Journal of Experimental Psychology: General
1988, Vol. 117, No. 4, 417–421

0096-3445/88/$00.75

Ambiguity in Perception and Experimentation

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

Bruno and Cutting (1988) varied four monocular cues to perceived depth in a factorial design. Subjects judged the distance between test objects. Given main effects in the analysis of variance, the authors concluded that the perceivers integrated the four different sources of information, as opposed to simply selecting a single source. Given no interactions in the analysis of variance, the authors concluded that the integration process was additive rather than multiplicative. The ambiguity inherent in Bruno and Cutting's experiments and analyses is discussed. As presented, their results did not provide evidence for integration of depth cues or evidence for additivity, independence, and parallel processing of the cues. An additional analysis of the distribution of the rating judgments given by their subjects, however, provides some evidence for integration of the cues. The fuzzy logical model of perception (FLMP) is extended to describe perceptual recognition of depth. The model assumes independence of the cues during feature evaluation and a nonadditive integration process in which the least ambiguous cues have the greatest impact on the judgment. The FLMP is contrasted with a model assuming additivity of the cues. Because both models describe the results equally well, it remains for future researchers to provide definitive tests between the models.

A person's impressive resolution of the visual world might be a consequence of having multiple sources of information about it. Among other things, these sources of information or cues specify the (exocentric) distance between two objects. Although these cues probably overspecify distance, it is not surprising that typical experiments underspecify what the cues are and how they are processed. Bruno and Cutting's (1988) experimental analysis of exocentric distance was a step forward because it improved on the classic single-factor design. The prototypical experiment in visual perception has been to eliminate or to hold constant all potential cues but one, vary the cue of interest, and observe its behavioral consequences. The single-factor design is weak in information value, however, because it cannot measure the relative salience of the cues or address the processing of multiple cues. Given that all cues but one are neutralized, the single-factor design might also overestimate the contribution of the cue being manipulated. Bruno and Cutting's use of a factorial design has the potential to address several important issues that cannot be addressed by the single-factor design (Anderson, 1981; Massaro, 1985, 1987).

Eliminating Cues to Depth?

Bruno and Cutting (1988) varied four cues to perceived depth in a factorial design and asked subjects to judge the distance between test objects. On the basis of significant effects of the depth cues, Bruno and Cutting concluded that they were integrated. However, their analysis was critically dependent on their assumption that each of the cues was either present or absent in the display. It is possible that the picture cues were always present, in which case additional tests are needed to test for integration. It is important to distinguish between the presence or absence of a cue and whether the cue simply supports one depth or another. One cannot really eliminate many of the cues to depth, if by eliminate we mean make the cues uninformative. The absence of the size cue, for example, simply refers to the squares having identical projective sizes (see Figure 1, Bruno & Cutting, 1988). This identity in projective size provides information that the objects are equal in depth.

Height in the picture plane also suggests one distance or another, rather than being present or absent. Because each of the test squares had some perceptible height, height in the visual field was not eliminated as a cue to depth. Although it seems reasonable to say that occlusion is either present or absent, it is just as reasonable to say that the perceptual system evaluates the degree of overlap among objects. Given this perspective, it is as informative to observe no overlap as it is to observe overlap. The lack of overlap might support no depth between two adjacent objects in the same manner that overlap suggests some depth between the objects. The motion cue appears to differ from the picture cues because a good argument can be made for its being present or absent. Motion was, in fact, present or absent because observers viewed either a static display or a display in motion. The static display did

---

The author would like to thank Nicola Bruno and James E. Cutting for graciously making available the raw data of their study, Michael Cohen for the model analyses and graphs, Michael Cohen and Nancy McCarrell for helpful discussions, and Norman H. Anderson and Geoffrey R. Loftus for comments on a previous draft of the commentary. This commentary is necessarily brief, and the reader is referred to Anderson (1981, 1982) and Massaro (1987) for more comprehensive treatments of the issues addressed in the article.

