the difference in ratings given n and $n - 1$ sources of information supporting depth. If integration were additive, as assumed by the AMP, there should be no difference among the difference scores. The analysis indicated a significant difference and, therefore, this ANOVA provides evidence against the AMP.

Cutting et al. (1992) were concerned because the model fits to the average group data and the ANOVA on the difference scores did not appear to be in complete agreement on the question of the nature of the integration process. They concluded that ". . .with respect to the additive model and FLMP, the cluster of results is inconsistent" (p. 374). To help resolve this discrepancy, they studied the two models' ability to predict different functions and random variability.

We comment on this analysis in the section Superpower of FLMP. First, it is important to note, however, that there is no meaningful discrepancy between (a) the ANOVA on the difference scores and (b) model fits of individual subjects. Both of these analyses provide some evidence against additive integration. An apparent discrepancy is observed only when group averages are considered meaningful. We propose and support the hypothesis that average group results are not necessarily valid. If the meaningfulness of the average group results is rejected, there is no inconsistency in Cutting et al.'s results.

Cutting et al. (1992) were misled by group results. They noted a larger variability in the individual than in the group data and interpreted the group functions as more meaningful than the those from individuals. "First, when one model may have a moderate advantage over another (such as FLMP compared with the additive model), it may use that advantage in its fits to data of individuals but not of groups" (p. 380). It is well known, however, that a group function might not correspond to any of the individual functions making up the group. For this reason, previous investigators (Sidman, 1952) and recent textbooks (Massaro, 1989a) have cautioned against the use of group functions. We trust that Cutting et al. agree that our charge is to describe individual behavior and not an average score that does not necessarily exist in any real person. Although there are many examples of the sins of averaging in the literature, we illustrate how averaging the results across individuals can distort the results in favor of the AMP. Individual functions conforming to the FLMP predictions, when averaged together, might give a function that is more consistent with the AMP predictions.

To illustrate the dangers of averaging across subjects, we have done for the AMP and the FLMP what previous investigators have demonstrated for all-or-none and incremental models of learning. Individuals who learn in an all-or-none manner, when averaged together, can give results
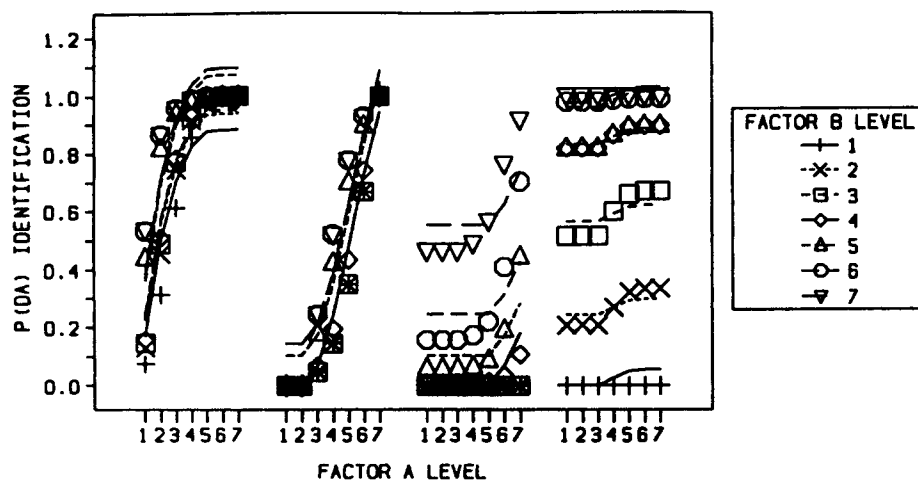

Figure 2. Simulated fuzzy logical model of perception data (points) and predicted additive model of perception fits (lines). (The four sections give the individual results for 4 pseudo-subjects as a function of factors A and B. $P[DA]$ = probability of identifying stimuli as /da/.)

predicted by the incremental model. In our analysis, we generated 4 ideal FLMP subjects performing in a factorial design with two independent variables, A and B, with seven levels per factor. The individual results are shown as the points in Figure 2. The individual results and the mean group results were fit by both the FLMP and AMP, using the parameter estimation program STEPIT (Chandler, 1969). Table 1 gives the root mean squared deviation (RMSD) values and the parameter values for the fit of the FLMP to the results. As expected, the FLMP gave almost a perfect description of the individual results, but the fit of the group results was significantly poorer, with an RMSD of 0.02254. The predictions of the FLMP are nonlinear, and averaging several FLMP subjects will not necessarily give an ideal FLMP subject. The fit of the AMP to these same results gave a very different outcome. Table 2 gives the RMSD values and the parameter values for the fit of the AMP to the results. In contrast to the fit of the FLMP, the AMP gave a poor description of the individual subjects (mean RMSD = 0.070) and a much better description of the group mean (RMSD = 0.00440). Figure 2 gives the predictions of the AMP for the 4 different FLMP subjects. As the figure shows, the AMP gives a poor description of the individual results. Figure 3 gives the fit of the FLMP and AMP models to the mean group data. The fit of the AMP to the group mean was 5 times better than the fit of the FLMP even though the AMP gave a very poor fit of each of the individual subjects. As Figure 3 shows, the group mean of FLMP subjects can resemble the predictions of the AMP. Thus, Cutting et al. (1992) were not justified in using the AMP's fit of the group data as evidence for additive integration. Although they concluded that "the modeling results of the group mean data favored the additive model" (p. 374), our demonstration reveals that their conclusion is not valid.

