

## Classification and perceived similarity of compound gratings that differ in relative spatial phase

MICHAEL J. KAHANA and PATRICK J. BENNETT  
*University of Toronto, Toronto, Ontario, Canada*

Discrimination studies suggest that two, and only two, channels encode relative spatial phase shifts in compound gratings (Bennett & Banks, 1991; Field & Nachmias, 1984). The more sensitive channel consists of even-symmetric filters and responds best to cosine phase shifts (e.g.,  $0^\circ$ - $180^\circ$ ); the other consists of odd-symmetric filters and responds best to sine phase shifts (e.g.,  $90^\circ$ - $270^\circ$ ). The present experiments investigated whether the two-channel model generalizes to suprathreshold perceptual tasks. Experiment 1 examined classification learning of compound gratings, consisting of a fundamental ( $f$ ) and second harmonic ( $2f$ ), that differed in  $2f$  contrast and relative phase. Experiments 2 and 3 measured the perceived similarity of  $f+2f$  gratings. The results of Experiment 1 were broadly consistent with the predictions of the two-channel model. Specifically, the classification data were best explained by assuming that classification was based on the responses of differentially sensitive even- and odd-symmetric filters. In Experiments 2 and 3, two-dimensional multidimensional scaling solutions provided a good account for the similarity judgments. In Experiment 2, Dimension 1 was strongly correlated with cosine phase, and Dimension 2 was moderately correlated with sine phase. In Experiment 3, cosine phase was again strongly related to Dimension 1, whereas the absolute value of sine phase was strongly related to Dimension 2. Overall, these results suggest that the two-channel model of phase discrimination provides a useful framework for interpreting classification and similarity judgments of compound gratings.

A basic problem in understanding human vision is to determine the information necessary to discriminate, identify, and categorize naturalistic images. There is considerable evidence that the visual system initially encodes pattern information by performing spatial frequency analyses on local patches of the retinal image (Wilson, Levi, Maffei, Rovamo, & DeValois, 1990). Such frequency analyses must represent in some manner the contrast and phase (defined in the following paragraph) of the local frequency components because *both* are important for pattern discrimination and recognition. For example, the demonstrations of Oppenheim and Lim (1981) and Pitrowski and Campbell (1982) suggest that much of the information for discrimination and identification of com-

plex patterns is contained in their global phase spectra: Image identification is severely disrupted by scrambling the phase spectrum, but remains virtually unchanged after scrambling the amplitude spectrum. Morgan, Ross, and Hayes (1991) performed similar analyses at different spatial scales within an image and showed that phase also carries a significant amount of recognition information about medium-sized image patches, although the amplitude spectrum carries most of the information at the smallest scales. Consistent with the results of these computational studies, psychophysical evidence also suggests that the representation of spatial phase is crucial for pattern discrimination and recognition. For example, it is well known that observers often fail to discriminate easily detectable patterns presented in the peripheral visual field, and these discrimination failures are thought to be caused by phase-encoding deficits that are also found in the periphery (e.g., Bennett & Banks, 1991). Phase-encoding anomalies may also underlie discrimination deficits commonly observed in the amblyopic fovea (Bennett, 1989) and in young human infants (Bennett, 1989; Braddick, Atkinson, & Wattam-Beil, 1986).

Figure 1 illustrates the construction of compound gratings that differ in spatial phase. The bottom lines in each panel represent the luminance profiles of two sine-wave gratings with frequencies that differ by a factor of 2, and the top line depicts the profile of the sum of the two sine

Portions of this work were presented at the 1992 meeting of the Association for Research in Vision and Ophthalmology, Sarasota, FL. We are grateful to Ingo Rentschler for providing us with the data from Caelli, Rentschler, and Scheidter (1987) and to J. O. Ramsey for providing us with his Multiscale software. Thanks are also due Allison B. Sekuler for comments on an earlier draft of this manuscript and Ian Spence for helpful discussions on multidimensional scaling. This research was supported by National Science and Engineering Research Council of Canada Grant OG0042133 to P.J.B. Correspondence concerning this article should be addressed to M. J. Kahana, Department of Psychology, Harvard University, Cambridge, MA 02138.

—Accepted by previous editor, Charles W. Eriksen

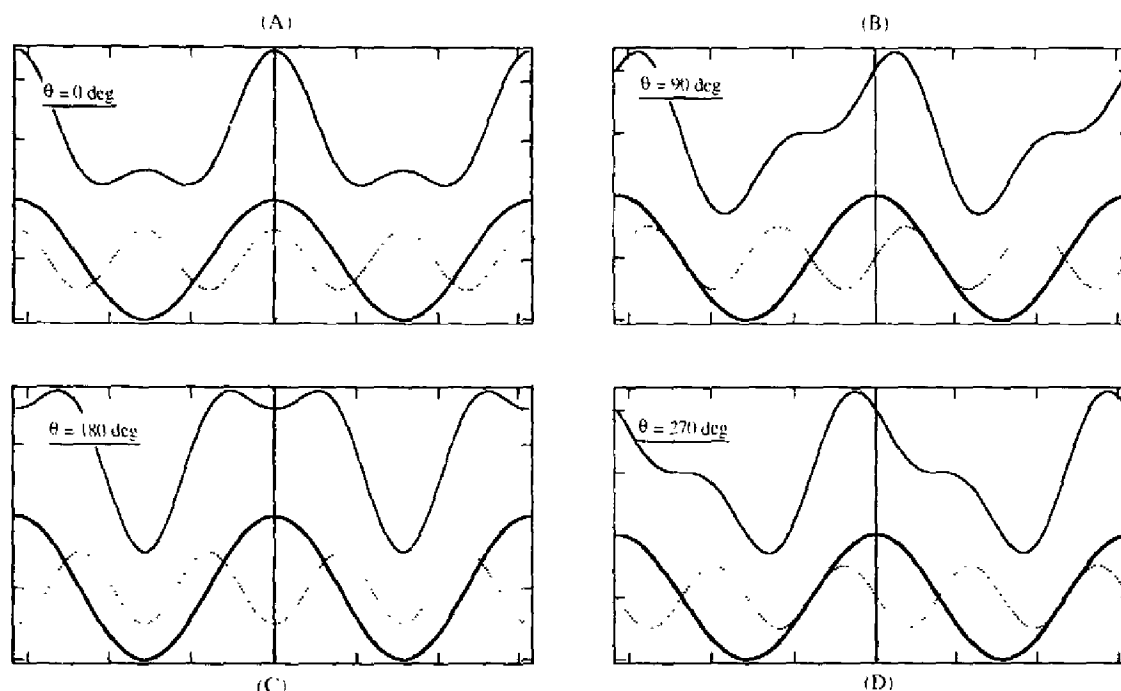


Figure 1. An illustration of the effects of relative phase shifts. In each panel, position and luminance are plotted in arbitrary units on the horizontal and vertical axes, respectively. The origin in each panel is indicated by the vertical line. The bottom curves in each panel show the luminance profiles of two sine-wave gratings with frequencies that differ by a factor of two. The phase of the higher frequency varies across panels and is indicated in the legends. The top line depicts the sum of the two sine waves and has been shifted vertically for clarity.

waves. *Absolute spatial phase* refers to the position of a grating relative to some fixed origin. *Relative spatial phase* refers to the positions of grating components relative to each other rather than to some fixed position in space. In the case of compound gratings composed of a fundamental ( $f$ ) and second harmonic ( $2f$ ), relative spatial phase refers to the position of  $2f$  relative to the peak of  $f$ . By convention, a sinusoidal grating whose peak is lined up with the peak of  $f$  is said to be in  $0^\circ$  phase (Figure 1A); a grating whose trough is lined up with the peak of  $f$  is said to be in  $180^\circ$  phase (Figure 1C). Gratings whose zero-crossings are lined up with the peak of  $f$  are said to be in either  $90^\circ$  or  $270^\circ$  phase (Figures 1B and 1D). Note that a  $180^\circ$  phase shift of an individual frequency component corresponds to a contrast reversal of that component.

Figure 1 also illustrates that relative phase shifts produce significant changes in the luminance profiles of compound waveforms even when they are composed of identical frequency components. For the waveforms shown in the figure, a cosine phase shift (i.e.,  $0^\circ$ – $180^\circ$ ) results in a change in the polarity of the most prominent bars in the grating. A sine phase shift (i.e.,  $90^\circ$ – $270^\circ$ ) alters the direction of the steepest luminance gradient. It is important to note, however, that both cosine and sine phase (or their correlates) must be represented in order to discrim-

inate and recognize most natural patterns, not just two-component gratings (see Oppenheim & Lim, 1981, for details).

One experimental approach that has yielded some insight into the way phase is encoded is to measure discrimination thresholds for  $180^\circ$  relative phase shifts in compound gratings composed of a fundamental ( $f$ ) and second harmonic ( $2f$ ) (Bennett & Banks, 1987, 1991; Field & Nachmias, 1984). In this procedure, thresholds are measured by fixing the contrast of  $f$  at some suprathreshold value and adjusting the contrast of  $2f$  until the  $180^\circ$  shift is discriminable. The advantage of using  $f+2f$  gratings is that the overall contrast of these waveforms is not altered by a  $180^\circ$  relative phase shift and therefore cannot be used as a discrimination cue.