Correspondence concerning this article should be addressed to Dominic W. Massaro, Program in Experimental Psychology, University of California, Santa Cruz, California 95064.

not suggest one depth and the moving display another. Although Bruno and Cutting (1988) described their stimuli in terms of the presence or absence of four cues to depth, three of the four cues are always present in each of the displays. Each of the three picture cues simply suggests one depth or another, whereas the motion cue was present or absent.

Questioning and Testing Integration

This reinterpretation of Bruno and Cutting's (1988) experimental design is important because it opens to question their conclusion that the observers integrated the cues rather than simply selected a subset of them. It is possible, however, that the subjects did not integrate (use all four cues) on each trial in the Bruno and Cutting study. As argued earlier, the size, height, and occlusion cues are always present in the display; each cue suggests either no difference or some difference in the distance between the objects. With static displays, a perceiver might sample only one of the these three potential cues to depth and base his or her decision on just this cue (the selection strategy rejected by Bruno and Cutting). The same process could occur with the moving displays, except that one of the four available cues would be sampled. The greater proportion of cues in the display suggesting some difference in distance would lead to a greater likelihood of sampling a cue suggesting such a difference. This strategy is consistent with Bruno and Cutting's findings that perceived distance increased with increases in the number of cues suggesting a difference in depth.

Given the possibility that the observer samples only a single cue on each trial, how can this selection process be tested and contrasted with an integration process? The distribution of the rating judgments to each stimulus can be analyzed to test between these two processes (Massaro, 1987). The selection process implies that a given rating response is a function of only a single (probabilistically selected) cue. In this case, the rating to a given stimulus will be sampled from the distribution of the cue selected on that trial. The integration process, on the other hand, implies that a given rating response is based on the outcome of the integration of all available cues. Thus, the ratings for a given stimulus will come from a single distribution reflecting the outcome of the integration of the cues in that stimulus.

To test between integration and selection, I analyzed the distribution of rating judgments in the Bruno and Cutting (1988) study. Each subject gave 10 ratings to each display condition, and these ratings can be analyzed to see if they are better described by selection of a single cue or by integration of multiple cues. The ratings were between 0 (no depth) and 99 (maximum depth). Histograms of these ratings were made by plotting the proportion of ratings in each of 10 equally spaced intervals. Figure 1 gives the distribution of ratings pooled across all subjects for each of the 16 displays.