It is also important to note that averaging across different individuals who give ideal AMP results can never give results that are more accurately described by the FLMP. Averaging AMP results will always give ideal AMP results and will necessarily be more poorly fit by the FLMP than

the AMP. A good fit of the FLMP to group results is meaningful, whereas a good fit of the AMP is not. Thus, previous good fits of the FLMP to group results (e.g., Massaro, 1987, Chapters 2 and 9) are still meaningful.

Cutting et al. (1992) observed that "In general, group data are smooth, individual data are more noisy" (p. 380). However, this finding has little to do with individuals and groups and is simply a function of the number of observations per data point. The law of large numbers states that variance decreases with increases in the number of observations. One can also obtain smooth curves for individuals if the number of observations is increased. We illustrate this fact along with several other new observations in the next section.

Superpower of FLMP

A potentially devastating charge of Cutting et al. (1992) is that the FLMP cannot be proven false. Somehow the FLMP seems capable of predicting a plethora of functions and also has the magic power to absorb random noise. We criticize their demonstrations, logic, and interpretations in Fitting Functions and Fitting Random Data. In this section, we directly address the issue of falsifiability—what Massaro (1988b) called *superpower*. Cutting et al. stated "Because FLMP fits random noise better than the additive model, it will be at an advantage in the noisier, individual comparisons" (p. 380).

To provide an empirical test of whether the FLMP absorbs random variability and the AMP does not, we evaluated the goodness of fit of these models as a function of the number of observations per data point. Cutting et al. (1992) appeared to attribute the better fit of the AMP to the group results to the fact that the group results are less variable. However, group results are not necessarily less variable than the results of individual subjects. The total number of observations per data point is important in both cases. (In the previous section, we showed that the group results can be misleading and do not necessarily reflect the performance of any subject making up the group.) We expect the goodness

Table 1
RMSD Values and Parameter Values for the Fit of the FLMP to the Results of 4 Hypothetical FLMP Subjects

Subject	RMSD	Factor	Level						
			1	2	3	4	5	6	7
1	0.00004	A	.21561	.60488	.84155	.95199	.99525	.99918	.99990
		B	.22990	.35312	.38466	.40482	.74289	.80591	.80829
2	0.00009	A	.00010	.00010	.09889	.26973	.54000	.81336	.99912
		B	.31545	.31550	.31550	.39578	.66618	.74701	.74704
3	0.00002	A	.05385	.05386	.05387	.05950	.07938	.17492	.41827
		B	.00010	.00010	.00010	.14305	.52742	.76663	.93840
4	0.00007	A	.54170	.54191	.54211	.62394	.68055	.69266	.69296
		B	.00010	.18253	.47261	.79009	.79012	.97266	.99990
G	0.02254	A	.23448	.31559	.40479	.49999	.59523	.68490	.76488
		B	.23501	.31488	.40480	.49987	.59567	.68478	.76501
I	0.00006	A	.20281	.30019	.38411	.47629	.57380	.67003	.77756
		B	.13639	.21281	.29322	.43343	.68165	.82305	.87341

Note. G refers to the fit of the mean group results, whereas I corresponds to the mean of the individual subject fits. The two factors in the factorial design, A and B, have seven levels each. RMSD = root mean squared deviation; FLMP = fuzzy logical model of perception.

Table 2
RMSD Values and Parameter Values for the Fit of the AMP to the Results of 4 Hypothetical FLMP Subjects

Subject	RMSD	Factor	Level 1	2	3	4	5	6	7
1	0.08546	A	.02000	.34723	.55401	.67191	.72489	.72995	.73097
		B	.15152	.20792	.22022	.22780	.34271	.36591	.36683
2	0.08384	A	.00010	.00010	.06959	.24362	.47822	.72892	.94261
		B	.00010	.00010	.00010	.00010	.10544	.14386	.14387
3	0.08442	A	.00010	.00010	.00010	.00010	.00010	.07086	.18110
		B	.00010	.00010	.00010	.00010	.10243	.24657	.55418
4	0.02830	A	.00010	.00010	.00010	.02577	.04982	.05505	.05516
		B	.00010	.24332	.56425	.82915	.82917	.95571	.97340
G	0.00440	A	.00010	.08193	.16511	.24842	.33173	.41516	.49643
		B	.00337	.08495	.16826	.25136	.33512	.41819	.49975
I	0.07051	A	.00507	.08688	.15595	.23535	.31326	.39619	.47746
		B	.03796	.11286	.19617	.26429	.34494	.42801	.50957

Note. G refers to the fit of the mean group results, whereas I corresponds to the mean of the individual subject fits. The two factors in the factorial design, A and B, have seven levels each. RMSD = root mean squared deviation; AMP = additive model of perception; FLMP = fuzzy logical model of perception.

of fit of any reasonably accurate model to improve with increases in the number of observations per data point.

