Journal of Experimental Psychology: General
1988, Vol. 117, No. 2, 161-170

0096-3445/88/$00.75

Minimodularity and the Perception of Layout

Nicola Bruno and James E. Cutting
Cornell University

SUMMARY

In natural vision, information overspecifies the relative distances between objects and their layout in three dimensions. Directed perception applies (Cutting, 1986), rather than direct or indirect perception, because any single source of information (or *cue*) might be adequate to reveal relative depth (or local depth order), but many are present and useful to observers. Such overspecification presents the theoretical problem of how perceivers use this multiplicity of information to arrive at a unitary appreciation of distance between objects in the environment.

This article examines three models of directed perception: selection, in which only one source of information is used; addition, in which all sources are used in simple combination; and multiplication, in which interactions among sources can occur. To establish perceptual overspecification, we created stimuli with four possible sources of monocular spatial information, using all combinations of the presence or absence of relative size, height in the projection plane, occlusion, and motion parallax. Visual stimuli were computer generated and consisted of three untextured parallel planes arranged in depth. Three tasks were used: one of magnitude estimation of exocentric distance within a stimulus, one of dissimilarity judgment in how a pair of stimuli revealed depth, and one of choice judgment within a pair as to which one revealed depth best.

Grouped and individual results of the one direct and two indirect scaling tasks suggest that perceivers use these sources of information in an additive fashion. That is, one source (or cue) is generally substitutable for another, and the more sources that are present, the more depth is revealed. This pattern of results suggests independent use of information by four separate, functional subsystems within the visual system, here called minimodules. Evidence for and advantages of minimodularity are discussed.

How do humans perceive distances between objects? Since the work of John Locke and George Berkeley there has been ample discussion of *signs* to depth, later called *cues* by William James and Edward Titchener (see Cutting, 1986; Pastore, 1971). Since Gibson's (1950) work, there has been ample discussion of information as it reveals layout (see Cutting, 1987). Depending on the framework within which one works, humans perceive objects in depth because cues suggest the relations among objects, or because the information specifies the layout. On the former view the relation between cues and object properties is stochastic (Brunswik, 1956), and the mapping between them many-to-many; on the latter the relation between information and object properties is deterministic and the mapping between them either one-to-one (Gibson, 1979) or many-to-one (Cutting, 1986).

But regardless of whether one traffics in cues or in information, the environment within which human perception evolved is rich. This richness can serve the visual system well, but it poses theoretical problems seldom addressed by the perception researcher (Epstein, 1977). In this article, we explore the conjunction of four sources of monocular information about depth: relative size, height in the projection plane, occlusion, and motion parallax. How is this multiplicity dealt with? Does the visual system optimize its efficiency by using only a single best source? Does it combine sources? Does it use interactive combinations? Does its strategy vary according to context, or according to information strength?

We cannot answer all these questions (particularly the last) but the following picture emerges from our data and consideration of results elsewhere: Information is gathered by separate visual subsystems—here called *minimodules*—and it is added together in the simplest manner, without regard to feedback across subsystems. The linchpin of the argument is additivity, and to discuss additivity we must start with the theory of information integration.

Oral versions of this article were read before the Fourth International Conference on Event Perception and Action, Trieste, Italy in August 1987, and the 28th Annual Meeting of the Psychonomic Society, Seattle, Washington in November 1987.

This research was supported by the National Institutes of Mental Health Grant MH37467 to James E. Cutting.

The authors are indebted to Geoffrey Loftus, Dominic Massaro, and an anonymous reviewer for commenting on an earlier version of the manuscript.

Correspondence concerning this article should be addressed to James. E. Cutting, Department of Psychology, Uris Hall, Cornell University, Ithaca, New York 14853-7601.

Integrating Information

The problem of how perceivers combine multiple sources of information to produce a response has been studied in many ways (e.g., see Ashby & Townsend, 1986; Birnbaum, Wong, & Wong, 1976; Brunswik, 1956; Garner, 1973; Massaro, 1987; Massaro & Cohen, 1983), but the framework adopted here is a modification of information integration theory (Anderson, 1974a, 1974b). This framework describes integration processes with algebraic models and tests them with functional measurement. Factorial designs are preferred over correlational studies, because correlations can provide misleading indices of fit (Anderson & Shanteau, 1977; Birnbaum, 1973). Functional measurements are used to validate both the model and the response scale, yielding a description of the integration process.

Within information integration theory, models divide into two major classes. One consists of adding, subtracting, and averaging models that express the integration process as a weighted sum of information components; the other consists of multiplying and dividing models that use joint addition and multiplication rules. Analysis of variance provides a straightforward test for both; regression can estimate subjective weights. In the following analysis, we borrow from information integration theory to describe how sources of depth information might be combined.

As an entrée, consider the following: In a situation where two sources of depth information s_1 and s_2 are present, an integration model could be written in its most general form as

$$d = f(c\,[w_1 * s_1,\ w_2 * s_2]), \tag{1}$$

where d is an exocentric distance to be determined, w_1 and w_2 are weights assigned to the sources, c is a combination rule, and f is the function that maps combined information to percept.

As can be seen, there are several components involved in assessing information integration. Massaro (1987) suggests that in these situations one must consider three factors: the evaluation process (which we take here to be equivalent to assessing weights, w), the integration process (which is reflected in the combination rule, c), and the classification process (which generates the response and is represented by f). However, the major concern in this article is the combination rule. Classification effects are set aside by varying responses across three scaling tasks. If the same set of results accrue across varied tasks, then the classification component contributes little to the results. Systematic assessments of weights are also set aside because the relative importance of information in depth perception seems subject to many factors, including subjective set (Gilinsky, 1951), context (Dosher, Sperling, & Wurst, 1986), state of adaptation (Wallach, 1976), and task (Cutting & Millard, 1984), not to mention the scale values of each source (the amount of difference in the optic array between relative size, height in plane, occlusion, and parallax). Our strategy to study the combination rule, then, is to present readily discriminable differences between sources that are present or absent in the display. Consider three combination rules together with their predictions: selection, additivity, and multiplication.