Thresholds for various  $180^\circ$  shifts (e.g.,  $0^\circ$ – $180^\circ$ ,  $45^\circ$ – $225^\circ$ ,  $90^\circ$ – $270^\circ$ ) of  $2f$  are typically plotted as in Figure 2. Each point in Figure 2 represents the discrimination threshold for a particular  $180^\circ$  shift. The contrast of  $2f$  at discrimination threshold is represented by the distance from the point to the origin. The phase shift (e.g.,  $0^\circ$ – $180^\circ$ ,  $45^\circ$ – $225^\circ$ , etc.) is represented by the angle between the point and the abscissa. Thus, threshold for a  $0^\circ$ – $180^\circ$  shift is represented by a point on the abscissa, threshold for a  $90^\circ$ – $270^\circ$  shift is represented by a point on the ordinate, and thresholds for other shifts lie in be-

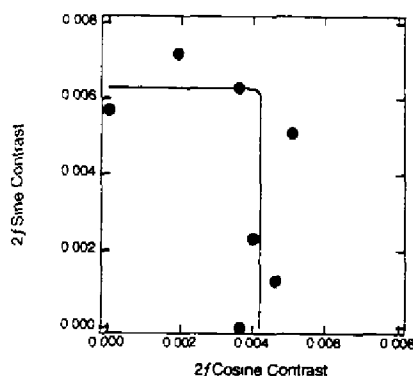


Figure 2. Examples of  $180^\circ$  relative phase discrimination thresholds measured with  $f+2f$  gratings. Each point represents the  $2f$  contrast needed to discriminate a  $180^\circ$  shift of  $2f$  phase (e.g.,  $0^\circ$ – $180^\circ$ ,  $45^\circ$ – $225^\circ$ ,  $90^\circ$ – $270^\circ$ , etc.). The  $2f$  contrast is represented as the length of a vector joining each point to the origin, and the phase is represented as the vector's angle to the horizontal axis. The rectangular coordinates of each point are the cosine and sine components of  $2f$  at discrimination threshold. In this plot, all points on a vertical contour have equal cosine values and all points on a horizontal contour have equal sine values. Notice that the thresholds fall along equal-cosine (vertical) and equal-sine (horizontal) contours. The solid lines are the predictions of the two-channel model. Data are from Bennett and Banks (1987).

tween. The Cartesian coordinates of the figure represent the sine and cosine components of  $2f$  at threshold.<sup>1</sup> Most experiments find that discrimination thresholds for different  $180^\circ$  relative phase shifts fall along equal-cosine and equal-sine contours (Bennett, 1993; Bennett & Banks, 1991; Burr, Morrone, & Spinelli, 1989; Field & Nachmias, 1984), thus indicating that  $180^\circ$  shifts are discriminated when the change in either the cosine or the sine component of the shift exceeds some criterion. This pattern of results holds across a wide range of fundamental frequency and contrast (Bennett & Banks, 1991) and for different harmonic frequencies (Bennett, 1993). One interpretation of this result is that relative phase is encoded by two spatial filters: Odd symmetric filters that are optimally sensitive to sine (i.e.,  $90^\circ$ – $270^\circ$ ) phase shifts, and even symmetric filters that are most sensitive to cosine (i.e.,  $0^\circ$ – $180^\circ$ ) shifts (Bennett & Banks, 1987; Burr et al., 1989; Field & Nachmias, 1984).

Of course, discriminating very subtle differences in relative phase is only one small aspect of phase perception. It is important, therefore, to determine whether the two-channel model provides a framework for understanding how phase is encoded in more quotidian conditions. Although the two-channel model provides a reasonable account of phase discrimination in compound gratings, it is by no means certain that it can be applied successfully to suprathreshold perceptual tasks. As others have pointed out (e.g., Badcock, 1984), relative phase shifts alter many spatial features. A phase shift in a two-component compound grating, for instance, can alter peak-to-trough am-

plitude, luminance maxima and minima, luminance gradients, and the widths of individual bars. Bennett (1993) has argued that changes in these and other features do not play a significant role in threshold tasks, but they may greatly influence classification or perceived similarity among highly discriminable patterns. In the present experiments, we examined the extent to which the two-channel model of phase discrimination can account for perceptual classification and similarity judgments.

We examined whether three notable features of the two-channel model of phase discrimination generalize to suprathreshold classification learning and similarity judgments. First, we tested whether two mechanisms—one that encodes cosine amplitude and another that encodes sine amplitude—could account for classification and similarity judgments. Second, we examined whether the greater sensitivity of the cosine-sensitive mechanism (e.g., Bennett & Banks, 1987, 1991; Rentschler & Treutwein, 1985) generalizes to the classification and similarity judgment tasks. Third, we examined whether the independence of these two mechanisms found in discrimination studies generalizes to the classification and similarity judgment tasks. In particular, does classification performance conform to a supremum metric (i.e., whichever mechanism gives the largest response is used by the system in making the classification or discrimination judgments; see Figure 2)? Classification and perceived similarity judgments provide complementary paradigms to test the generality of the discrimination findings.

## EXPERIMENT 1

Classification tasks provide an elegant link between discrimination and identification/recognition paradigms. While discrimination tasks generally involve stimuli that are presented near threshold and in close temporal proximity, classification tasks involve suprathreshold judgments. Classification tasks also rely more heavily upon cognitive processes involving memory and decision.

There have been numerous models proposed to explain classification performance. Prominent among these are *prototype*, *exemplar*, and *decision bound* models (Ashby & Maddox, 1993). Each type of model assumes that a stimulus is represented by a point, or distribution of points, in a multidimensional feature space. These models differ from one another primarily in the manner in which the relevant information is retrieved (see Ashby & Maddox, 1993). Prototype models (e.g., Nosofsky, 1986; Reed, 1972) assume that a category is represented by the average position (in space) of its constituent stimuli. A probe stimulus is then classified in the category whose prototype is closest in feature space. Exemplar models (e.g., Nosofsky, 1986) propose that categories are not represented by a single point in feature space, but rather are derived from the positions of their constituent members. According to these models, the proximity of a target item to all of the stored items is determined, and the item is classified into the category whose items are most similar to the probe. Finally, decision bound models (e.g., Ashby

& Perrin, 1988) assume that each stimulus is represented by a distribution of points in feature space. These distributions, usually assumed to be multivariate normal, reflect perceptual noise in the system. A region in feature space is associated with a particular categorical response. Decision bounds dividing response regions are determined so as to maximize classification performance.

### Models of Classification

Three models will be compared: Nosofsky's (1986, 1992) exemplar-based generalized context model, a probabilistic prototype model (Nosofsky, 1987; Reed, 1972), and Ashby and Perrin's (1988) generalized recognition theory—a decision bound model.

**Generalized context model (GCM).** The exemplar-based GCM (Nosofsky, 1986) is an extension of Medin and Schaffer's (1978) context model of classification. Following Luce's (1963) similarity choice rule, the GCM assumes that the probability of a stimulus  $i$  being assigned to a category  $J$  is given by

$$P(R_J|S_i) = \frac{\beta_J \sum_{j \in C_J} \eta_{ij}}{\sum_K \beta_K \sum_{k \in C_K} \eta_{ik}}, \quad (1)$$

where  $0 \leq \beta_J \leq 1$  and  $\sum_J \beta_J = 1$ . Here,  $\beta_J$  represents the bias toward category  $J$ , and  $\eta_{ij}$  represents the similarity of exemplars  $i$  and  $j$  (see Nosofsky, 1986). Similarity is defined to be a monotonic transformation of distance in feature space. If one allows for differential weighting of the dimensions in feature space, the formula for the distance between two exemplars  $i$  and  $j$  that vary along  $n$  dimensions is given by

$$d_{ij} = c \left( \sum_{k=1}^n w_k |x_{ik} - x_{jk}|^r \right)^{1/r}, \quad (2)$$

where each dimension ( $k$ ) is weighted by a factor  $w_k$  such that  $0 \leq w_k \leq 1$  and  $\sum_k w_k = 1$ . A scaling parameter,  $c$ , adjusts the overall distance scale. Summation between dimensions is governed by the exponent of the Minkowski metric,  $r$ . For example, if  $r = 2$ , and the dimensions are equally weighted, then the distance corresponds to a simple Euclidean metric. Other popular metrics are the city-block metric ( $r = 1$ ) and the supremum metric ( $r = \infty$ ). The similarity between two exemplars,  $\eta_{ij}$ , is often related to distance,  $d_{ij}$ , by the power function (see Nosofsky, 1992)

$$\eta_{ij} = e^{-d_{ij}^p}. \quad (3)$$

When the similarity exponent ( $p$ ) equals 1, Equation 3 is Shepard's (1987) exponential decay model of generalization.