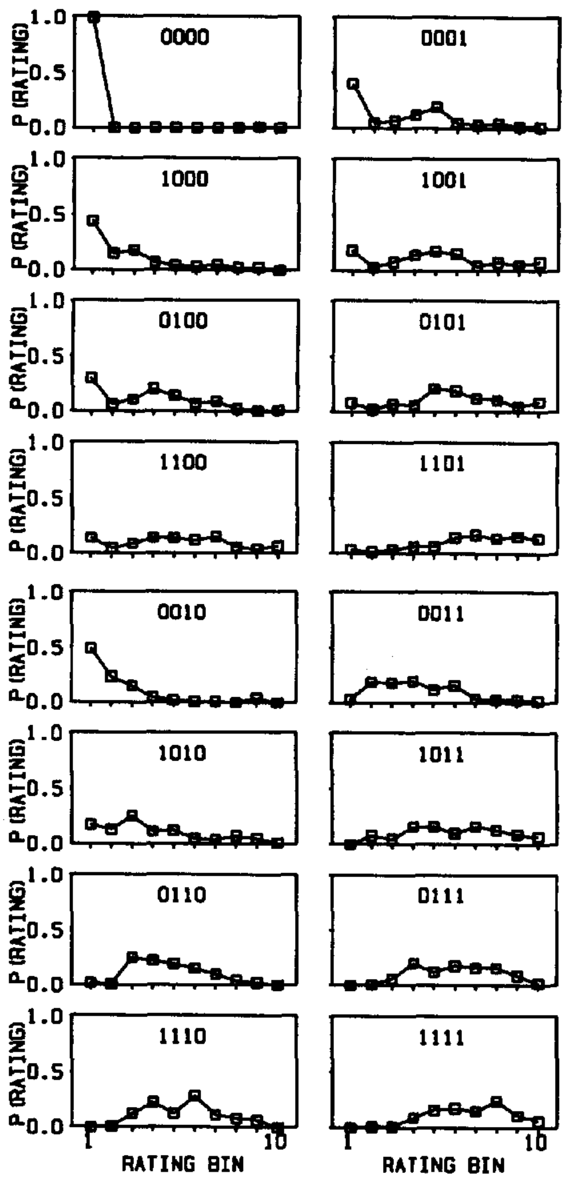

Figure 1. Probability (p) of a rating within each of 10 rating bins for 10 subjects (Bruno and Cutting, 1988, Experiment 1). The label in each cell gives the values of the independent variables size (s), height (h), occlusion (o), and motion parallax (p). The variables are in the order *shop* and can take on the value 0 (no depth) or 1 (depth).

It appears that selection can be rejected even on the basis of the group results. Consider the top four panels in the left hand column of Figure 1, which give the four static displays when there is no occlusion of the test squares. When the size and height indicate no depth (0000), the ratings are consistently near zero. When only size (1000) or only height (0100) provides depth information, about 30% or 40% of the ratings are near zero. The selection strategy says that ratings to the displays with both size and height indicating depth (1100) should be a mixture resulting from these other three conditions (0000, 1000, 0100). It is easy to see that the distribution of ratings to the 1100 display could not have resulted from this type of mixture. There are very few ratings near zero and a significant number of depth ratings above those obtained in the other three conditions. A similar conclusion can be reached given the distribution of ratings to the static display with all three picture cues indicating depth (1110). The distribution of ratings to this display could not have resulted from a mixture of the ratings to Displays 0000, 1000, 0100, and 0010.

To provide a formal test between integration and selection, mathematical models of these two processes were formulated and tested against the rating judgments. The selection model assumed that the rating judgment was determined by only one source of information. Each source had some fixed probability of being sampled on every trial. The rating judgment determined by a given source was assumed to have a mean and variance. Each of the two levels of each of the four sources could have an unique mean and variance. This selection model requires 19 free parameters to predict the distribution of rating judgments. Three parameters are needed for the probabilities of sampling the four sources. (Only 3 rather than 4 parameters are necessary because the four probabilities must add to 1). Sixteen parameters are needed for the eight means and eight variances for the two levels of each of the four cues.

The integration model was based on the fuzzy logical model of perception (FLMP), which is described in the next section of this article. Each level of each source of information was associated with a mean and variance, giving a total of 16 free parameters. The predicted distributions of ratings for each of the 16 conditions resulted from the integration of the four cues, following the integration rule of the FLMP.

The two models were applied to the individual results of each of the 10 subjects. The ratings for each subject were partitioned into 10 equally spaced bins. The models predicted the number of observations in each bin. Given 16 experimental conditions and 10 bins per condition, there are 160 points to be predicted (or 144 independent observations). The predictions of the models were computed by using the parameter estimation routine STEPIT (Chandler, 1969), which finds the parameters that minimize the differences between the predicted and observed values. The criterion for goodness of fit is the root mean square deviation (RMSD).

The integration model gave a better description of the distribution of ratings than the selection model for 9 of the 10 subjects. On the basis of an analysis of variance, the RMSD values for the selection and integration models differed significantly, $F(1, 9) = 21.85, p < .001$. The average RMSD was .1027 for the integration model and .1210 for the selection model. Both the group histograms shown in Figure 1 and the model tests on the individual results support an integration of the cues to depth.

Questioning and Testing Additivity

Bruno and Cutting (1988) concluded that "perceivers use these sources of information in an additive fashion" (p. 161). Additive integration implies that a given cue makes a contribution that is independent of the information value of the other cues. In contrast, research in a variety of domains has shown that the integration process is nonadditive. A given cue makes a bigger contribution to the extent that the other potential cues are ambiguous (Massaro, 1987).