Selection

Selection is, in fact, not a strategy of integration; observers may simply use the single most effective available source and disregard the others. Nevertheless, it is a strategy worthy of our interest. It is one that some animals appear to adopt in spatial localization (Knudsen & Konishi, 1979), and that humans may adopt for the classification of surfaces as curved (Cutting & Millard, 1984; Todd & Akerstrom, 1987) or for judgments of flatness of surfaces that translate as opposed to rotate (Cutting, 1986). It is also the strategy that can be most closely allied with Gibson's (1979) idea of invariant pickup. Such a strategy is useful if limited neural resources can be allocated to a task, or if one source of information is more stable (Koffka, 1935), like an invariant, than others.

Algorithmically, if d is the distance to be determined and s_1 and s_2 are sources of information with s_1 being the only one that is useful, we can write a selection hypothesis as

$$d = f(s_1) \tag{2}$$

Statistical support for selection would be the reliability of one main effect with no interactions in analysis of variance.

Addition

Observers might process all information sources, weight them, and then add results to form the percept. (In this context, selection could be regarded as a special case of additivity where some weights are zero, an idea that we will return to later.) Evidence for additive combination of spatial information has been found elsewhere (Cutting & Millard, 1984; Dosher et al., 1986; van der Meer, 1979). Because it allows for parallel processing, additivity is the most efficient model of computation (pickup) of information.

Symbolically, the simplest additive strategy is

$$d = f(w_1 * s_1 + w_2 * s_2) \tag{3}$$

In the terminology of systems analysis, this is an instance of the superposition principle, a criterion for linearity of a system (Kaufman, 1974). Analysis of variance diagnoses additivity in the data (Anderson, 1974a); reliable main effects without interactions is the pattern to look for.

Multiplication

Observers might use some sources to correct information from other sources. Motion parallax, for example, might be used with other cues to derive absolute distances (Nakayama, 1983; Ono, Rivest, & Ono, 1986), or to disambiguate depth ordering in parallel projections (Farber & McConkie, 1979; Braunstein, Andersen, Rouse, & Tittle, 1986). Multiplicative rules offer the best way for such interactions, because they may provide ways for information to be sensitive to contexts, as in the many Ames demonstrations (Ames, 1955).

The simplest multiplication may be written as

$$d = f(w_1 * s_1 * w_2 * s_2). \tag{4}$$

More realistic models may combine addition and multiplication in various ways, such as $d = f(w_1 * s_1 + w_1 * s_1 * w_2 *$

s_2), or even $d = f(w_1 * s_1 + w_2 * s_2 + w_1 * s_1 * w_2 * s_2)$. Interactions in analysis of variance support multiplicative models. Joint inspection of interaction patterns and main effects indicates whether observers used simple multiplication or a combination of multiplication and addition.

Optics of the Simulation

In this article, perceived exocentric distances are assessed as a function of relative size, height in the projection plane, occlusion, and motion parallax. We chose this combination for a number of reasons: It includes both static and motion information, which might be regarded as fundamentally different; it includes sources that may seem to rely on cognitive assumptions (such as relative size) as well as those that do not; and most important, to our knowledge it is a longer list than has ever before been rigorously assessed in visual perception.

We used polar projective views of a simplified environment simulated with a computer-controlled graphics display, with any given source either present or absent. This yields 2^4, or 16, stimuli. It will now be obvious why this is a modification of functional measurement: Functional measurement demands at least 3 values along a given dimension. To follow this stricture would entail an unwieldy value of 3^4, or 81, stimuli in each experiment.

The simplified environment had three square panels, (initially) 2 arbitrary units on a side, and no visible texture on their surfaces. These were laid out parallel to the projection plane in depth at equal exocentric distances, $d_{12} = d_{23} = (d_{13})/2 = 3$ units (where subscripts enumerate panels, near to far). If the observer were stationary, his or her location was 21 units away along the z axis orthogonal to the center of the middle panel. If the display showed differences in *height in plane*, then the observer's eye was 6 units above the ground plane (along the x axis), but if there were no eye height information the observer's eye was 0 units above the ground plane.

If the observer was moving, his or her linear path ran parallel to the planes, ending at the same point as that for stationary stimuli and starting 21 units to the right along the x axis. Observer movement was smooth across 24 frames with eye fixation at the center of the middle panel as if the eyes were following it with smooth pursuit; such a situation creates *motion parallax*. These displays provided a simulated-fixation experience somewhat similar to watching a film sequence taken with a mobile camera, mounted in the right front seat of a slowly moving car, and panning to keep a stationary object in the middle of the image (Cutting, 1986; Regan & Beverley, 1982).

If *occlusion* information was present in the display, the three panels were slid closer in alignment, each having edges that overlapped its neighbor along the x axis by 1 unit; if no occlusions were present, the edges of the panels were separated along the x axis by .2 units.

If *relative size* information was present, the projected differences in retinal size were appropriate for planes 2 units on a side; if no relative size information was present, the middle and far planes were enlarged so that their optic projections were identical to that of the near panel from the observation point of the static stimuli, which was also the last observer position for moving stimuli.

Figure 1 illustrates observer movement with respect to the three planes, Figure 2 shows the eight static stimuli, and Figure 3 shows the starting positions of the eight motion stimuli. Static stimuli consisted of 24 identical frames; motion stimuli were generated by moving the observer .913 units along the path. End frames of all stimuli subtended about 8° both vertically and horizontally.

Throughout the presentation, results are collapsed across stimuli according to the number of sources of information in them. There is one stimulus with no sources (which we label 0), four with one source (S for that with relative size alone, H for height alone, O for occlusion alone, and P for motion parallax alone), six with two sources (SH, SO, SP, HO, HP, and OP), four with three sources (SHO, SHP, SOP, and HOP), and one with four sources (SHOP).

Stimuli were generated on a Hewlett-Packard 1000L series computer and displayed on a Hewlett-Packard 1350s vector-plotting system with a P31 phosphor. They were seen binocularly in a moderately lit room, with the sides of the monitor visible and thus forming a frame around the images.

Experiment 1: Direct Judgments of Distance

Perceived exocentric distances (those between objects in the environment) were assessed with a magnitude estimation procedure.