**Probabilistic prototype model (PPM).** The PPM (Nosofsky, 1987; Reed, 1972) is essentially a distance-based model of classification. Unlike exemplar models, prototype models assume that a category is represented by a single point in feature space representing the average value of all of the category exemplars on each of the dimensions (e.g., Reed, 1972). As such, prototype models

estimate response probabilities using the distance between each stimulus and the prototype of the response category. The PPM, like the GCM, uses Equation 1 to predict category judgments, but the summation over exemplars is replaced by the similarity of the probe stimulus ( $i$ ) to the category prototype ( $J$ ). The probability of a stimulus  $i$  being assigned to a category  $J$  is then given by

$$P(R_J|S_i) = \frac{\beta_J \eta_{iJ}}{\sum_K \beta_K \eta_{iK}},$$

where  $0 \leq \beta_J \leq 1$  and  $\sum_J \beta_J = 1$ . As with the exemplar model,  $\beta_J$  represents the bias toward category  $J$  and  $\eta_{iJ}$  represents the similarity of exemplar  $i$  to the prototype of category  $J$ . Similarity is defined exactly as in the GCM (see Equations 2 and 3). Although this model is similar to the GCM, it is still possible to distinguish between them. Nosofsky (1992), in a detailed comparison of exemplar and prototype models, concluded that across a wide range of tasks, the exemplar-based GCM provided a better fit to the data than did the PPM.

**Generalized recognition theory (GRT).** GRT is the multidimensional extension of signal detection theory developed by Ashby and his colleagues (e.g., Ashby & Gott, 1988; Ashby & Perrin, 1988). In GRT, stimuli are represented by distributions of points in feature space and categories are associated with regions within feature space. Decision bounds separating these regions determine the probability with which a stimulus is classified in a given category. Although GRT is an extremely general model that can be shown to subsume some distance-based models as special cases, we will be working with a simple version of GRT that contains a relatively small number of free parameters.

In GRT, a physical stimulus may give rise to different representations on different trials due to encoding variability. This variability is typically assumed to be multivariate normal. A probe stimulus is assigned to a given category if it falls within a region in feature space associated with that category. Classification errors may result from the representational variability of the stimuli.

In contrast to the exemplar and prototype models, GRT assumes a deterministic process of response selection. Decision bounds—estimated from the data—separate the regions of perceptual space associated with each response, and a stimulus is always classified according to the region in feature space in which it falls. In this paper, we adopt two simplifying assumptions that typically are used when working with GRT—namely, that the decision bounds are linear and stable across trials.

### Classification of Compound Gratings That Differ in Relative Phase

The two-channel model predicts that a classification model utilizing a two-dimensional space corresponding to the cosine and sine components of the phase-shifted harmonic should provide a reasonably good account of the data. Also, discrimination studies show that observers are generally less sensitive at discriminating 90°–270° rel-

ative phase shifts than they are at discriminating  $0^\circ$ - $180^\circ$  relative phase shifts. Therefore, the two-channel model predicts that subjects should make more errors along the sine dimension than along the cosine dimension. Because classification models can account for these differences by differentially weighting the dimensions (in the GCM and the PPM) or by assuming differential variability along dimensions (in GRT), the classification models should assign a lower weight or greater variability to the sine dimension than to the cosine dimension. Finally, the discrimination data suggest that there is little summation between the two phase-encoding channels and therefore predict that the exponent of the Minkowski metric should be significantly greater than 2. Contrary to these predictions, Caelli, Rentschler, and Scheidter (1987) proposed a nonweighted Euclidean distance model to account for the classification of  $f+3f$  gratings differing in relative phase. The following experiment examines whether the two-channel model of phase discrimination generalizes to classification of  $f+2f$  compound gratings.

### Method

**Subjects.** Sixteen subjects, all introductory psychology students at the University of Toronto, participated in the experiment for optional course credit.

**Stimuli.** Stimuli consisted of a set of 20  $f+2f$  gratings segregated into four classes. These classes were defined by the relative phase

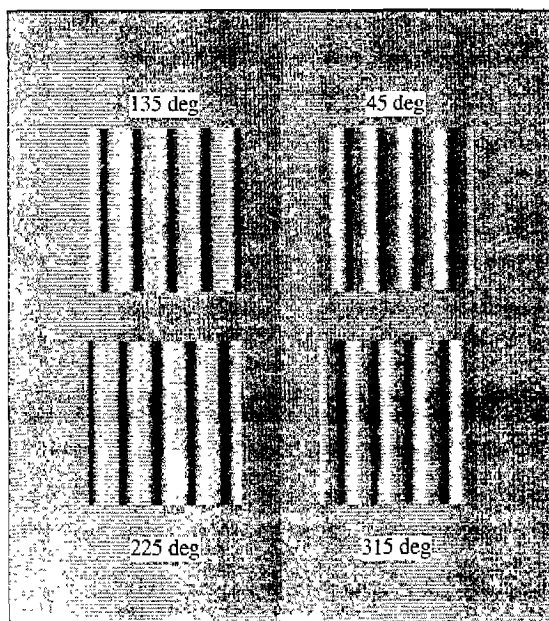


Figure 3. Examples of the prototypical patterns used in Experiments 1 and 2. The  $2f$  phase is indicated next to each compound grating. Gratings on top of one another (e.g.,  $135^\circ$  and  $225^\circ$ ,  $45^\circ$  and  $315^\circ$ ) differ only along the sine dimension, whereas gratings that are next to one another (e.g.,  $135^\circ$  and  $45^\circ$ ,  $225^\circ$  and  $315^\circ$ ) differ only along the cosine dimension. In the actual experiments, the patterns were displayed in a circular aperture that showed approximately four cycles of the fundamental. The reproduction process has altered the contrast of the patterns.

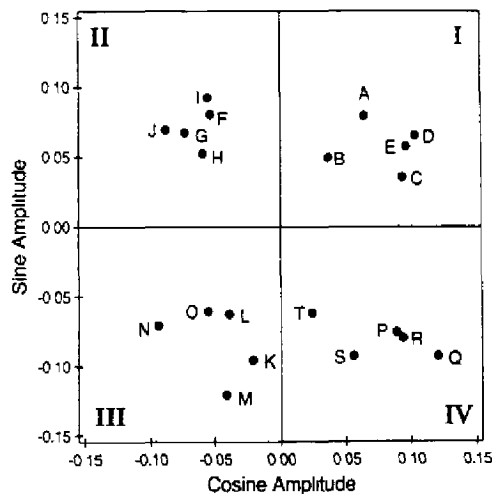


Figure 4. The stimuli used in Experiments 1 and 2. Each point represents the cosine and sine components of  $2f$  in an  $f+2f$  grating. The contrast and phase of  $2f$  are represented as the length of the vector joining each point to the origin and the angle of the vector, respectively. The four classes of stimuli in Experiment 1 correspond to the four quadrants in the figure. Finally, the cosine and sine values of all the stimuli are listed in Table 1.

and contrast of the  $2f$  component. Luminance profiles for these stimuli are given by the following equation:

$$L(x,y) = L_{avg} \{1 + A_1 \cos(2\pi fx) + A_2 \cos(2\pi 2fx - \phi)\},$$

where  $f$  is spatial frequency in cycles per degree,  $A_1$  and  $A_2$  are Michelson contrasts, and  $\phi$  is relative phase. Fundamental frequency ( $f$ ) and contrast ( $A_1$ ) were 1.75 cycles per degree (cpd) and 0.2, respectively. The mean contrast ( $A_2$ ) for the  $2f$  component was 0.1. Average luminance ( $L_{avg}$ ) was 14 cd/m<sup>2</sup>. The four classes of stimuli were centered around phases ( $\phi$ ) of  $45^\circ$ ,  $135^\circ$ ,  $225^\circ$ , and  $315^\circ$ . Figure 3 illustrates the prototype of each category. The five exemplars within each class were generated by adding Gaussian noise to the prototypical phase angle and contrast. The sine and cosine amplitudes of the  $2f$  component of the stimuli are shown in Figure 4.

Stimuli were presented on an Apple color monitor. Display size was  $640 \times 480$  pixels (72 pixels per inch). The frame rate was 67 Hz, noninterlaced. Luminance was linearized with lookup tables, and the entire 8-bit range produced a maximum Michelson contrast of 0.5. Stimuli were presented individually within a circular aperture that was centered within the display and subtended  $6.8^\circ$ . Both the stimulus and the uniform background surrounding it had the same average luminance and chromaticity.

**Procedure.** Each subject was given 10 study-test blocks to learn the classification scheme for the stimuli in Figure 4. In both the study and the test phase within a block, the subjects were shown each of the 20 gratings for 1 sec. During the study phase, a number indicating class membership was shown prior to each grating. To facilitate learning, the subjects were instructed to type the number after viewing each stimulus. During the test phase, the gratings were shown without their class numbers, and the subjects pressed a key indicating the stimulus class. After each response, a high-frequency tone indicated a correct classification, whereas a low-frequency tone indicated an incorrect classification. Feedback was also provided during the study trials to ensure that the subjects were pressing the right buttons.