Determining the nature of the integration process is best carried out in the framework of specific models of pattern recognition. One theoretical framework for the present analysis is the FLMP, developed and tested in a variety of pattern-recognition domains (Massaro, 1987). All of these domains have involved categories, such as segmental categories in speech perception, letter and word categories in reading, semantic and syntactic categories in sentence interpretation, and object categories in concept identification. Central to the FLMP are summary descriptions of the categories in the task at hand. These summary descriptions are called prototypes and they contain a conjunction of various properties called features. A prototype is a category and the features of the prototype correspond to the ideal values that an exemplar should have if it is a member of that category. Prototypes are generated for the task at hand. During the first operation in the model, the features are evaluated in terms of the prototypes in memory. For each feature and for each prototype, featural evaluation provides information about the degree to which the feature in the stimulus matches the featural value of the prototype.

Given the necessarily large variety of features, it is necessary to have a common metric representing the degree of match of each feature. Both projective size and height in the picture plane influence perceived depth. These two features must share a common metric if both of them are going to be related to one another or to another category. To serve this purpose, fuzzy truth values (Zadeh, 1965) are used because they provide a natural representation of the degree of match. Fuzzy truth values lie between 0 and 1, corresponding to a proposition being completely false and completely true. The value .5 corresponds to a completely ambiguous situation, .7 would be more true than false, and so on. Fuzzy truth values, therefore, can represent not only continuous rather than categorical information, they can also represent different kinds of information. Although projective size and height are different kinds of information, assigning truth values to them puts them on an equal footing. Another advantage of fuzzy truth values is that they couch information in mathematical terms (or at least in a quantitative form). This allows the natural development of a quantitative description of the phenomenon of interest.

The FLMP postulates three operations between the presentation of a stimulus display and its perceptual interpretation, as illustrated in Figure 2. Feature evaluation provides the degree to which each feature supports some particular depth. The overall degree of depth, as indicated by all of the features, is determined by feature integration. The features, or actually the degrees of matches, corresponding to each depth are combined (or conjoined in logical terms). The outcome of feature integration consists of the degree to which each depth configuration matches the input display. In the

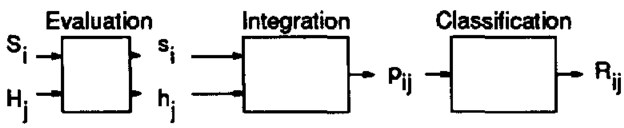

Figure 2. Schematic representation of the three operations involved in perceptual recognition. (The values s_i and h_j refer to the projective size and height cues, respectively; p_{ij} represents the degree to which each depth configuration matches the input display; R_{ij} represents the relative goodness of each depth configuration.)

model, all features contribute to the final value representing the degree of match. The third operation, pattern classification, determines the merit of each relevant depth category relative to the sum of the merits of all the relevant depth categories. This relative goodness of match gives the proportion of times the display is identified as having a particular depth. Similarly, the relative goodness of match can be mapped linearly into a rating judgment indicating the degree to which the display indicates depth.

The application of the FLMP to the Bruno and Cutting (1988) study is relatively straightforward. The four independent variables of size, height, occlusion, and motion parallax support either no depth or some depth. In addition to these cues, there is also a conglomeration of other cues not being manipulated by the experimenter. This conglomeration of cues might be expected to bias perception in one direction or the other. Given that subjects viewed a computer screen with binocular vision, one can expect the bias (b) to be in the direction of no depth. In this case, the value of b should be less than .5. Thus, the two possible interpretations of depth can be defined as a function of the four cues being manipulated and the constant background of other cues.

Depth: (appropriate size) and (appropriate height) and
(appropriate occlusion) and (appropriate motion parallax)
and (appropriate background) = s and h and o and p and b.

The interpretation of no depth can also be defined as a function of the same five cues. In its simplest form, the perception of no depth can be described as the conjunction of the negations of the five cues for the perception of depth.

No depth = $(1 - s)$ and $(1 - h)$ and
$(1 - o)$ and $(1 - p)$ and $(1 - b)$

In this version of the FLMP, five parameters are necessary to predict the 16 data points corresponding to the 16 unique displays. One parameter is necessary to represent the degree to which each of the five cues supports depth.