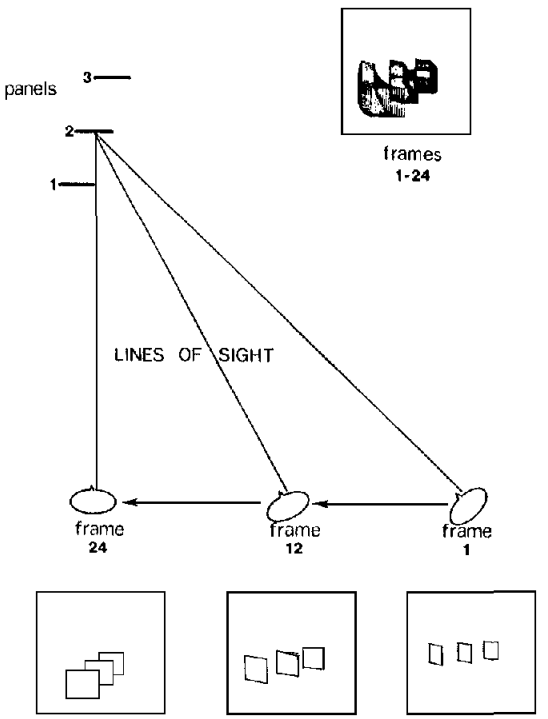

Figure 1. Movement of an observer with respect to the three planes. (Frames, 1, 12, and 24 are reproduced below the corresponding positions; all frames are superimposed in the upper-right inset panel.)

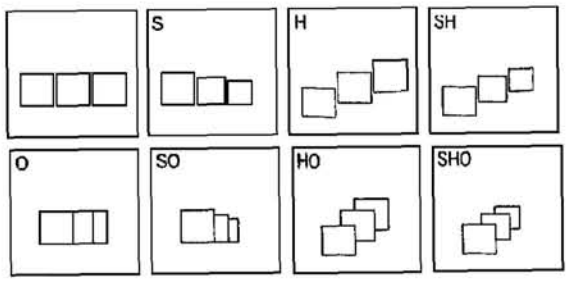

Figure 2. The eight static stimuli, with relative size, height in plane, and occlusion manipulated orthogonally. (S = relative size; H = height in plane; O = occlusion).

Method

Ten members of the Cornell University community participated individually for about an hour and received payment. All had prior experience with psychological experiments, but none were familiar with the purposes of the study. Viewers were told that the task involved estimating the relative distance between three square panels on a 100-point scale, with 0 indicating no distance and 99 the maximum possible exocentric separation. Viewers were shown several times the stimulus with all four sources of information to adjust their scales and then given a 20-trial practice block. The experimental session presented the 16 displays 10 times each, randomly ordered, totalling 160 trials. Each display was presented for 2,088 ms, one frame lasting 87 ms. The procedure allowed each stimulus to be presented many times during a given trial at the option of the observer. Observers were encouraged to inspect the stimulus as many times as needed.

Analysis, Results, and Discussion

Distance ratings for every stimulus were averaged within each observer for each stimulus. The 10 data sets were then intercorrelated to assess consistency among viewers. Correlations were satisfactory (mean $r = .565$, $p < .05$). Next, the individual means for each of the 16 stimuli were entered in a four-way analysis of variance. Three main effects were found: relative size, height in plane, and motion parallax, $Fs(1, 9) = 48.26, 48.35, 27.26$, respectively, all $ps < .001$. Occlusion provided no reliable effect by this measure, $F(1, 9) = 1.7$, $p = .22$. More important, however, is the fact that no first-, second-, or third-order interactions were reliable (all $ps > .09$).

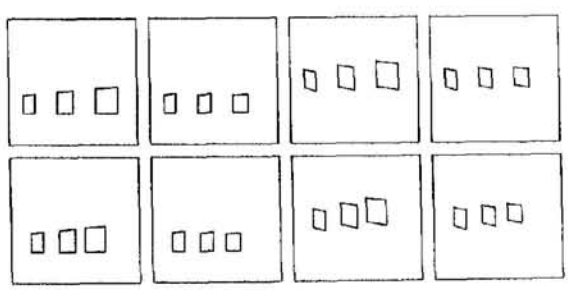

Figure 3. The first frame (of 24) of the eight stimuli with motion parallax. (The final frames of each were identical to the static stimuli shown in Figure 2.)

Next, mean distance judgments for the 16 stimuli were regressed against the four dichotomous variables. The multiple correlation for the four experimental variables was high ($R = .98$). Each source of information was a reliable predictor of the mean data (all $ps < .003$; $\beta = .41, .61, .24$, and .60) for size, height, occlusion, and parallax, respectively. The reliability of an occlusion effect here, but not in the analysis of variance results, suggests somewhat more variability across observers in judgments of stimuli with occlusions; but the same pattern can be seen in the effects for motion parallax as well.

The proportion of variance accounted for by each source in the regression equation is shown in the top panel of Figure 4. Whereas these values reflect the relative importance of the four sources in this particular task and with these particular displays, they should be taken only as an indication that all four sources were used. Again, the purpose here is not to assess the general relative importance of each source, only to assess how different and discriminable sources are used.

As suggested earlier, the best way to display the mean data is to collapse across those stimuli that have the same number of sources of information. If additivity holds, then a linear function should be found. Results for this study are shown in the left panel of Figure 5, with reasonable linearity shown. A more powerful test, however, is to test individual means. Judgments within observers should follow an ordered set of inequalities: $0 < 1 < 2 < 3 < 4$, where each integer indicates means for those stimuli with that number of sources of information. The a priori likelihood of such a sequence is small, 1/5! (or 1/120), but 9 out of 10 of our observers yielded this pattern.