Table 1  
Classification Data From Experiment 1

Stimulus Category	Stimulus Code	Probability of Classifying Stimulus in Category				Stimulus Coordinates	
		Category 1	Category 2	Category 3	Category 4	x	y
1	A	0.556	0.106	0.144	0.194	0.065	0.079
1	B	0.600	0.131	0.113	0.156	0.037	0.049
1	C	0.556	0.144	0.038	0.263	0.094	0.035
1	D	0.575	0.050	0.113	0.263	0.104	0.065
1	E	0.594	0.113	0.056	0.238	0.097	0.057
2	F	0.088	0.675	0.169	0.069	-0.053	0.080
2	G	0.088	0.706	0.175	0.031	-0.073	0.067
2	H	0.169	0.619	0.150	0.063	-0.059	0.052
2	I	0.113	0.656	0.169	0.063	-0.055	0.092
2	J	0.150	0.613	0.163	0.075	-0.088	0.069
3	K	0.119	0.156	0.563	0.163	-0.021	-0.096
3	L	0.088	0.306	0.556	0.050	-0.039	-0.063
3	M	0.119	0.213	0.569	0.100	-0.041	-0.121
3	N	0.063	0.313	0.569	0.056	-0.094	-0.071
3	O	0.088	0.238	0.575	0.100	-0.055	-0.061
4	P	0.238	0.069	0.106	0.588	0.089	-0.076
4	Q	0.244	0.038	0.150	0.569	0.121	-0.093
4	R	0.213	0.094	0.069	0.625	0.094	-0.080
4	S	0.181	0.081	0.188	0.550	0.056	-0.093
4	T	0.181	0.113	0.225	0.481	0.024	-0.062

Note.—Each value represents the proportion of trials that a given stimulus was assigned to a given class. The data were averaged across all trials and subjects.

All testing was done with the room lights off. Viewing was binocular through natural pupils. Viewing distance was 1 m. Head position was stabilized with a chin/forehead rest.

### Results and Discussion

Table 1 lists the probability of a given stimulus being classified in each of the four classes. Because stimuli within a class are clustered close to each other and are very far from the exemplars of other categories (see Figure 4), most of the variance in classification performance is between-class variance. Therefore, a correlation between distance and classification accuracy will be driven by the mean difference between correct and incorrect classifications. In other words, the correlation just tells us that subjects are capable of doing the task. To properly evaluate various classification models, it is necessary to measure the goodness of fit, rather than the correlation, between the model's predictions and the classification probabilities for each stimulus. We have done this in two ways: For data pooled across subjects, we minimized the chi-square statistic, whereas in fitting the models to data from individual subjects, we minimized the root-mean-square (RMS) difference scores. Chi-square tests allow for the comparison of two models when one model is a generalization of the other. In contrast, by obtaining RMS scores for each subject, one can make direct statistical comparisons between completely different models. In addition, fitting the models to data from individual subjects takes into account the possibility that subjects differ on the model parameters.

### Classification Models

Three different classification models are compared: the exemplar-based GCM, the PPM, and GRT.

**Generalized context model (GCM).** The GCM provides a reasonably good fit to the data from Experiment 1. Figure 5A shows the predicted and observed classification probabilities for all response types to each of the stimuli. The predicted values here are based on the data averaged across subjects. A chi-square test showed that the deviations from the predictions were just statistically significant ( $\chi^2 = 92.20$ ,  $df = 72$ ,  $p = .05$ ), suggesting that GCM does not account for all of the observed variance. In fitting GCM separately to the data from individual subjects, the average RMS value was 0.11. Considering the high levels of variability in individual subject data (each response probability represents only 10 responses), the RMS values obtained were quite good.

Best-fitting parameter values for the averaged data are shown in Table 2, and mean parameter values for the model fits to individual subject data are shown in Table 3. Some constraints were placed on the parameter values in fitting the data: The sine and cosine weights could range from 0.01 to 0.99, the value of the Minkowski metric

Table 2  
Parameter Estimates for the Generalized Context Model (GCM) and the Probabilistic Prototype Model (PPM) Applied to Averaged Classification Data From Experiment 1

Parameters	GCM	PPM
Minkowski exponent ( $r$ )	2.0	1.9
Similarity exponent ( $p$ )	1.0	1.0
Scaling parameter ( $c$ )	13.2	12.4
Bias $\beta_1$	0.25	0.26
Bias $\beta_2$	0.27	0.26
Bias $\beta_3$	0.25	0.24
Bias $\beta_4$	0.23	0.24
Cos/Sin Ratio	3.3	2.7

could range from 1 to 20, and the similarity exponent could range from 1 to 4. These constraints were employed to keep the model within bounds so that a solution could be obtained. Function minimization was accomplished using a downhill simplex method (Nelder & Mead, 1965). A tolerance of 0.01 on RMS values and 0.1 on chi-square values was used to terminate the simplex routine.

The cosine dimension was consistently weighted more heavily than the sine dimension. For the averaged data, the ratio of the cosine to sine weights was 3.3. For the individual subject data, the ratio of the mean cosine to sine weights was 2.6, with all but 2 of the 16 subjects weighting the cosine dimension more heavily than the sine dimension. These results indicate that the subjects were more sensitive to changes in cosine amplitude. They also show that this aspect of the discrimination data generalizes to classification tasks. The best-fitting Minkowski metric was nearly Euclidean ( $r = 2.0$ , for averaged data;  $r = 2.52$ , for fits to individual subjects), suggesting that, unlike the discrimination data, summation of sine and cosine information was taking place. In all of the fits, the response bias parameters were nearly equal, indicating that they were not critical in fitting the data.

**Probabilistic prototype model (PPM).** The PPM also provides a reasonably good fit to the data. Figure 5B shows the correspondence between the observed and predicted classification probabilities for the PPM. Parameter values for the averaged data are shown in Table 2, and the mean parameter values for the fits to individual subjects data are shown in Table 3. Overall, these results are quite similar to those obtained by the GCM. The cosine dimension was weighted between two and three times as heavily as the sine dimension, the best-fitting metric was nearly Euclidean, and the bias parameters were nearly equal. On the basis of the chi-square tests, the PPM fit the data marginally better than did the GCM; however, the mean RMS values were identical for the two models. Again, the deviations from the model predictions were quite small, though nearly significant ( $\chi^2 = 88.6$ ,  $df = 72$ ,  $p = .09$ ). Overall, both the PPM and the GCM provided reasonably good fits to the classification data.

Table 3  
Means and Standard Errors of Parameter Estimates for the Generalized Context Model (GCM) and the Probabilistic Prototype Model (PPM) Applied to Individual Subject Classification Data From Experiment 1

Parameters	GCM		PPM	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Minkowski exponent ( <i>r</i> )	2.52	0.74	2.39	0.57
Similarity exponent ( <i>p</i> )	1.44	0.25	1.57	0.26
Scaling parameter ( <i>c</i> )	15.06	2.41	13.44	1.92
Bias $\beta_1$	0.24	0.02	0.26	0.02
Bias $\beta_2$	0.27	0.02	0.26	0.02
Bias $\beta_3$	0.27	0.03	0.24	0.02
Bias $\beta_4$	0.23	0.02	0.24	0.02
Cosine weight (0-1)	0.72	0.06	0.68	0.06
Sine weight (0-1)	0.28	0.06	0.32	0.06
RMS	0.11	0.01	0.11	0.02

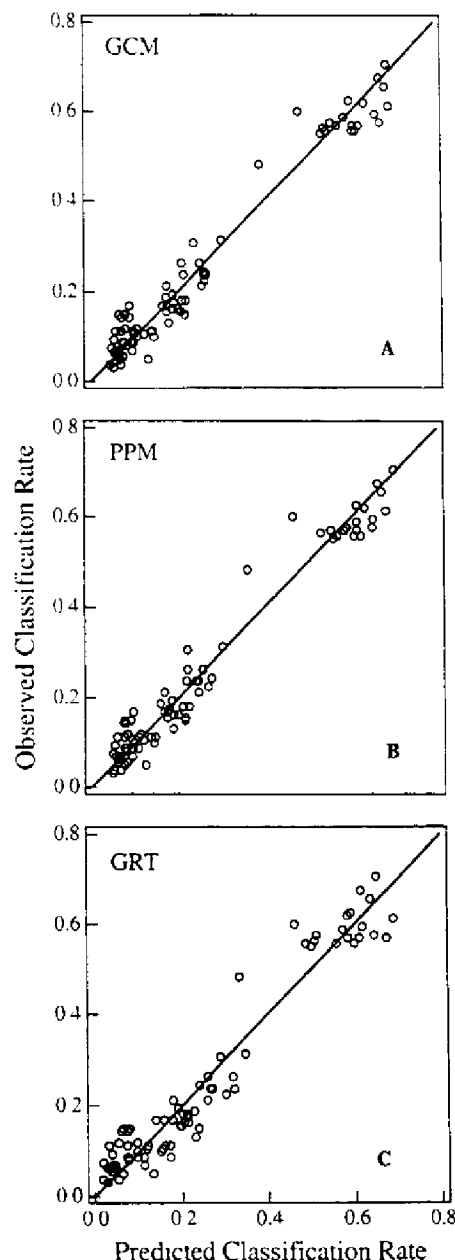


Figure 5. Relationship between observed and predicted classification rates. Fits are shown for the GCM (panel A), PPM (panel B), and GRT (panel C). Each point represents the proportion of trials in which a stimulus was classified into a given category. The model accounts for the data to the extent that the points fall along the diagonal line.