Given their conclusions about the smoothness of group relative to individual functions, Cutting et al. (1992) would seem to predict that the advantage of the FLMP over the AMP would decrease as the number of observations per data point is increased. On the other hand, we predict that the goodness of fit of the FLMP (and any reasonably accurate model) should improve with increases in the number of observations. To test between these predictions, we tested the models against the results of an experiment (Massaro, Tzusaki, Cohen, Gesi, & Heredia, in press) on bimodal speech perception. An expanded factorial design was used to manipulate auditory and visual information in a speech

perception task. The novel design illustrated in Figure 4 provides a unique method to address the issues of evaluation and integration of audible and visible information in speech perception. In this experiment, five levels of audible speech varying between /ba/ and /da/ were crossed with five levels of visible speech varying between the same alternatives. The presentation of the auditory synthetic speech was synchronized with the visible speech for the bimodal stimulus presentations. The audible and visible speech also were presented alone, giving a total of 25 + 5 + 5 = 35 independent stimulus conditions.

All of the test stimuli were recorded on videotape for presentation during the experiment. Six unique test blocks were recorded with the 35 unique test items presented in
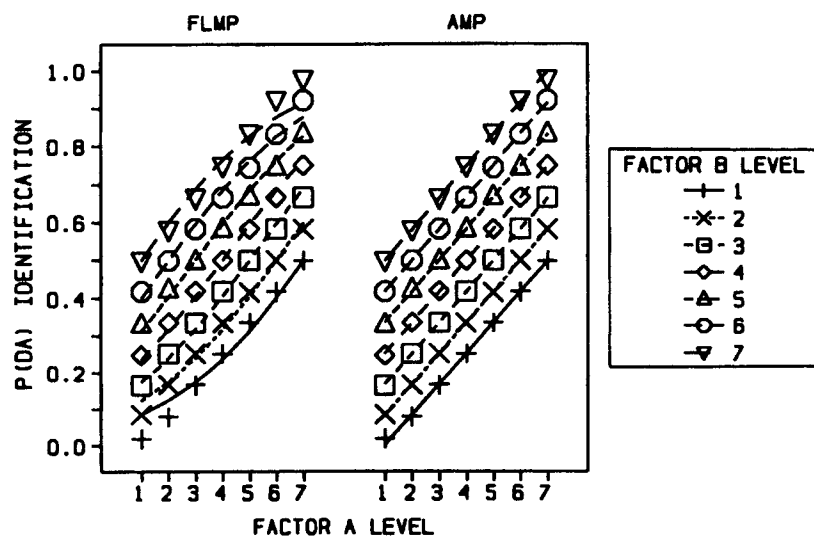

Figure 3. Simulated fuzzy logical model of perception (FLMP) average group data (points) and predicted FLMP (left panel) and additive model of perception (AMP; right panel) fits (lines) as a function of arbitrary factors A and B. (Average group data are from the 4 simulated FLMP pseudo-subjects shown in Figure 2. P[DA] = probability of identifying stimuli as /da/.)

Visual

Figure 4. Expansion of a typical factorial design to include auditory and visual conditions presented alone. (The five levels along the auditory and visible continua represent auditory and visible speech syllables varying in equal physical steps between /ba/ and /da/.)

each block. Twenty-one college students were tested in the experiment. Subjects were instructed to listen and to watch the speaker and to identify the syllable as either /ba/ or /da/ during a 3-s response interval. Each of the 35 possible stimuli were presented a total of 24 times during four sessions, with six observations per stimulus condition in each session.

Figure 5 shows the mean proportion of identifications across subjects. The identification judgments changed systematically with changes in the audible and visible sources of information. The likelihood of a /da/ identification increased as the auditory speech changed from /ba/ to /da/ and changed analogously for the visible speech. Each source had a similar effect in the bimodal conditions
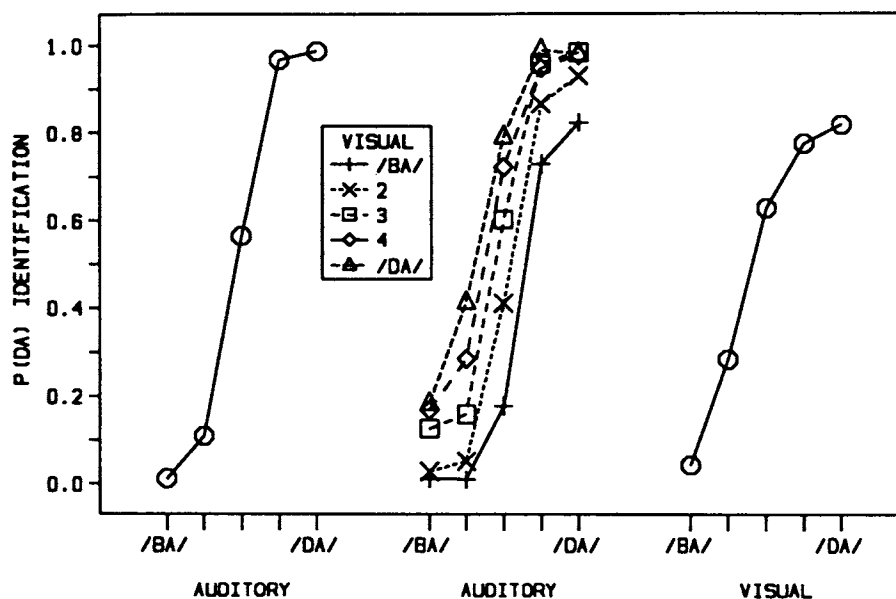