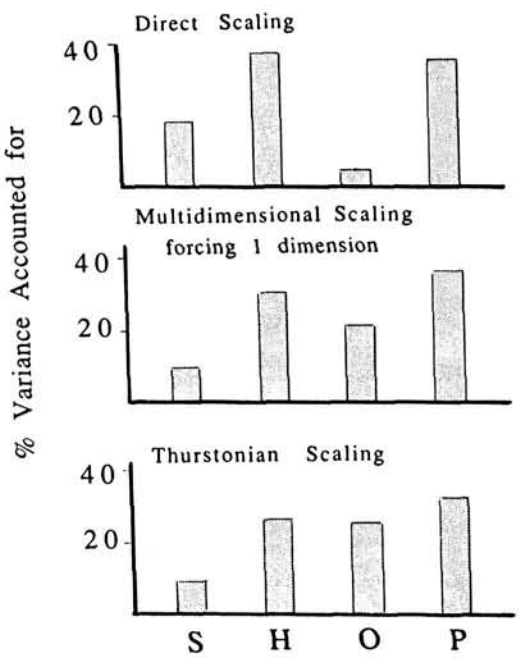

Figure 4. The variance accounted by the four sources of information in a multiple regression on the mean scale values in each of the three experiments. (S = size; H = height; O = occlusion; P = motion parallax).

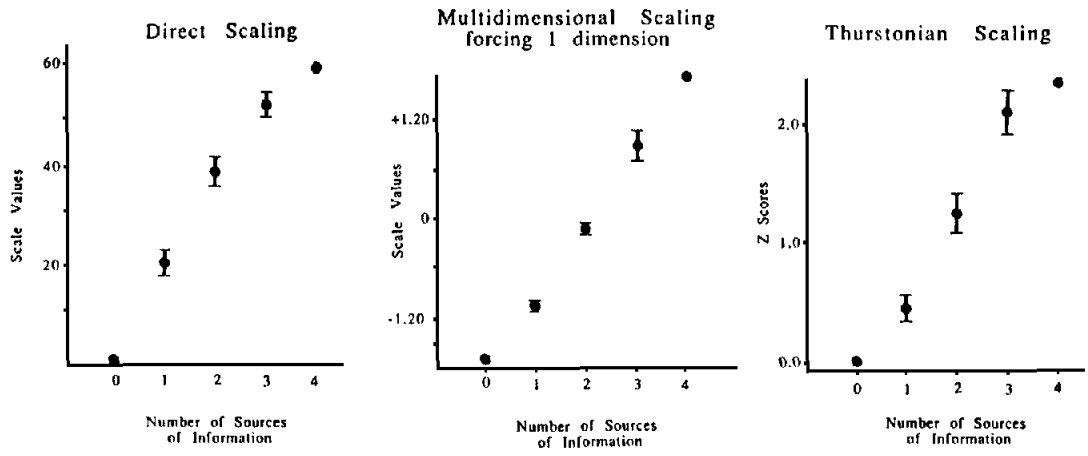

Figure 5. The mean scale values (and standard errors) for stimuli with 0 through 4 sources of information in each of the three experiments.

Now consider these results with respect to the three models discussed earlier. First, because no single source of information dominated the pattern of results, direct scaling values provide no support for a strategy of selection. The relatively high interobserver correlations, together with a detailed inspection of individual means, showed a consistent pattern across observers, ruling out the possibility that different observers used different selection strategies. Second, the lack of interactions impugns a multiplication model. Thus, these results are fit best by an additive strategy, with all sources being relatively strong and noninteracting determinants of the distance judgments. Furthermore, these results agree well with previous assessments of additivity, particularly those of Holway and Boring (1941), Harker (1958), and Jameson and Hurvich (1959), who used some of these sources of information but without an information integration approach.

Nonetheless, the additive results here and earlier could merely reflect the demand character of the experimental situation (Orne, 1962). Because the four sources of information are easily discerned in the displays, participants could have approached the experimental situation as a task of counting sources and allotting scale values accordingly. Such a strategy would not reveal anything important about the stimuli, but would reflect only the ability of participants to take discernable sources into account in a situation of uncertainty. To help assure that the results were not due to such an artifact, we performed two more experiments that used indirect assessments of perceived distances.

Experiment 2: Indirect Assessment of Distance Using Dissimilarity Ratings

Observers were asked to compare exocentric distances in pairs of stimuli and provide a dissimilarity rating of them. Comparisons were then scaled in various dimensions. If the results of Experiment 1 were a reflection of the experimental demand of direct judgments, a different pattern might be found using an indirect-judgment procedure. On the other hand, if the results of Experiment 1 are a proper measure of multiple information use, a similar pattern should be found.

Dissimilarities, of course, can be manifested in many ways. Because multidimensional scaling allows results to be scaled as more than just a linear vector, we might expect to see intrepretable patterns in the scaled data in as many as four dimensions, corresponding to those of stimulus manipulation. Such a solution, however, would not be informative because it would simply reflect the dimensions that were built into the stimuli. To provide a useful assessment of multiple information use, the multidimensional scaling solution should be interpretable as reflecting perceived distances. Thus, a reasonable collapse of all stimulus dimensions into a one-dimensional solution would provide the strongest test of additivity.

Method

Ten different members of the Cornell community were paid for their individual participation. Again, all had prior experience with psychological experiments, but none were familiar with the purpose of the study. Stimulus displays and apparatus were the same as Experiment 1. Unlike Experiment 1, however, pairs of stimuli were presented sequentially.

The experiment consisted of 240 randomly ordered trials, corresponding to the 120 combinations of 16 stimuli in the two possible orders of successive presentation (AB and BA). No stimulus was paired with itself. For each trial, observers were asked to estimate the dissimilarity of the pair of stimuli as they revealed exocentric distance and enter the judgment on the console using a 9-point scale, a judgment of 1 representing minimal or no difference, and 9 maximal difference. As in Experiment 1, trials could be repeated at the observer's discretion. The experiment lasted approximately 1.5 hr, depending on the number of presentations required by the observer before reaching a decision in each trial.

Analysis, Results, and Discussion

Raw data from each observer began as a 16 × 16 matrix with the major diagonal missing. Because no order effects or asymmetries were apparent, each matrix was then folded, providing 10 half matrices of mean dissimilarity ratings. Consistency between viewers was then assessed by computing

correlation coefficients between all half matrices. Although the mean correlation was reliable ($r = .35$, $p < .05$), 13 of the 45 individual correlations were not ($\alpha = .05$). Thus, there is a certain amount of noise in the data, mostly likely because there were only two pairings of stimuli in each cell of each half matrix.