**Generalized recognition theory (GRT).** A simplified version of GRT was used to fit these data. Equal covariance matrices for each exemplar and linear decision bounds with no criterial noise were assumed. This version of GRT has the same number of free parameters

Table 4  
Parameter Estimates for Generalized Recognition Theory (GRT)  
Applied to Classification Data From Experiment 1

Parameters	GRT
$\sigma_{\cos}$	00.071
$\sigma_{\sin}$	00.135
Correlation $\rho$	00.006
Vertical bound orientation	86.3°
Vertical bound $x$ intercept	-0.01
Horizontal bound orientation	7.4°
Horizontal bound $y$ intercept	-0.02
$\sigma_{\sin}/\sigma_{\cos}$	1.9

(seven) as the GCM and the PPM, thus providing a basis for comparisons among these models. Under these constraints, GRT did not fare as well as the distance-based models (see Figure 5C). For the best-fitting parameter fits to the averaged data (see Table 4), the model deviated substantially from the data ( $\chi^2 = 154$ ,  $df = 72$ ,  $p = .001$ ). In contrast, for the individual subject fits, the model did not do as poorly (mean RMS = 0.12). As in signal detection theory, GRT may be particularly sensitive to changes in the decision contour from subject to subject. For this reason, it is particularly important to fit the data separately for each subject. Parameter values for these fits are shown in Table 5.

In GRT, differences in observers' sensitivity to each dimension are modeled by the variability parameter associated with each dimension. In fitting the model to the averaged classification data, the variability of the sine dimension was nearly twice that of the cosine dimension, and there was essentially no covariance between the two dimensions. Similar results were obtained in fitting the GCM to individual subject data.

### Evaluation

In order to statistically compare the relative predictions of these different models, an analysis of variance (ANOVA) was conducted on the RMS values for each subject. The ANOVA revealed a reliable difference in the

Table 5  
Means and Standard Errors of Parameter Estimates for  
Generalized Recognition Theory (GRT) Applied to  
Individual Subject Classification Data  
From Experiment 1

Parameters	$M$	$SE$
$\sigma_{\cos}$	0.08	0.01
$\sigma_{\sin}$	0.11	0.01
Correlation $\rho$	0.04	0.01
Vertical bound orientation	83.1°	2.6
Vertical bound $x$ intercept	0.01	0.00
Horizontal bound orientation	12.1°	3.7
Horizontal bound $y$ intercept	-0.02	0.01
$\sigma_{\sin}/\sigma_{\cos}$	1.92	0.28
RMS	0.12	0.01

model fits [ $F(2,15) = 28.8$ ,  $p < .001$ ]. Fisher's least significant difference test was used to test the significance of the individual comparisons. This test is appropriate for comparisons among three variables (Howell, 1992). Results indicated that both the GCM and the PPM model provided a better fit to the classification data than did the version of GRT that assumes linear bounds, equal covariance matrices for all stimuli, and no criterial noise ( $p < .001$ ). No significant difference in the fits of the PPM and GCM models was detected. Because GRT is a very powerful model, it is possible that a more general version of GRT would at least match, if not improve on, the predictions of the GCM and PPM models. Nonetheless, the GCM and PPM models capture some essential characteristics of the classification data. Under more specialized experimental conditions, it may be possible to dissociate these different models more clearly.

To demonstrate that the quality of the fit of each of the models is largely due to the differential noise or weighting of the cosine and sine dimensions, we tested a version of each model in which the dimensions were treated identically while allowing all other parameters to vary. With this constraint, none of the models adequately fit the classification data ( $\chi^2_{\text{GCM}} = 221.1$ ;  $\chi^2_{\text{PPM}} = 201.2$ ;  $\chi^2_{\text{GRT}} = 293.5$ ). For all of the models, weighting the sine and cosine dimensions equally resulted in a significantly worse fit to the data ( $p < .001$ ). These results are in qualitative agreement with previous phase discrimination studies that found that highly practiced observers are two to three times more sensitive to cosine ( $0^\circ$ - $180^\circ$ ) than to sine ( $90^\circ$ - $270^\circ$ ) phase shifts (Bennett & Banks, 1991).

### MDC Model and the Principle of "Virtual Prototypes"

Caelli et al. (1987) studied classification performance of subjects using  $f+3f$  gratings. They suggested that a simple prototype model (called the *minimum distance classifier*, or MDC) provides a good fit to their classification data. In this section, we examine Caelli et al.'s MDC model and a recent extension that employs "virtual prototypes" (Rentschler & Caelli, 1990). Unlike the more general classification models reviewed here, the MDC model does not allow for differential weighting of stimulus dimensions (an important aspect of the data) and assumes a reciprocal relationship between similarity and Euclidean distance. According to the MDC, the probability of a stimulus  $i$  being assigned to a category  $J$  is given by

$$P(R_i | S_i) = \frac{\eta_{iJ}}{\sum_K \eta_{iK}},$$

where  $\eta_{iJ}$  represents the similarity of exemplar  $i$  to the prototype of category  $J$ , and  $\eta_{iK}$  represents the similarity of exemplar  $i$  to the prototype of category  $K$ . Similarity is defined as the reciprocal of distance in an unweighted Euclidean space (e.g., Equation 2, with  $c = 1$ ,  $r = 2$ , and no weighting parameters). When applied to our clas-



sification data, the MDC model did not provide an adequate fit ( $\chi^2 = 277$ ,  $df = 80$ ,  $p < .001$ ).<sup>2</sup>

Rentschler and Caelli (1990) proposed an extension of their MDC model in which virtual prototypes rather than physical prototypes are used to determine class membership. In this version of their model, distance is computed between the exemplar and the category's *virtual* prototype. Virtual prototypes are computed by manipulating the coordinates of the physical prototype so as to minimize the error of the model. The MDC virtual prototype model was still unable to provide an adequate account for our data ( $\chi^2 = 217$ ,  $df = 72$ ,  $p < .001$ ). Nonetheless, the difference in goodness of fit between the virtual prototype MDC model and the regular MDC model was statistically significant ( $\Delta\chi^2 = 60$ ,  $\Delta df = 8$ ,  $p < .001$ ).

There are several possible reasons why both the MDC and the virtual prototype MDC models fare poorly when compared with the PPM. First, it may be that the differential weighting of the dimensions is critical. This would account for the failure of the MDC model, but not for the failure of the virtual prototype MDC model. With virtual prototypes, the dimensions can be stretched to reflect the differential sensitivity of the sine- and cosine-sensitive mechanisms. Another factor that may account for the difference between the models is the variable Minkowski metric in the PPM and GCM. This, however, cannot explain the difference because the best-fitting versions of PPM and GCM result in a Euclidean metric ( $r = 2$ ). Finally, it could be that the transformation from distance to similarity is what distinguishes these models. In the MDC, the function is simply the reciprocal of distance, whereas in the PPM it is an exponential function (Equation 3). To test this hypothesis, we evaluated the virtual prototype MDC model with an exponential decay function relating similarity and distance (Equation 3, with  $p = 1$ ). This version of the virtual prototype model provided an excellent fit to the data ( $\chi^2 = 76$ ,  $df = 71$ ,  $p > .50$ ), suggesting that the transformation between distance and similarity is a particularly important part of a classification model.

Consider how the virtual prototype model with an exponential function relating similarity to distance might account for our data. The solid squares in Figure 6 depict the prototypes—defined as the average of the five exemplars—for the four categories in our study. The virtual prototype model shifts the cosine and sine coordinates of each prototype to account for the classification performance. The virtual prototypes that provide the best fit to the data are indicated by the open symbols in Figure 6. Note that the major difference between the physical and virtual prototypes is the contraction along the sine dimension: The sine dimension is contracted 2.06 times relative to the cosine dimension. Thus, the virtual prototype model ends up being very similar to the general classification models (GCM, PPM, GRT) in so far as it weights the cosine dimension more than the sine dimension.

### The Effects of Learning

In their paper on virtual prototypes, Rentschler and Caelli (1990) attributed the differential scaling of the sine and cosine dimensions to the effects of perceptual learning. Specifically, they argued that learning improves sensitivity to changes along the sine dimension more than

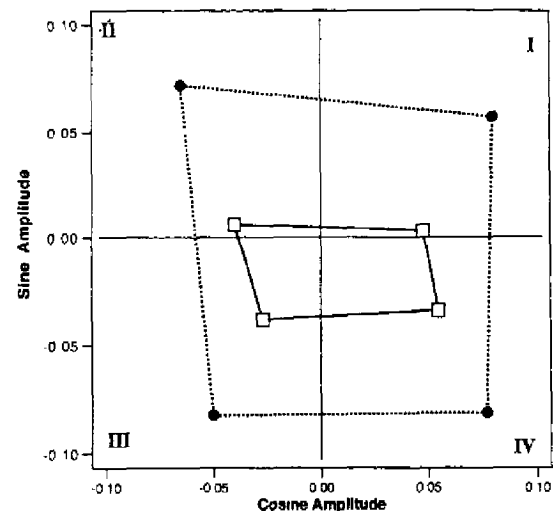


Figure 6. The cosine and sine amplitudes of the physical (filled circles) and virtual (open squares) prototypes of the stimuli shown in Figure 4.