Figure 5. Proportion of /da/ identifications for the auditory-alone (left panel), the factorial auditory–visual (center panel), and the visual-alone (right panel) conditions as a function of the five levels of the synthetic auditory and visual speech varying between /ba/ and /da/. (Results from the 21 English speakers in Massaro, Tsuzaki, Cohen, Gesi, & Heredia, in press. P[DA] = probability of identifying stimuli as /da/.)

relative to the corresponding unimodal condition. In addition, the influence of one source of information was greatest when the other source was ambiguous.

As mentioned in the description of this study, the subjects were tested in four sessions of six blocks of 35 trials in each block. We were interested in the goodness of fit of the models as a function of the number of observations per data point. Therefore, we tested the models with 6, 12, and 24 observations per data point. For each subject, there were seven sets of data. Four sets had 6 observations per data point, two sets had 12 observations, and one set had 24 observations. These sets were created by simply pooling the observations across the appropriate number of trials. The FLMP and AMP were fitted to the results. An RMSD value was determined for each subject at each of the three conditions. An ANOVA was carried out on these RMSD values with number of observations and model as two independent variables. Figure 6 shows the obtained RMSD values for the FLMP and AMP fits as a function of the number of observations. The FLMP gave a significantly better fit than the AMP, $F(1, 20) = 1,126, p < .001$, and the goodness of fit of both models improved with increases in the number of observations, $F(2, 40) = 19, p < .001$. There was no interaction between these variables, $F(2, 40) = 0.037, p = .96$. Given that the goodness of fit improved with increases in the number of observations, Cutting et al. would have to predict that the advantage of the FLMP over the AMP should have been larger with fewer observations and a poorer overall fit. However, this was not the case, and this result casts doubt on Cutting et al.'s premise that the FLMP absorbs random variability.

One might question why the fit of the AMP improved with increases in the number of observations if the AMP is

Figure 6. The average root mean squared deviation (RMSD) values of the fit of the fuzzy logical model of perception (FLMP) and additive model perception (AMP) to results based on 24, 12, or 6 observations per condition to the 21 English speakers in Massaro, Tsuzaki, Cohen, & Gesi, Heredia (in press).

indeed the wrong model. The answer is that the AMP is not completely wrong. For example, it correctly predicts that the functions across the auditory and visual speech should be monotonic. Increased variability decreases the likelihood of observing this monotonicity. However, the AMP gives a consistently poorer fit than the FLMP because it fails to predict the quantitative interaction between the auditory and visual speech. The empirical results indicate that the contribution of one source is largest when the other source is ambiguous.

The FLMP is deterministic at the feature evaluation and integration stages and becomes stochastic at the decision stage (Massaro & Friedman, 1990). The variability at the decision stage is due to the relative goodness rule (RGR), in which the probability of a response is equal to the merit of that alternative relative to the sum of the merits of all relevant alternatives. For example, given an RGR value of .8, that alternative is chosen .8 of the time. A strong prediction of the model is that the observed variability should be equal to that expected on the basis of simple binomial variance. Thus, the FLMP not only predicts probability but also predicts that its expected variability should decrease with increases in the number of observations. It is possible to determine the expected binomial variability as a function of the number of observations. With this measure of variability we can ask if the fit of a model is poorer than what would be expected from binomial variability. One way to determine the expected variability is to generate simulated results of the models based on the parameters of the models determined in the tests of the models.

To determine the expected variability predicted by the FLMP as a function of the number of observations, the following simulations were carried out. Using the parameter values from the fit of the FLMP to the results of the Massaro et al. (in press) study (with 24 observations per data point), we generated the 35 predicted response probabilities for each of the 21 subjects. Using these predictions, we generated 50 simulated subjects from the predictions of each real subject. The FLMP predicts the probability of a /da/ or /ba/ response at each experimental condition based on the outcome of the RGR in the model. For example, the prediction for a /da/ response might be .876 for a given condition. The prediction for a /ba/ response would be the additive complement, or .124. For the simulated subject, a uniform random number between 0 and 1 was drawn. If the number was less than or equal to .876, then the simulated response would be a /da/. If the number was greater than .876 then the simulated response would be a /ba/. This computation was carried for each of the 35 data points with 6, 12, 24, 48, 96, and 192 simulated trials. Fifty simulated subjects were created for each of the 21 real subjects.