Dissimilarities were normalized within each observer to correct for possible differences in use of the scale, then averaged across the 10 individuals, yielding a single half matrix that formed the basis for multidimensional scaling (ALSCAL). Nonmetric scaling was selected and the data scaled in one, two, and three dimensions. Kruskal's Formula 1 stress values were .248 for one dimension, .148 for two dimensions, and .089 for three dimensions, indicating that the greatest reduction in stress was brought about by the two-dimensional solution. This solution is shown in Figure 6. The vertical dimension captures the numbers of information sources and a correlated diagonal dimension divides static and moving stimuli.

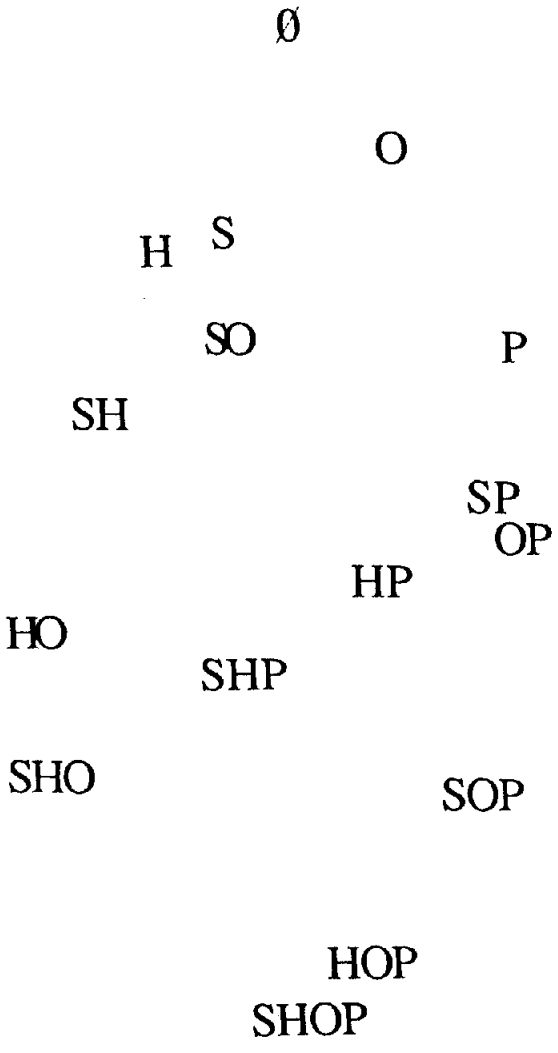

Figure 6. The two-dimensional scaling solution from mean dissimilarity ratings.

Because of the correlation of these two dimensions we find a one-dimensional solution preferable, and rationalize it on three grounds. The first concerns statistical reliability. Monte Carlo studies with randomly generated entries in half matrices of the size we used (De Leeuw & Stoop, 1984; Klahr, 1969; Levine, 1978; Wagenaar & Padmos, 1971) indicate that a very reliable one-dimensional structure is present in our data. The second concerns interpretability (Shepard, 1974, 1980). The correlated dimensions in Figure 6 are redundant in that both reduce to number of sources of information in the stimuli. The third concerns the high correlation of these one-dimensional scale values with the magnitude estimations of the first experiment ($r = .92$). Thus, we feel confident that the one-dimensional solution reflects the scaling of distances. Collapsing scale values across those stimuli with the same number of information sources, the linear trend of the data is clear, as shown in the central panel of Figure 5. Such a pattern is again consistent with additivity of information.

As a further test for additivity the half matrices of each observer were scaled in one dimension. These individual solutions had a median stress of .267, only slightly above that for the pooled data. When scale values were compared across all observers the intersubject correlation was much higher (mean $r = .516$, $p < .001$) than for the individual half matrices. In 7 out of 10 observers the set of inequalities concerning scale values and number of information sources ($0 < 1 < 2 < 3 < 4$) was upheld, demonstrating that the pattern shown in the middle panel of Figure 5 is representative of most observers.

Individual scale values were then entered in an analysis of variance in a fashion analogous to Experiment 1. The results replicated closely those of the previous experiment, with all main effects now reliable: size, height, occlusion, and parallax, Fs(1, 9) = 13.8, 25.0, 30.6, 13.8, respectively, all ps $< .005$. Again, there were no reliable interactions (all ps $> .19$). Multiple regression of the four sources on mean scale values across observers based on one-dimensional dissimilarity scaling yielding excellent fit ($R = .96$), leaving essentially no variance accountable for by interactions. Patterns of weights are shown in the middle panel of Figure 4.

These results complement those of the previous study. Again, there is strong evidence for additivity and none for either selection or multiplication of information sources. Still, to be assured of the generality of the results, we conducted a simpler indirect scaling experiment, one that also reflects better what viewers are likely to do with distance information in natural situations.

Experiment 3: Indirect Assessment of Distance Using a Choice Task

Another set of judgments was obtained by having observers choose the stimulus in each pair that revealed the greater distance. Such a task resembles more closely a naturalistic evaluation of what the layout affords: which two trees are closest together, what aperture is possible to walk through (Warren & Whang, 1987). Given the consistency seen previously, we expected the scaled solution here to have the same pattern as that observed in Experiments 1 and 2.

Method

The same observers as in Experiment 2 participated, using a combined procedure identical to that of Cutting and Millard (1984). The stimuli and the apparatus were also the same, and the procedure was identical as well except that instead of evaluating dissimilarities, observers were simply asked to select in every pair the one stimulus that depicted the greater exocentric distance.

Analysis and Results

As in Experiment 2, each individual data matrix was checked for order effects. It was then folded into a half matrix containing the frequencies of selection of the *i*th over the *j*th stimulus (resulting in values of 0, 1, or 2). Phi coefficients, a measure of correlation related to χ^2 ($\phi^2 = \chi^2/n$), and χ^2 tests of association were computed for each half matrix when paired with all others. These revealed good consistency among subjects (mean $\phi = .374$, $p < .0001$), with 43 of 45 intersubject tests reliable ($\alpha = .05$). Frequencies were then summed across matrices yielding a single half matrix and then scaled according to Thurstone's Case 5 (Dunn-Rankin, 1983), which yields a one-dimensional solution. The scale values were averaged across those stimuli with equal numbers of sources of information, and the plot of these data is shown in the right panel of Figure 5.