The parameter values are assumed to be fuzzy truth values between 0 and 1 that indicate the degree to which each of the five cues support depth. According to the FLMP, the conjunction of the five cues is assumed to be multiplicative, followed by the relative goodness determination given by the pattern classification operation. The first experiment of Bruno and Cutting (1988) asked for rating judgments indicating an estimation of the relative distance between the test squares. The average rating judgment is assumed to reflect the relative goodness of the alternatives determined by the pattern classification operation. Similarly, the rating could be assumed to reflect the degree of depth. In both of these cases, the rating is predicted to be

$$R(\text{Depth}) = \frac{shopb}{shopb + (1-s)(1-h)(1-o)(1-p)(1-b)}. \qquad (1)$$

Subjects chose a rating between 0 (no depth) and 99 (maximum depth). The rating judgment can be transformed into the relative goodness values predicted by the model by dividing the rating by 100. Observed and predicted values will be between 0 and .99.

Within the framework of the FLMP, each of the three picture cues supports either one depth or another. As argued earlier, the motion parallax cue is either present or absent. Thus, the value of p in Equation 1 must be set to .5 in the static-display conditions. Setting the truth value to the completely ambiguous value .5 is mathematically identical to eliminating the cue (Massaro, 1987, p. 167).

This framework can also be extended to formalize an adding model (favored by Bruno and Cutting, 1988). In this case, the cues are simply added.

$$R(\text{Depth}) = s + h + o + p + b. \qquad (2)$$

In the adding model, each of the five parameters takes on some positive value if the depth is supported by the corresponding source of information and is zero otherwise. For example, the average rating to the display 0100, with only height indicating depth, would be equal to $h + b$.

Both the FLMP and the adding model require five free parameters, making their comparison straightforward. The FLMP and the adding model were tested against the results of Bruno and Cutting's (1988) first experiment. Each of the 10 subjects made 10 ratings of each of the 16 test displays. The mean rating for each subject under each of the 16 experimental conditions served as the dependent variable to test the models. The predictions of the models were computed as in the test between the selection and integration models.

Table 1 gives the average observed results and the average predictions of the two models. The RMSD values for the two models are given in Tables 2 and 3. The two models gave

Table 1

Average Ratings as a Function of Depth Cues Size (s), Height (h), Occlusion (o), and Motion Parallax (p) in the Bruno and Cutting (1988) Study; the Rating Judgments Reflect the Perceived Degree of Depth and the Average Predictions of the FLMP and Adding Model

	Model			
Depth cues	FLMP		Adding	
s h o p	Obs	Pre	Obs	Pre
0 0 0 0	.009	.125	.009	.067
0 0 0 1	.253	.261	.253	.255
0 0 1 0	.139	.179	.139	.171
0 0 1 1	.342	.327	.342	.359
0 1 0 0	.271	.257	.271	.260
0 1 0 1	.495	.471	.495	.448
0 1 1 0	.382	.317	.382	.364
0 1 1 1	.523	.533	.523	.552
1 0 0 0	.176	.212	.176	.194
1 0 0 1	.411	.392	.411	.382
1 0 1 0	.306	.280	.306	.298
1 0 1 1	.515	.466	.515	.486
1 1 0 0	.423	.396	.423	.388
1 1 0 1	.607	.619	.607	.575
1 1 1 0	.464	.460	.464	.491
1 1 1 1	.591	.684	.591	.679

Note. FLMP = fuzzy logical model of perception; Obs = average observed degree of depth rating; Pre = average predicted degree of depth rating.

fairly similar descriptions of performance, with a slight edge in favor of the FLMP. An analysis of variance on the RMSD values showed no difference between the two models. The model tests show, therefore, that the results do not warrant a rejection of nonadditivity, as instantiated in the FLMP.

The parameter values for the fit of the FLMP are shown in Table 2. The parameter values correspond to the degree of support for seeing a depth difference for the level indicating a difference. The parameter value for the level indicating no depth difference is simply 1 minus the parameter value, except for the motion variable, which is .5 for the static displays. The parameter values are reasonable and give an index of the magnitude of the effect of each variable for each subject. These values show significant individual differences that would tend to dilute the group results. Three subjects (4, 7, and 9) appear to be responsible for the lack of a group effect for occlusion. These subjects actually perceived (or at least rated) depth as being greater for nonoccluded than for occluded test squares. The parameter values for *b* are smaller than .5 for all subjects, which supports the earlier speculation that the constant background of cues should support no depth.