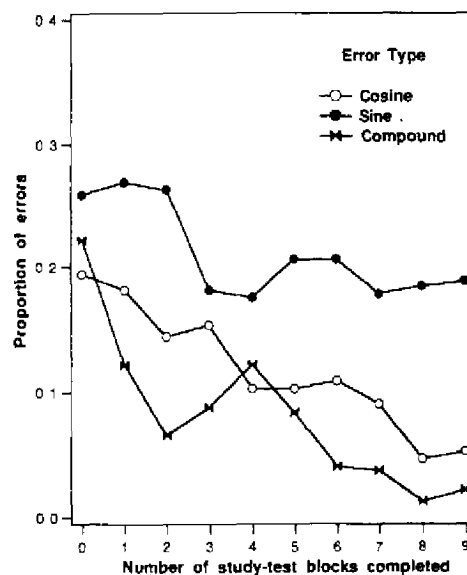


Figure 7. Learning data for Experiment 1. See text for details.

along the cosine dimension (see also Rentschler, 1990, cited in Shapley, Caelli, Grossberg, Morgan, & Rentschler, 1990). We examined the hypothesis that the differential weighting of the cosine and sine dimensions obtained in the present experiment was due to the inability of the subjects to attend to the sine dimension during early stages of learning. The average proportions of three different types of errors within each study-test block are shown in Figure 7. A sine error is a confusion between classes that differ only in sine amplitude (e.g., Classes I and IV in Figure 4). A cosine error is a confusion between classes that differ only in cosine amplitude (e.g., Classes III and IV in Figure 4). Finally, a compound error was defined as a confusion between classes that differ in both cosine and sine amplitude (e.g., Classes I and III in Figure 4). Figure 7 shows that sine errors were more prevalent than cosine errors in all study-test blocks. In fact, the difference between cosine and sine sensitivity appears to increase slightly during the course of the experiment. These results are inconsistent with the learning hypothesis. Of course, one could argue that the differential weighting of cosine and sine would disappear with sufficient learning. However, the results of discrimination studies suggest that this hypothesis is probably incorrect: Differential sensitivity to cosine and sine phase shifts persists even after many thousands of practice trials (Bennett & Banks, 1991).

The best-fitting Minkowski exponent for both the GCM and the PPM model was approximately 2 (see Tables 2 and 3). Furthermore, the subjects were able to combine sine and cosine information throughout the experiment, as indicated by the low classification error rates when the stimuli differed along both dimensions simultaneously (see Figure 7). Both of these results indicate that the subjects were able to integrate sine and cosine information in our classification task. In contrast, discrimination studies show that there is little summation between sine and cosine dimensions: At least when averaged across subjects, the Minkowski exponent is significantly greater than 2 (Bennett, 1993). The present findings suggest that summation between cosine and sine channels increases at supra-threshold levels of  $2f$  contrast.

## EXPERIMENT 2

Although subjects are differentially sensitive to sine and cosine information, Experiment 1 demonstrated that they are capable of utilizing both types of information to classify stimuli. Nevertheless, subjects often report that the sine dimension is much less salient than the cosine dimension. Indeed, informal examination of the stimuli used in Experiment 1 reveals that stimuli that differ only along the sine dimension ( $45^\circ$  vs.  $315^\circ$  in Figure 3) look similar, despite the fact that they are clearly discriminable. Such informal and introspective reports suggest that although the magnitude of the sine component can be used, when necessary, to make discriminations, it does not greatly influence perceived similarity (see Rentschler & Caelli, 1990, for a similar proposal). It is quite possible, for example, that similarity between  $f+2f$  gratings is based

on features or dimensions that are only indirectly related to the sine and cosine dimensions. One might expect, therefore, that similarity judgments measured in a context devoid of external reinforcement for attending to specific sine and cosine dimensions would not be strongly related to those dimensions. In Experiment 2, we examined the judged similarity of compound sinusoidal gratings.

## Method

**Stimuli.** The stimuli consisted of the same  $20f+2f$  gratings used in the classification experiment. The sine and cosine components of these stimuli are depicted in Figure 4. The patterns were presented on a high-resolution SuperMac 21-in. Monochrome Display (Model MM2136ASM). Display size was  $1,000 \times 1,000$  pixels (77 pixels per inch). The frame rate was 77 Hz, noninterlaced. Three stimuli were presented in a triangular configuration on each trial. Each stimulus was presented within a circular aperture that subtended  $4.4^\circ$ . The entire display subtended  $17^\circ$ . The average luminance ( $14 \text{ cd/m}^2$ ) and chromaticity of the stimulus patches and surrounding uniform field were identical.

**Procedure.** The subjects made similarity judgments on all possible combinations of three stimuli chosen from the total stimulus sample. The order in which the stimuli within a triad were examined and the duration of looking devoted to individual stimuli were not controlled. Therefore, we obtained similarity judgments for only one spatial configuration of each triad, rather than obtaining judgments for all six spatial configurations. For each triad, the subjects were asked to select the "odd one out." This procedure, known as the *triangular* method of triads (Ennis, Mullen, Frijters, & Tindall, 1989), requires only a single judgment per triad—namely, subjects are told to select the stimulus that they perceive to be most different from the other two (e.g., Romney, Brewer, & Batchelder, 1993; Weller & Romney, 1990; Wexler & Romney, 1972). Each possible stimulus pair (A,B) is presented with every other stimulus (C). A similarity matrix is then constructed by counting the number of times the stimulus C is selected as the "odd one out" in the triad (A, B, C). If stimulus C is always selected as the "odd one out," then stimuli A and B are deemed highly similar. On the other hand, if stimulus C is never selected, then A and B are highly dissimilar.

All testing was done with the room lights off. Viewing was binocular through natural pupils. Viewing distance was 1 m. Head position was stabilized with a chin/forehead rest. The subjects made their selections by clicking on the chosen item with a mouse and indicated that they were satisfied with their selection by clicking on a button drawn in the center of the display. No time limit was placed on the responses, but a typical trial lasted only several seconds.

**Subjects.** Five subjects made similarity judgments on the 1,140 stimulus triads.<sup>3</sup> The subjects, who were researchers working in the lab, were naive to the specifics of this experiment. A single session took approximately 3 h.

## Results and Discussion

The proximity matrix derived from the triadic judgments is shown in Table 6. Summary statistics for the multidimensional scaling (MDS) solutions are shown in Table 7. To determine the appropriate number of dimensions to accept in the MDS solution, we adopted Schwarz's (1978) Bayesian information criterion (BIC).<sup>4</sup> Using the BIC statistic addresses the problem of overestimation that occurs when using a maximum likelihood criterion: Maximum likelihood estimation will always prefer the largest number of dimensions that the data will per-

Table 6  
Proximity Matrix for Experiment 2

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
A	—																			
B	0.64	—																		
C	0.66	0.69	—																	
D	0.88	0.56	0.86	—																
E	0.88	0.66	0.93	0.87	—															
F	0.13	0.17	0.09	0.10	0.12	—														
G	0.12	0.12	0.11	0.08	0.11	0.89	—													
H	0.12	0.22	0.17	0.10	0.12	0.73	0.86	—												
I	0.19	0.18	0.07	0.09	0.10	0.89	0.81	0.63	—											
J	0.12	0.09	0.11	0.09	0.13	0.84	0.88	0.79	0.82	—										
K	0.08	0.11	0.06	0.00	0.01	0.51	0.38	0.40	0.59	0.37	—									
L	0.00	0.11	0.06	0.01	0.04	0.52	0.50	0.56	0.38	0.51	0.84	—								
M	0.04	0.03	0.03	0.02	0.01	0.51	0.44	0.37	0.53	0.49	0.90	0.57	—							
N	0.00	0.04	0.09	0.04	0.06	0.47	0.53	0.53	0.43	0.67	0.62	0.79	0.69	—						
O	0.01	0.06	0.08	0.02	0.04	0.42	0.52	0.63	0.43	0.57	0.64	0.94	0.62	0.90	—					
P	0.59	0.40	0.46	0.62	0.56	0.03	0.01	0.00	0.07	0.00	0.14	0.10	0.13	0.10	0.08	—				
Q	0.46	0.18	0.37	0.54	0.49	0.02	0.00	0.00	0.07	0.01	0.13	0.11	0.12	0.10	0.11	0.84	—			
R	0.58	0.37	0.46	0.67	0.52	0.03	0.00	0.00	0.07	0.00	0.14	0.08	0.13	0.09	0.10	0.90	0.88	—		
S	0.53	0.40	0.34	0.47	0.37	0.01	0.00	0.00	0.03	0.00	0.33	0.12	0.32	0.10	0.11	0.89	0.83	0.91	—	
T	0.37	0.68	0.24	0.26	0.24	0.09	0.02	0.03	0.09	0.03	0.51	0.34	0.30	0.11	0.18	0.63	0.41	0.64	0.84	—

mit, whereas the BIC statistic penalizes models that employ an excessive number of free parameters. Formally, the BIC is given by

$$\text{BIC} = \log L - \frac{k \log n}{2},$$

where  $k$  is the number of dimensions of the model,  $n$  is the number of observations, and  $L$  is the log likelihood for the total sample (see Ramsay, 1991, for details).