Three results are noteworthy: First, the average discrepancy for the scaled solution was very low (.069), indicating excellent fit of the data. Second, the scaled solution for the 16 stimuli correlates well both with the one-dimensional dissimilarity data and with the direct distance estimates (both $rs = .95$), suggesting that the same combination rule was used in all three psychophysical judgments. Third, the multiple regression of the four sources of information on the scale data yielded another high correlation ($R = .98$), with essentially no variance unaccounted for and with little possibility of contribution by interactive factors. Weights are shown in the right panel of Figure 4. This set of outcomes provides further support to the additivity hypothesis. In addition, no evidence is provided for either selection or multiplication strategies.

Discussion

The results of the three studies taken in consort would appear to confirm that observers were seeing distances in our stimuli. On this basis two objections can be rejected. First, one might claim that the informative content in some of the stimuli was drastically reduced, sometimes imposing an unnatural constraint on the viewer, such as a viewpoint at ground level. As a consequence, the participants of Experiment 1 might have been forced to perform an unnatural task, applying numbers to distances that were indeterminate. However, when performing a more natural task, such as deciding which one of two distances was greater, observers scaled the stimuli the same way. We feel confident therefore that the direct judgments were indeed distance judgments. An analysis of standard deviations supports this conclusion, indicating that a greater amount of information did not imply greater certainty in the judgment.

Second, and more radically, whereas one might acknowledge that the three sets of distance judgments were reliable, one might still claim that they are not informative. When humans move about in the environment, for example, their task is not generally one of making distance judgments; instead, they regulate their motor responses with respect to the layout in space of the objects to be acted upon (Jeannerod, 1983). Thus, the strategies used to perform visuomotor spatial tasks could be completely different from those considered in this article. But, whereas this idea may have some merit, the argument that phenomenal experience of a spatial layout depends on action seems difficult to defend. Even if attaching a number to this experience might be unusual or require training it remains true that to judge distance, one first has to see it. The mechanism involved might be different from the one used in visuomotor tasks, or it might not. Understanding that mechanism will still have important consequences for instances of spatial vision, including but not limited to pictorial depictions of space or quasipictorial spaces in flight simulators or other control displays.

On Independence

Consider the general integration model described in our introduction: $d = f(c[w_1 * s_1, w_2 * s_2])$, where the combination rule c can be selective, additive, or multiplicative. Although our data provide a robust conflation of indicators that c is additive, we are presently unable to provide an equally strong test that the sources did not interact before c was applied. Independent weighting of the sources could be assessed if the response to s_1 and s_2 together is a linear function of s_1 and s_2 in isolation (Massaro, 1987). The fit of such a linear model to our data is reflected in part by the multiple regression coefficients, and is very good. However, correlations, multiple or otherwise, can be misleading indices of linear fit (Birnbaum, 1973). Ashby and Townsend (1986) discussed many other ways that perceptual independence might be ascertained. Their general approach, however, deals with (most simply) two physical dimensions as they map onto two psychological dimensions. Because we are interested here in how two or more physical dimensions map onto one single psychological dimension, the criteria proposed by Ashby and Townsend are of little help here.

Although we do not have strong data to support independence, the assumption that individual mechanisms are independent from one another seems reasonable. It has proved especially useful in computational vision because it allows the computational theorist to concentrate on a specific subsystem (for example, stereopsis) without having to understand the system as a whole. The principle of modular design (Marr, 1981) allows for a system that can be more simply implemented and debugged. But stronger reasons for independence, we believe, come from considering additive integration.

On Additivity

In this article we investigated multiple specification of depth relations by four sources of information and their processing by human observers. A clear pattern emerged. Integration of

depth information from size, height in plane, occlusion, and motion parallax seems best described by the simplest possible additive model:

$$d = w_s * S + w_h * H + w_o * O + w_p * P. \quad (5)$$

But why should one prefer a visual system organized on a principle of additivity? First, additive strategies are generally consistent with parallel processing models of the visual system, and may be easiest to implement in both neural and computer hardware. If the task of natural vision is to integrate multiple information into a coherent representation of the environment, then having several subsystems working in parallel toward a common result is a most efficient way to accomplish this task.

Second, additivity is useful for a system that employs several depth perception mechanisms that emerge at different times during development. Yonas and Granrud (1985) summarized the evidence for infant perception of objects using various sources of information, including some of those that we investigated. Kinematic information, which includes motion parallax, is used by infants earlier than 3 months of age; stereopsis seems next to emerge, between 3.5 and 5 months; and the use of relative size and occlusions comes by 7 months, perhaps even at different times, with size being used by 5.5 months. Given this evidence for a maturational sequence, specialized processing mechanisms may best serve the visual system of the developing infant. Each mechanism may develop independently along its own time course without having to wait for the development of others. The perception of objects in depth can proceed throughout this development, forming percepts by simply adding outputs from processing subsystems as they mature.

Third, additivity is one possible outcome of cooperative-competitive neural mechanisms (Dosher et al., 1986); that is, additivity can be the descriptive counterpart of computational models involving parallel, hierarchical interactions between processing subsystems (Dev, 1975; Grossberg, 1987a, 1987b; Julesz, 1971; Marr & Poggio, 1976; Sperling, 1970). In this vein, Todd (1985) suggested that although there are limitations to the idea of modules within the visual system, there are ways to consider how they might work together. One is to assume the following:

> Objects and events in a natural environment can be multiply specified by many different sources of information, each of which is detected by a specialized processing module with its own individual limitations. In any given situation, we would expect to obtain erroneous outputs from some of these modules because of inappropriate viewing conditions, but it would be most unlikely for two or more of them to fail in exactly the same way. (p. 708)

Different subsystems could be designed to excite one another (in this case, add) when their inputs are compatible, and inhibit one another (subtract) when not. As a result, the system would converge on the correct interpretation of the information.