Figure 7 gives the expected average RMSD values of the fit of the FLMP and AMP to simulated results based on the predictions of the FLMP as a function of the number of observations per data point. An ANOVA was carried out on the RMSD values with model and number of observations as factors. All significant differences were for $p < .001$. As Figure 7 shows, the fit of both models improved with increases in the number of observations per data point. The
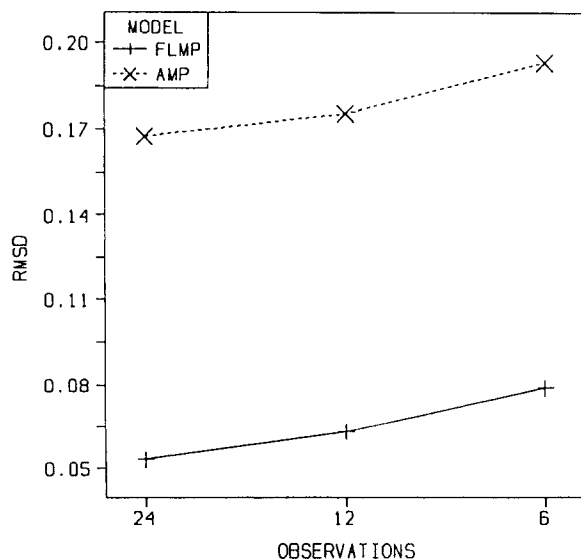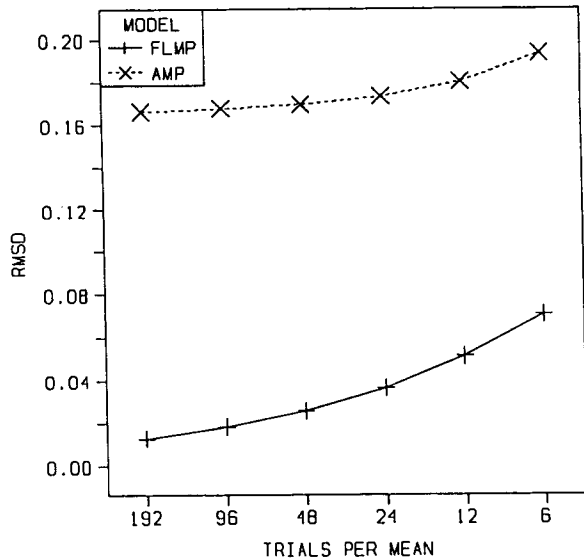
Figure 7. The expected average root mean squared deviation (RMSD) values of the fit of the fuzzy logical model of perception (FLMP) and additive model of perception (AMP) to simulated results based on the predictions of the FLMP as a function of the number of observations per data point. (The predictions were generated for 1,050 hypothetical subjects based on the parameter values obtained from the fit of the FLMP to the 21 English speakers in Massaro, Tsuzaki, Cohen, Gesi, & Heredia, in press. Fifty hypothetical subjects were generated for each of the 21 sets of parameters.)

FLMP gave a significantly better fit, and the advantage of the FLMP increased with increases in the number of observations per mean.

These results can be compared with the RMSD values from the observed results in Figure 6. The RMSDs from the observed data are of roughly the same order of magnitude as the RMSDs from the simulated data. In both real data and simulated data, the fit of the FLMP improved with increases in the number of observations per data point. The fit of the AMP also improved but leveled off at an RMSD of about 0.167.

We expected and found that observed variability (as measured by the RMSD values) would be somewhat greater than the predicted binomial variability. For the real data, the RMSD values for the FLMP were 0.0787, 0.0631, and 0.0530 for 6, 12, and 24 observations, respectively. For the simulated data, the RMSD values for the FLMP were 0.0704, 0.0510, and 0.0364 for 6, 12, and 24 observations, respectively. One reason the observed RMSDs might be slightly larger than the predicted RMSDs is that the parameter values might fluctuate somewhat over the course of the experiment. That is, the support of one source of information for a given alternative might be slightly different in one block of trials than in another block. The fit of the model, of course, assumes that the parameter values are constant across all blocks of trials. Some variability in the parameter values across trial blocks is consistent with the observation that the difference between the observed and simulated

RMSDs increased somewhat with increases in the number of observations. We might expect that the parameter values to be less stable across as the number of trial blocks increases.

The AMP also introduces variability at the decision stage. An analogous simulation was carried out for the AMP by simulating hypothetical AMP subjects based on the parameter values of the fit of the AMP. Figure 8 gives the expected mean RMSD values of the fit of the FLMP and AMP to simulated results based on the predictions of the AMP as a function of the number of observations per data point. Analogous to the FLMP analyses in Figure 7, the predictions for the AMP were generated for 1,050 hypothetical subjects based on the parameter values obtained from the fit of the AMP to the 21 subjects in the Massaro et al. (in press) study. Fifty hypothetical subjects were generated for each of the 21 sets of parameters. As Figure 8 shows, a very similar pattern was obtained. The fit of both models improved with increases in the number of observations per data point. The AMP gave a significantly better fit, and the advantage of the AMP increased with increases in the number of observations per mean. This analysis reveals that the FLMP and AMP are identifiably different (Massaro & Friedman, 1990) and that the models are best distinguished when tested against individual subjects with a large number of observations per data point.