The two-dimensional solution minimized the BIC statistic and accounted for significantly greater variance than the one-dimensional solution. Although adding a third dimension significantly improved the fit ( $\Delta\chi^2 = 60$ ,  $df = 17$ ,  $p < .001$ ), the BIC statistic increased, suggesting that the improvement in fit was due to an excessive number of free parameters. The two-dimensional MDS solution with a Euclidean metric ( $r = 2$  in Equation 1) is shown in Figure 8.

Linear multiple-regression analyses were conducted to determine to what extent the MDS dimensions mapped onto the physical dimensions of sine and cosine components of  $2f$ . For the two-dimensional solution, MDS Dimension 1 was significantly related to cosine amplitude (adjusted  $R^2 = 0.88$ ,  $t = -11.83$ ,  $p < .001$ )<sup>5</sup> and MDS Dimension 2 was significantly related to sine amplitude (adjusted  $R^2 = 0.38$ ,  $t = 3.69$ ,  $p < .01$ ). There was no hint of a relationship between Dimension 1 and sine amplitude ( $t = -0.62$ ) or Dimension 2 and cosine amplitude ( $t = 0.01$ ). The relation between Dimension 1 and cosine amplitude and the relation between Dimension 2 and sine amplitude are shown in Figure 9.

The cosine advantage typically observed in discrimination studies, and also exhibited in our classification data, seems to generalize to the present similarity study. As is shown in the MDS solution in Figure 8, the stimuli are stretched out along Dimension 1 (which is highly correlated with cosine amplitude) but not along Dimension 2

(which is related to sine amplitude). These data provide evidence that the cosine dimension is more salient than the sine dimension in judging the similarity of compound gratings.

### EXPERIMENT 3

Although the results of Experiment 2 seem to conform to the predictions of the two-channel model, the stimuli used here were segregated into distinct categories. One might argue that the categorical structure of the stimuli suggested the dimensions to be used in making similarity judgments. To test the generality of these results, we conducted a third experiment using stimuli that varied continuously along the sine and cosine dimensions.

#### Method

**Stimuli.** Stimuli were 16 compound  $f+2f$  gratings. The frequency and contrast of  $f$  were the same as in Experiments 1 and 2. The phase of  $2f$  was generated by randomly selecting a phase angle from a uniform distribution ranging from  $0^\circ$  to  $360^\circ$ . The contrasts of  $2f$  were selected from a Gaussian distribution with a mean of 0.05 and a standard deviation of 0.016. The sine and cosine components of these stimuli are depicted in Figure 10. Note that the stimuli were more evenly distributed in phase space than were those used in Experiment 2.

**Procedure.** The procedure in this experiment was identical with that of Experiment 2. Each subject viewed all 560 unique combi-

Table 7  
MDS Results for Similarity Judgments Collected  
in Experiment 2

Analysis	Log Likelihood	BIC Statistic	Number of Parameters
One dimensional	-925	2,057	32
Two dimensional	-767	1,855	50
Three dimensional	-737	1,904	67

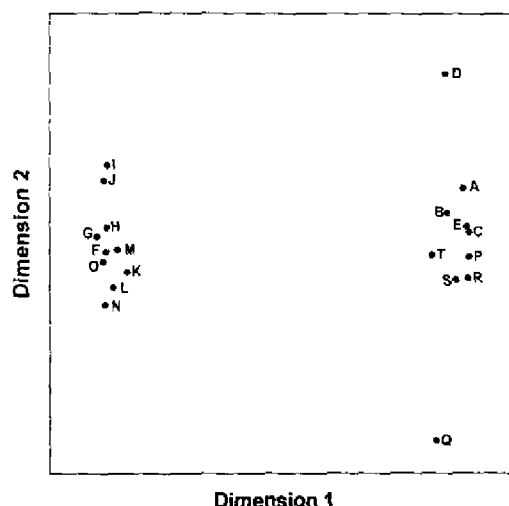


Figure 8. The MDS solution for the similarity judgments of Experiment 2. The letters are the stimulus labels (see Figure 4).

nations of three stimuli. As in the previous experiment, we obtained similarity judgments for one spatial configuration of each triad, rather than obtaining judgments for all six spatial configurations.

**Subjects.** The subjects were 7 psychology graduate students. Each subject was paid \$15 (Canadian) for participating in the experiment. Each session lasted approximately 1.5 h.

### Results and Discussion

The proximity matrix derived from the triadic judgments is shown in Table 8. Summary statistics for the fit of one- through three-dimensional solutions are shown in Table 9. As in Experiment 2, the two-dimensional solution minimized the BIC statistic and accounted for significantly greater variance than did the one-dimensional solution. Again, the three-dimensional solution significantly improved the fit ( $\Delta\chi^2 = 70$ ,  $df = 13$ ,  $p < .001$ ); however, the BIC statistic favored the two-dimensional solution (shown in Figure 11).

Linear multiple-regression analyses were conducted to determine to what extent the MDS dimensions mapped onto the sine and cosine components of  $2f$ . For the two-dimensional solution, MDS Dimension 1 was significantly correlated only with cosine amplitude (adjusted  $R^2 = 0.94$ ,  $t = 15.68$ ,  $p < .001$ ); however, MDS Dimension 2 was not correlated with either sine or cosine amplitude (adjusted  $R^2 \approx 0.00$ ). Although sine amplitude was not linearly correlated with Dimension 2, inspection of the data suggested that the subjects perceived stimuli with positive and negative sine amplitude as being similar to one another but different from stimuli with sine amplitudes near zero. To test for this possibility, we correlated the absolute value of the sine amplitude with Dimension 2 and found a substantial correlation ( $R^2 = 0.68$ ,  $F = 30.25$ ,  $p < .001$ ). This suggests that subjects are sensitive to the presence of sine contrast, but not to its sign. For example, gratings in  $45^\circ$  and  $315^\circ$  phase (see Figure 3) are judged as similar to one another but different from grat-

ings at  $0^\circ$  phase. The relation between Dimension 1 and cosine amplitude and the relation between Dimension 2 and the absolute value of the sine amplitude are shown in Figure 12.

This experiment extended the results of Experiment 2 by demonstrating that even when stimuli vary continuously along the sine and cosine dimensions, a two-dimensional solution provides a reasonable account for perceived similarity judgments. Nonetheless, the results were not entirely consistent across these experiments. In Experiment 2, similarity judgments were linearly corre-

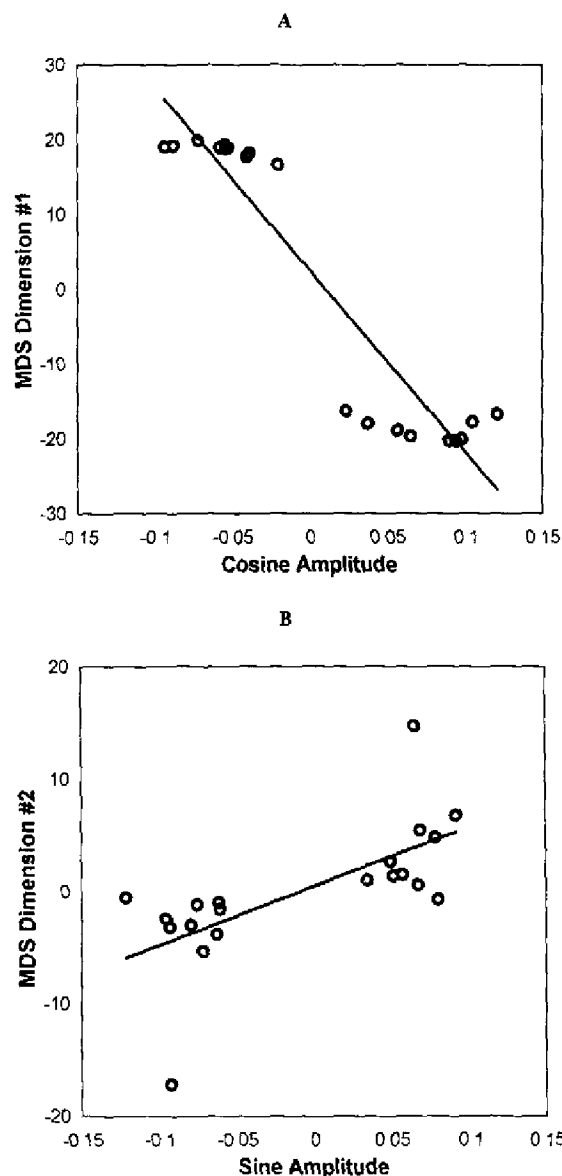


Figure 9. The relation between cosine amplitude and MDS Dimension 1 (panel A) and the relation between sine amplitude and MDS Dimension 2 (panel B). The dimensions are the results of the two-dimensional MDS solution for Experiment 2.