Nonetheless, given our orientation, there is at least one potential problem. It normally arises in situations of what is called *cue conflict*. That is, whereas we generated stimuli with either no information in one source or considerable information, one could easily do an experiment in which sources are put into conflict, one source indicating one depth order and the other reversing it. Consider the situation shown in Figure 7, and assess its depth. Here a more complex strategy could be required, because occlusion is in conflict with size and height in plane. This might even imply a nonlinear combination.

However, one could still attain additivity by allowing negative weights, so that linear weighting would result in subtraction. If this is the case, three predictions should hold true. One is that perceived depth should appear diminished in cases of source conflict. This means that in the case of Figure 7 the panels should appear to be more like pieces of paper lying flat on top of one another. Furthermore, when present in the image, occlusion should dominate local depth ordering. And finally, in situations of conflict that have occlusion as one source, the additive model may reduce to something quite like the selection model with negative or zero weights. In this manner, a selection strategy could be viewed as a degenerate additive strategy.

Conclusion

It seems that directed perception of distances is carried out by specialized processing subsystems. We propose to call them minimodules, because they are all part of a putative vision module (Dennett, 1984; Fodor, 1983). From our results and those of other researchers, these minimodules may be described as having three properties: independence, ontogenetic sequencing, and additivity. As discussed earlier, such a minimodular organization of the visual system would have many advantages, both developmentally and for efficient and accurate use of visual information. Minimodules would provide the fastest, most efficient way for a visual system to combine converging sources of information. Furthermore, their combined outputs would also attain the most economical description of the stimulus by the most economical process.

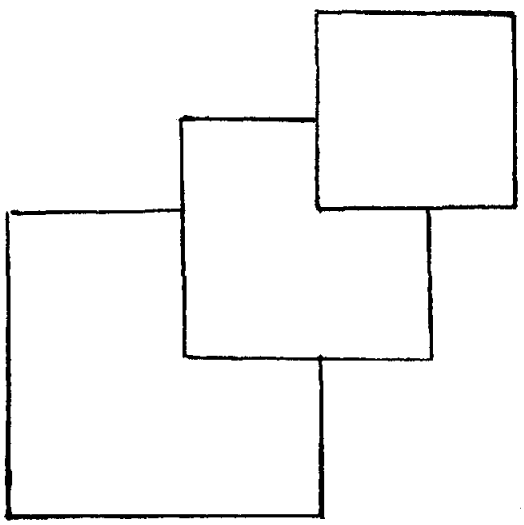

Figure 7. An image with conflicting information, and with occlusion dominating relative size and height in plane.

Such a pattern also follows the notion of operationism, noted by Garner, Hake, and Eriksen (1956), but here the convergence is not on the part of the scientist, but on the part of the visual system. More concretely, as noted by Todd (1985), converging measurements from separate devices are less likely to be mistaken than single measurements. Results from such operations might turn out in many cases to be also the best possible "bet" based on the incoming information. In this fashion, minimodularity is consistent with long-debated concepts such as the minimum principle (Hatfield & Epstein, 1985) or the likelihood constraint that underlies neo-Helmholtzian approaches to perception.

It is important, however, that when assessing visual performance in future research certain properties identified here are kept separated. Additivity, for instance, does not imply independent processing: As evidenced by Massaro (1987), dependent weighting of each source may occur, but the outputs of processing mechanisms could still undergo additive combination. Although this scheme is logically possible, its disadvantage in the perception of layout is that additivity of dependent mechanisms would be much less effective than minimodularity in providing reduction of noise. Conversely, independent weighting may apply, but the combination rule could be multiplicative. Against this latter hypothesis our data provide very sound evidence, but we can be less confident in rejecting the possibility of dependent additivity. As far as the evidence presented here is concerned, independence of additive minimodules is not falsified, but further, stronger tests are desirable.

By the same token, additivity does not imply ontogenetic sequencing. Although independent minimodules would serve best a developing visual system as discussed earlier, additivity could turn out useful also in instances where sequencing does not occur in development. We have presently no basis on which to speculate, but we suspect that the critical point here is that of mapping. Minimodules, in sum, are adaptive when more than one source of information is available. This implies a many-to-one mapping between optical information and distal world and occurs in directed perception (Cutting, 1987). It could be, then, that minimodularity is confined to certain stimulus dimensions or certain information contexts, namely, those where directed perception holds. This is both a theoretical and an empirical question, because it involves both analysis of information as it maps back to distal properties and empirical assessments of its use in the appropriate contexts.

References

Ames, A. (1955). *An interpretive manual for the demonstrations in the Psychology Research Center, Princeton University: The nature of our perception, prehensions and behavior.* Princeton NJ: Princeton University Press.

Anderson, N. H. (1974a). Algebraic models in perception. In E.C. Carterette & P.M. Friedman (Eds.), *Handbook of perception* (Vol. 2, pp. 15–98). New York: Academic Press.

Anderson, N. H. (1974b). Information integration theory: A brief survey. In D.H. Krantz, R.C. Atkinson, R.D. Duce, & D. Suppes (Eds.), *Contemporary developments in mathematical psychology: Measurement, psychophysics, and neural information processing* (pp. 236–305). San Francisco: Freeman.

Anderson, N. H., & Shanteau, J. (1977). Weak inference with linear models. *Psychological Bulletin, 84*, 1155–1170.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93*, 154–179.

Birnbaum, M. H. (1973). The devil rides again: Correlation as an index of fit. *Psychological Bulletin, 93*, 154–179.

Birnbaum, M. H., Wong, R., & Wong, L. K. (1976). Combining information from sources that vary in credibility. *Perception & Psychophysics, 4*, 330–336.

Braunstein, M. L., Andersen, G. J., Rouse, M. W., & Tittle, J. S. (1986). Recovering viewer-centered depth from disparity, occlusion, and velocity gradients. *Perception & Psychophysics, 40*, 216–224.

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

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

Cutting, J. E. (1987). Perception and information. *Annual Review of Psychology, 38*, 61–90.

Cutting, J. E., & Millard, R. T. (1984). Three gradients and the perception of flat and curved surfaces. *Journal of Experimental Psychology: General, 113*, 198–216.