Massaro and Friedman (1990) already demonstrated that the FLMP and AMP were distinguishable from one another.
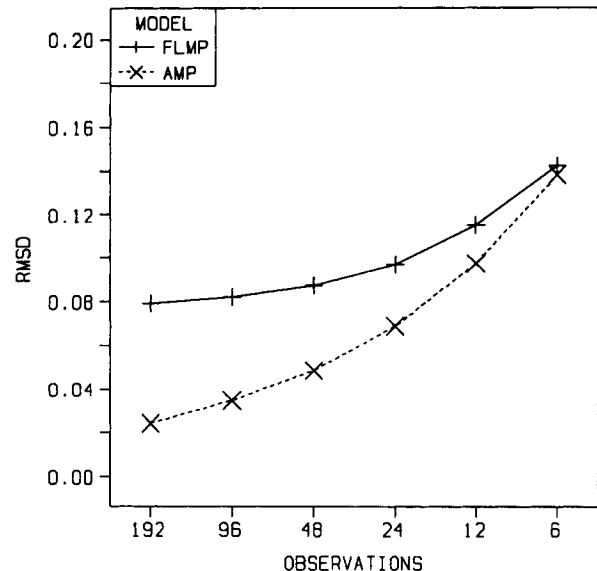

Figure 8. The expected average root mean squared deviation (RMSD) values of the fit of the fuzzy logical model of perception (FLMP) and the additive model of perception (AMP) to simulated results based on the predictions of the AMP as a function of the number of observations per data point. (The predictions were generated for 1,050 hypothetical subjects based on the parameter values obtained from the fit of the AMP to the 21 English speakers in Massaro, Tsuzaki, Cohen, Gesi, & Heredia, in press. Fifty hypothetical subjects were generated for each of the 21 sets of parameters.)

A comparison between the models requires that data generated by one model are fit by the predictions of the other model, as carried out by Massaro (1987, 1988b). Neither model is capable of describing the results of the other model. (It is true that the models are not identifiably different in the middle range of the response scale. Both models predict relatively additive results in the middle range of the response scale. Thus additive results in the middle range can be well described by the FLMP.) Thus, only data that give extreme responses (close to 0 or 1) provide a strong test between the two models. Perhaps one of the reasons that the Cutting et al. (1992) depth judgments did not provide a clear test between the models is that the results did not cover the complete range of the scale.

Fitting Different Functions

Cutting et al. (1992) also compared the AMP and the FLMP in terms of their ability to predict changes in rated depth as a function of the number of sources of information. In their comparison, each source of information was assumed to have the same effect as every other source. The only variable of importance was, therefore, the number of sources of information. Cutting et al. reduced the four-factor design to an analysis of scale values plotted as a function of 0 through 4 sources of information. They generated hypothetical functions describing how ratings change with additional sources of information. The functions were assumed to be negatively accelerating, linear, or positively accelerating, as illustrated in Figure 4 of Cutting et al. (p. 375). The FLMP and AMP were then fit to these results. They found that the FLMP gave a better description of a majority of these functions relative to the AMP. Cutting et al. took this outcome as an illustration of an inherent flexibility in the FLMP not shared by the AMP.

Cutting et al. (1992) discovered that the FLMP gave a better fit to more data sets than did the AMP. We agree completely with this outcome because they simply generated data sets that were more consistent with the FLMP than the AMP. As noted by Cutting et al., the AMP predicts that the change in scale values should be a linear function of the number of sources of information. Most of the functions that they generated were clearly nonadditive and therefore were better fit by the FLMP than the AMP. What Cutting et al. failed to note, however, was that the FLMP also makes the strong prediction that the change in scale values should be a linear function of the number of sources of information if the antilogistic transform of the scale values is taken. Most of the nonlinear functions generated by Cutting et al. are roughly linear when the antilogistic transform of the scale values is taken. There are many other functions that would have given poor fits of the FLMP and better fits of the AMP. What Cutting et al. accomplished, as shown in their Figure 4 was to generate many functions that were consistent with predictions of the FLMP and only a few functions that were consistent with the predictions of the AMP. Thus, it should not be surprising that the FLMP gave a better description. We could easily perform an analogous exercise with different functions and show an advantage for the AMP. Thus, this analysis of Cutting et al. does not show that the FLMP is superpowerful.

Fitting Random Data

Cutting et al. (1992) claimed that the FLMP is good at absorbing noise and fitting random data. They generated 1,000 sets of random data and fit them with the FLMP and AMP. Cutting et al. used a sum of least squares (SOLS) as an index of model fitting. The SOLS is equal to the RMSD squared multiplied by the number of data points (16 in the Cutting et al. study). The mean RMSD was 0.240 for the FLMP and 0.241 for the AMP. The FLMP showed a very slight advantage over the AMP in the model fits (a difference of 0.008 in SOLS and 0.001 in RMSD). Their simulations were interpreted to mean that the noisier the data, the better the FLMP fares. However, this conclusion is unjustified. Both models gave unacceptably large RMSDs to all of the sets of random data. The FLMP does not fit random data, as witnessed by the fact that the fits of the random data were about 8 times poorer than the fits of real data. The FLMP has never given such a poor description of actual results taken to support the model. Thus, the simulations with random noise are not relevant to the good performance of the FLMP. In fact, the simulations show that the FLMP is not superpowerful because it does not give an acceptable description of any possible result. It cannot give an adequate description of random data.