Table 3 gives the parameter values for the adding model.

Table 3

Root Mean Square Deviation (RMSD) Between Observed and Predicted Ratings, and Best Fitting Parameters of the Adding Model for the Constant Background Cue (b), and Four Depth Cues Size (s), Height (h), Occlusion (o), and Motion Parallax (p)

Subject	RMSD	*b*	Depth cues			
			s	*h*	*o*	*p*
1	.0803	.0824	.1175	.2375	.1378	.2645
2	.0443	.1011	.1008	.1593	.0760	.1708
3	.1169	.0001	.2346	.2971	.0429	.0936
4	.0954	.0001	.1382	.2266	.0001	.1926
5	.0967	.1662	.2182	.2540	.1011	.0620
6	.0714	.0648	.1512	.2943	.0628	.1723
7	.1404	.2565	.1340	.0945	.0001	.1393
8	.0365	.0016	.0596	.1389	.1021	.3694
9	.1217	.0001	.0365	.2104	.0001	.3662
10	.0728	.0001	.814	.0190	.5115	.0477
Average	.0876	.0673	.1272	.1932	.1035	.1878

Note. Each parameter represents the degree of support of the cue for a depth rating.

Parallel Processing

Bruno and Cutting (1988) saw an additive integration as generally consistent with parallel processing models of the visual system. It is important to acknowledge that there is no convincing evidence for parallel processing from their experiments, in the same way that there is no convincing evidence for additive integration. Even if integration were demonstrated unambiguously for Bruno and Cutting's results, methodological details (that might easily be overlooked) compromise any conclusion about parallel processing. Observers had the option of viewing such stimulus as many times as they wished before making a judgment about a particular stimulus. It remains a possibility that subjects did not process the cues in parallel but processed them serially one at a time across repeated presentations of the stimulus on a given trial and then integrated these separate views. Eliminating the possibility of repeated presentations in Bruno and Cutting's task would not have been sufficient, however, to address the issue of parallel processing. The reason is that each display was excessively long (2,088 ms). Observers would have plenty of time to switch among the cues over the course of the long stimulus presentation. A single look at about three frames of their display (261 ms) would have been a more reasonable test of parallel processing. Even if integration is demonstrated in their task with long displays and multiple views, serial processing of the cues remains a viable possibility.

Table 2

Root Mean Square Deviation (RMSD) Between Observed and Predicted Ratings, and Best Fitting Parameters of the FLMP for the Constant Background Cue (b), and Four Depth Cues Size (s), Height (h), Occlusion (o), and Motion Parallax (p)

			Depth Cues			
Subject	RMSD	*b*	*s*	*h*	*o*	*p*
1	.0829	.3242	.5704	.6319	.5759	.7677
2	.0542	.2692	.5529	.5842	.5415	.6777
3	.1139	.2370	.6646	.6984	.5159	.6352
4	.0810	.1471	.6118	.6490	.4249	.7647
5	.1003	.4621	.6146	.6323	.5535	.5632
6	.0819	.3198	.5850	.6542	.5393	.6693
7	.0951	.3618	.5741	.5501	.3874	.6485
8	.0441	.1515	.5315	.5840	.5677	.8598
9	.0976	.0635	.5566	.6806	.4514	.9345
10	.0691	.1936	.5652	.4914	.8259	.6094
Average	.0820	.2530	.5827	.6156	.5383	.7130

Note. Each parameter represents the degree of support of the cue for a depth rating. FLMP = fuzzy logical model of perception.

References

Anderson, N. H. (1981). *Foundations of information integration theory.* New York: Academic Press.

Anderson, N. H. (1982). *Methods of information integration theory.* New York: Academic Press.

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.

Massaro, D. W. (1985). *Brunswik and Gibson: Similarities and contrasts with implications for contemporary psychology* (Report no. 67/1985). Bielefeld, FRG: Research Group on Perception and Action, Zentrum für interdisziplinare Forschung, Universität Bielefeld.

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

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

Received January 26, 1988
Revision received April 13, 1988
Accepted April 25, 1988 ■