De Leeuw, J., & Stoop, I. (1984). Upper bounds for Kruskal's stress. *Psychometrica, 49*, 391–402.

Dennett, D. C. (1984). Carving the mind at its joints. *Contemporary Psychology, 29*, 285–286.

Dev, P. (1975). Perception of depth surfaces in random dot stereograms: A neural model. *International Journal of Man-Machine Studies, 7*, 511–58.

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.

Dunn-Rankin, P. (1983). *Scaling methods.* Hillsdale, NJ: Erlbaum.

Epstein, W. (1977). Historical introduction to the constancies. In W. Epstein (Ed.) *Stability and constancy in visual perception* (pp. 1–22). New York: Wiley.

Farber, J. M., & McConkie, A. B. (1979). Optical motions as information for unsigned depth. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 494–500.

Fodor, J. (1983). *Modularity of mind.* Cambridge, MA: MIT Press.

Garner, W. R. (1973). Attention: the processing of multiple sources of information. In E.C. Carterette & P.M. Friedman (Eds.), *Handbook of perception* (Vol. 2, pp. 23–59). New York: Academic Press.

Garner, W. R., Hake, H. W., & Eriksen, C. W. (1956). Operationism and the concept of perception. *Psychological Review, 63*, 149–159.

Gibson, J. J. (1950). *The perception of the visual world.* Boston: Houghton-Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

Gilinsky, A. S. (1951). The effect of attitude upon the perception of size. *American Journal of Psychology, 68*, 173–192.

Grossberg, S. (1987a). Cortical dynamics of three dimensional form, color, and brightness perception: I. Monocular theory. *Perception & Psychophysics, 41*, 87–116.

Grossberg, S. (1987b). Cortical dynamics of three dimensional form, color, and brightness perception: II. Binocular theory. *Perception & Psychophysics, 41*, 117–158.

Harker, G. S. (1958). Interrelation of monocular and binocular acuities in the making of equidistance judgment. *Journal of the Optical Society of America, 48*, 233–240.

Hatfield, G., & Epstein, W. (1985). The status of the minimum principle in the theoretical analysis of perception. *Psychological Bulletin, 97*, 155–186.

Holway, A. F., & Boring, E. G. (1941). Determinants of apparent

visual size with distance variant. *American Journal of Psychology, 54*, 21–37.

Jameson, D., & Hurvich, L. M. (1959). Note on the factors influencing the relation between stereoscopic acuity and observation distance. *Journal of the Optical Society of America, 49*, 639.

Jeannerod, M. (1983). How do we direct our actions in space? In A. Hein & M. Jeannerod (Eds.), *Spatially oriented behavior* (pp. 1–13). New York: Springer Verlag.

Julesz, B. (1971). *Foundations of cyclopean perception*. Chicago: University of Chicago Press.

Kaufman, L. (1974). *Sight and mind*. New York: Oxford University Press.

Klahr, D. A. (1969). Monte Carlo investigations of the statistical significance of Kruskal's scaling procedure. *Psychometrica, 34*, 319–330.

Knudsen, E. I., & Konishi, M. (1979). Mechanisms of sound localization in the Barn Owl (*Tyto Alba*). *Journal of Comparative Physiology, 133*, 13–21.

Koffka, K. (1935). *Principles of gestalt psychology*. New York: Harcourt.

Levine, D. M. (1978). A Monte Carlo study of Kruskal's variance-based measure of stress. *Psychometrica, 43*, 307–315.

Marr, D. (1981). *Vision*. San Francisco: Freeman.

Marr, D., & Poggio, T. (1976). Cooperative computation of stereo disparity. *Science, 194*, 283–287.

Massaro, D. W. (1987). *Speech perception by ear and eye*. Hillsdale, NJ: Erlbaum.

Massaro, D. W., & Cohen, M. M. (1983). Evaluation and integration of visual and auditory information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 753–771.

Nakayama, K. (1983). Motion parallax sensitivity and space perception. In A. Hein & M. Jeannerod (Eds.), *Spatially oriented behavior* (pp. 223–242). New York: Springer Verlag.

Ono, M. E., Rivest, J., & Ono, H. (1986). Depth perception as a function of motion parallax and absolute distance information. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 331–337.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17*, 776–783.

Pastore, N. (1971). *Selective history of theories of visual perception: 1650–1950*. New York: Oxford University Press.

Regan, D. M., & Beverley, K. I. (1982). How do we avoid confounding the direction we are looking and the direction we are moving? *Science, 215*, 194–196.

Shepard, R. N. (1974). Representation of structure in similarity data: Problems and prospects. *Psychometrica, 39*, 373–421.

Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science, 210*, 390–398.

Sperling, G. (1970). Binocular vision: A physical and a neural theory. *American Journal of Psychology, 83*, 461–534.

Todd, J. T. (1985). Perception of structure from motion: Is projective correspondence of moving elements a necessary condition? *Journal of Experimental Psychology: Human Perception and Performance, 11*, 689–710.

Todd, J. T., & Akerstrom, R. A. (1987). Perception of three-dimensional form from patterns of optical texture. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 242–255.

van der Meer, H. C. (1979). Interrelation of the effects of binocular disparity and perspective cues on judgments of depth and height. *Perception & Psychophysics, 26*, 481–488.

Wagenaar, W. A., & Padmos, P. (1971). Quantitative interpretation of stress in Kruskal's MDS technique. *British Journal of Mathematical and Statistical Psychology, 24*, 101–110.

Wallach, H. (1976). *On perception*. New York: Quadrangle/New York Times.

Warren, W. H., & Whang, S. (1987). Visual guidance of walking through apertures: Body-scaled information for affordances. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 371–383.

Yonas, A., & Granrud, C. E. (1985). The development of sensitivity to kinetic, binocular, and pictorial depth information in human infants. In D. J. Ingle, M. Jeannerod, & D. N. Lee (Eds.), *Brain mechanisms and spatial vision* (pp. 113–145). Dordrecht, The Netherlands: Martinus Nijhoff.

Received August 14, 1987
Revision received December 1, 1987
Accepted December 2, 1987 ■