Cutting et al. (1992) also found a negative correlation (−.16) between the difference between the fits of the two models and the SOLS (see Cutting et al., Figure 5, p. 376). This is a small correlation, and it is difficult to predict the direction and slope of the corresponding linear regression line on the basis of only the observed points. We argue that it is not reasonable to do a linear regression from the fit of random data to the fit of real data. The RMSDs come from two different types of data and, therefore, there is no reason to expect a linear relationship across the RMSD values. Small RMSDs only hold for real data, not random data, and there is no justification for extrapolating from random to real data. Just as we witnessed the sins of averaging results across subjects, we note that it is equally dangerous to extrapolate results from one domain to another, especially when the domains might be unrelated to one another.

We replicated Cutting et al.'s (1992) findings on the fit of the models to random data in a factorial design with two factors and seven levels per factor. We wondered if the results would differ for an expanded factorial design. As in the Cutting et al. analysis, 1,000 random data sets were generated and fitted by the FLMP and AMP. The FLMP gave a better fit of 527 of the 1,000 data sets. The average RMSD was only slightly smaller for the FLMP relative to the AMP (0.25407 vs. 0.25463). As we said, we see no significance in this small difference relative to the actual RMSD values.

What is extremely interesting about fitting the FLMP and the AMP to random data, however, is that their RMSDs are very similar for the fit of any given data set relative to the large range of RMSD values across the different data sets.

The correlation between the fit of the FLMP and the fit of the AMP was .971, accounting for over 94% of the variance. This result appears to contradict Cutting et al.'s conclusion that the FLMP absorbs variability relative to the AMP. The FLMP is offended by variability in the same way as the AMP. Random data take on a variety of forms, and what is generally good for the FLMP is also good for the AMP and vice versa. The FLMP does not have magical powers to fit various sets of random data that cannot be fitted by the AMP. Just the opposite must be the case. Very orderly data with small amounts of variability are necessary to distinguish between the FLMP and the AMP.

In much of our previous research, we have contrasted the FLMP with a weighted averaging model (WAM) or a categorical model of perception (CMP; Massaro, 1987, 1989b; Massaro & Cohen, 1990). (It should be noted that Cutting et al's "full averaging model" is not a weighted averaging model but an additive model with the additional constraint that the parameters corresponding to the manipulated variables sum to 1.) The WAM and CMP are mathematically identical with one another and also predict additivity as does the AMP. However, the WAM and CMP have an additional free parameter that allows a differential weighting of the two sources of information. Thus, it should be worthwhile to fit these models to random data and evaluate how they compare with the FLMP. In this case, the FLMP had the disadvantage. The WAM–CMP gave a better fit of 839 of the 1,000 sets of random data. The average RMSD was also smaller for the WAM–CMP relative to the FLMP (0.24757 vs. 0.25407). Thus, the previous victories of the FLMP over these models cannot have been the result of a magical advantage of the FLMP to absorb random variability. The FLMP has been a clear winner over models that should have had the advantage according to the criteria of Cutting et al.

Equation Length and Other Magical Properties

Cutting et al. (1992) hypothesized that equation length somehow accounts for the magical ability of the FLMP to fit results. However, they must remember that correlation does not imply causation. The curiously significant correlation between Super Bowl outcomes and the economy does not imply any causal link between the two. We have already noted that the FLMP provided a somewhat better fit of the 44 individual subjects across the three experiments. The FLMP is also the longer equation by Cutting et al.'s count. Thus, it cannot be surprising that equation length will be correlated with goodness of fit.

As pointed out earlier and elsewhere (Cohen & Massaro, 1992), the FLMP predicts additivity when the antilogistic transform of the response probabilities is taken. It is perfectly reasonable to have both models predict these transformed values. If equations of the FLMP and AMP are now derived to predict these transformed values, the AMP will necessarily be longer. Hence, equation length is to some extent flexible and can have very little relevance to the issues at stake. If the investigator desires a measure of equation length for each model, then it seems fair to allow each model to choose the optimal response measure. In this

case, the FLMP and AMP would have identical equation lengths (for linear and logistic data, respectively), and a direct comparison between them is justified.

New Findings

Although Cutting et al. (1992) focused their article on methodological issues, their new empirical findings provide valuable evidence concerning perception and pattern recognition. They extended their depth-judgment task to include two manipulations. First, the relative frequency of occurrence of the 16 stimuli was varied systematically. In most experiments on pattern recognition and perceptual judgment, the stimuli are presented an equal number of times. In the real world, their probabilities usually differ significantly. Thus, it is important to know whether the FLMP describes these results as well as it does in the case of equal probability. Second, a correlation between the sources of information was introduced. As emphasized by Rosch (1978) and others, in the study of natural categories, feature dimensions tend to be correlated with one another. These correlations are seldom introduced in our experiments on pattern recognition. The FLMP and AMP assume that the different dimensions are evaluated independent of one another. It is possible that this assumption does not hold when correlations are built into the experimental stimuli. Thus, these two experiments are important new tests of the models. Significantly poorer fits of the models to these results relative to the previous experiment of Bruno and Cutting (1988) would expose a serious limitation in the models. Based on the model fits (Cutting et al., Table 4), the goodness of fit of the models does not appear to differ across the three experiments. An ANOVA of the RMSD values (shown in Table 4 in Cutting et al.), with three experiments and two models as factors, indicated no significant effect of experiment or interaction of experiment and model. Thus, the processes involved in distance perception do not appear to change with dramatic differences in stimulus probability and feature correlations. These results increase our belief in the external validity of the findings.

FLMP Versus Directed Perception

Cutting (e.g., 1986, 1991) favors a theoretical stance that he calls *directed perception*. This theory is offered as an alternative to both indirect and direct perception. Like the FLMP, multiple sources of information are assumed to be available. In contrast to the FLMP, however, all sources are not always integrated. Sometimes some sources of information are simply ignored, and only one or several are selected. Additive integration is assumed when the sources are integrated, with different sources receiving different weights.

There is one inconsistency in this theory, however. It is assumed that perception occurs on the basis of multiple sources of information, "each singly and completely specifying what is to be perceived" (Cutting, 1991, p. 26). It is not clear how Cutting is using complete specification in this case. Consider the exocentric distance between the objects in the Bruno and Cutting (1988) study. There were several

picture cues specifying the distance between the objects. However, the results clearly show that no one cue completely specified the distance. The reason is that the distance that was perceived varied depending on the particular combination of cues on a given trial. The objects were rated as being farther apart when several picture cues specified some distance between the objects relative to just one of these cues specifying that distance.

Given that a cue might not always completely specify the perceptual event, advocates of directed perception must clarify the theoretical stance they intend to take. If one accepts the fact that the integration of several cues specifies a different perception from just one of these cues, then it cannot be assumed that each cue completely specifies what is to be perceived. If perception becomes more reliable and accurate with the use of more cues, then a single cue does not completely specify an event and is necessarily ambiguous (as described by Massaro, 1988a). In all other respects, except for the integration algorithm, Cutting's (1991) recent version of directed perception resembles the FLMP. Cutting (1991) made a distinction between ecological and functional information stressed by Brunswik (1956) and emphasized by Massaro (1985, 1987). He also adopted Brunswik's idea of functional equivalence, or the fact that two different sources of stimulus information can have the same effect on perception even though the sources are not equivalent.

In summary, the similarities between directed perception and the FLMP far exceed their differences. The main point of debate has to do with the integration rule. Although there is some question about the rule for visual perception of distance, no such ambiguity exists for other perceptual and cognitive domains (Massaro, 1992). In addition, the results of Dosher, Sperling, and Wurst (1986) provided evidence against additive integration of stereopsis and proximity—luminance covariance in the visual perception of three-dimensional structure. Their measures were transformed to Z scores before being added, which makes this integration roughly equivalent to that assumed by the FLMP (Massaro & Friedman, 1990).

References

Brunswik, E. (1956). *Perception and the representative design of psychological experiments.* Berkeley, CA: University of California Press.

Bruno, N., & Cutting, J. E. (1988). Minimodularity and the perception of layout. *Journal of Experimental Psychology: General, 117,* 161–170.

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

Cohen, M. M., & Massaro, D. W. (1992). On the similarity of categorization models. In F. G. Ashby (Ed.), *Probabilistic multidimensional models of perception and cognition* (pp. 395–447). Hillsdale, NJ: Erlbaum.

Cutting, J. E. (1986). *Perception with an eye for motion.* Cambridge, MA: MIT Press.

Cutting, J. E. (1991). Four ways to reject directed perception. *Ecological Psychology, 3,* 25–34.

Cutting, J. E., Bruno, N., Brady, N. P., & Moore, C. (1992). Selectivity, scope, and simplicity of models: A lesson from fitting judgments of perceived depth. *Journal of Experimental Psychology: General, 121,* 364–381.

Dosher, B. A., Sperling, G., & Wurst, S. A. (1986). Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. *Vision Research, 26,* 973–990.

Massaro, D. W. (1985). *Brunswik and Gibson: Similarities and contrasts with implications for contemporary psychology* (Report 67/1985). Zentrum fur interdisziplinare Forshung, Universitat Bielefeld, Bielefeld, Federal Republic of Germany, Research Group on Perception and Action.

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

Massaro, D. W. (1988a). Ambiguity in perception and experimentation. *Journal of Experimental Psychology: General, 117,* 417–421.

Massaro, D. W. (1988b). Some criticisms of connectionist models of human performance. *Journal of Memory and Language, 27,* 213–234.

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. (1992). Broadening the domain of the fuzzy logical model of perception. In H. L. Pick Jr., P. Van den Broek, & D. C. Knill (Eds.), *Cognition: Conceptual and methodological issues* (pp. 51–84). Washington, DC: 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., & Friedman, D. (1990). Models of integration given multiple sources of information. *Psychological Review, 97,* 225–252.

Massaro, D. W., Tsuzaki, M., Cohen, M. M., Gesi, A., & Heredia, R. (in press). Bimodal speech perception: An examination across languages. *Journal of Phonetics.*

Oden, G. C. (1981). A fuzzy propositional model of concept structure and use: A case study in object identification. In G. W. Lasker (Ed.), *Applied systems and cybernetics* (Vol. VI, pp. 2890–2897). Elmsford, NY: Pergamon Press.

Oden, G. C. (1984). Dependence, independence, and the emergence of word features. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 394–405.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.). *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Erlbaum.

Sidman, M. (1952). A note on functional relations obtained from group data. *Psychological Bulletin, 49,* 263–269.