Table 8  
Proximity Matrix for Experiment 3

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
A	—															
B	0.66	—														
C	0.65	0.31	—													
D	0.64	0.47	0.53	—												
E	0.42	0.18	0.59	0.56	—											
F	0.21	0.61	0.13	0.21	0.04	—										
G	0.78	0.41	0.56	0.72	0.52	0.15	—									
H	0.22	0.10	0.43	0.27	0.72	0.02	0.30	—								
I	0.70	0.64	0.80	0.62	0.38	0.36	0.54	0.12	—							
J	0.12	0.33	0.04	0.14	0.02	0.74	0.12	0.08	0.15	—						
K	0.24	0.51	0.07	0.22	0.02	0.83	0.15	0.07	0.23	0.80	—					
L	0.29	0.08	0.44	0.31	0.69	0.01	0.34	0.84	0.20	0.06	0.01	—				
M	0.26	0.04	0.36	0.14	0.57	0.01	0.19	0.90	0.14	0.06	0.06	0.83	—			
N	0.47	0.12	0.61	0.36	0.69	0.04	0.62	0.65	0.35	0.05	0.05	0.71	0.64	—		
O	0.04	0.35	0.04	0.09	0.01	0.69	0.10	0.10	0.11	0.87	0.78	0.06	0.08	0.02	—	
P	0.02	0.30	0.03	0.07	0.01	0.71	0.06	0.09	0.07	0.83	0.69	0.05	0.09	0.02	0.89	—

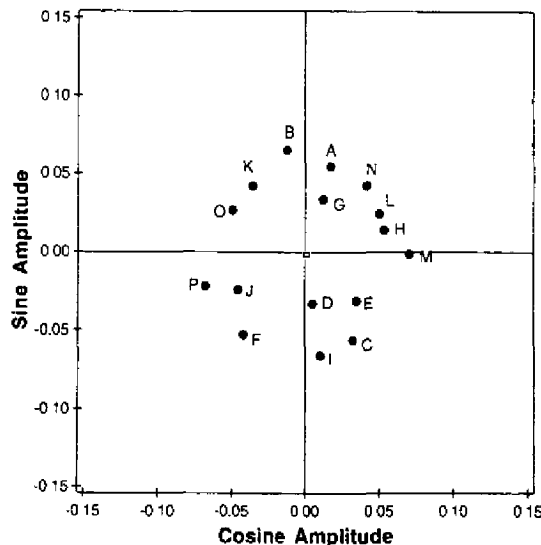


Figure 10. The stimuli used in Experiment 3. The figure conventions are the same as those for Figure 4.

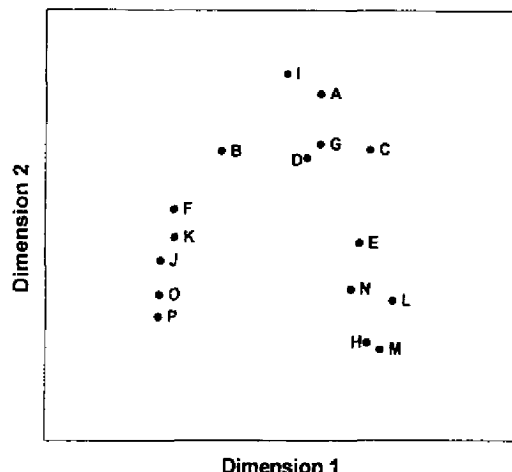


Figure 11. The MDS solution for the similarity judgments of Experiment 3. The letters are the stimulus labels (see Figure 10).

lated with sine amplitude, whereas in the present experiment, the subjects seemed to judge similarity according to the absolute value of the sine amplitude and not its sign. There are two possible explanations for this discrepancy. In order to make differences in sine and cosine amplitude more subtle in this experiment, the average contrast of  $2f$  was only half that of Experiment 2. Consequently, one might reason that the subjects were simply unable to discriminate the sign of the sine component. This explanation is unlikely, because Bennett and Banks (1991) have shown that when patterns are viewed in the fovea (as in our experiments), one can discriminate phase shifts as soon as  $2f$  is detected. An alternative explanation seems more plausible. In Experiment 2, stimuli were clustered around prototypes in phase space. In order to judge the four categories as distinct, the subjects had to attend to the sign of the sine and cosine amplitudes. For example, stimuli in Categories I and IV (see Figure 4) differed primarily in the directionality rather than the absolute value of their sine amplitude. In contrast, stimuli in Experiment 3 (see Figure 10) varied continuously and did not conform to any obvious categorical structure. In this case, the subjects could ignore the directionality of the sine amplitude if its absolute value were more salient.

## GENERAL CONCLUSIONS

These experiments investigated whether models of phase discrimination would generalize to suprathreshold phase perception in classification and similarity-judgment tasks. In Experiment 1, classification performance was consistent with the predictions of the two-channel model of phase discrimination (Bennett & Banks, 1991; Field & Nachmias, 1984). The subjects were more sensitive to the cosine than to the sine information, and classification models that base judgments on interstimulus similarities along independent sine and cosine dimensions provided a good fit to the data. Prototype and exemplar models provided nearly identical fits to the data, whereas a somewhat impoverished version of generalized recognition theory (matched to the other models in the number of free parameters) provided a slightly worse fit to the data.

Table 9  
MDS Results for Similarity Judgments Collected  
in Experiment 3

Analysis	Log Likelihood	BIC Statistic	Number of Parameters
One dimensional	-1,014	2,251	34
Two dimensional	-873	2,060	48
Three dimensional	-838	2,075	61

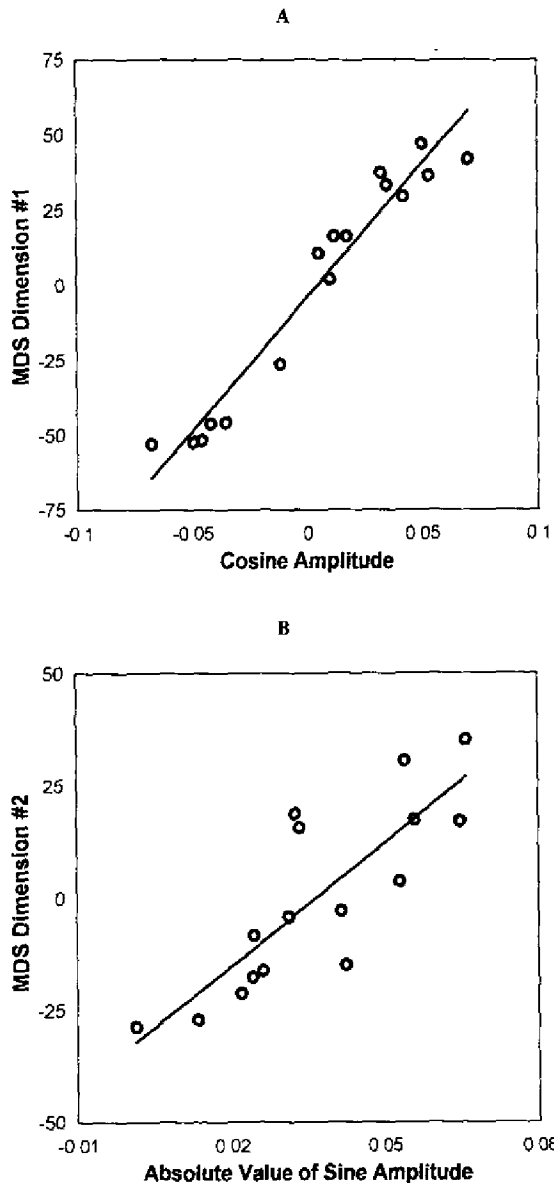


Figure 12. The relation between cosine amplitude and MDS Dimension 1 (panel A) and the relation between the absolute value of the sine amplitude and MDS Dimension 2 (panel B). The dimensions are the results of the two-dimensional MDS solution for Experiment 3.

In Experiments 2 and 3, a two-dimensional MDS solution provided an adequate characterization of the data. Higher dimensional solutions were rejected on the basis of the Bayesian information criterion (Schwarz, 1978). In both experiments, perceived similarity of compound gratings was influenced primarily by differences in cosine phase. Sine phase was also correlated with the multidimensional scaling solutions, but less strongly. Overall, these results suggest that the two-channel model of phase discrimination provides a useful framework for interpreting classification and similarity judgments of compound gratings.

## REFERENCES

- AKAIKE, H. (1974). A new look at the statistical identification model. *IEEE Transactions on Automatic Control*, 19, 716-723.
- ASHBY, F. G., & GOTT, R. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 14, 33-53.
- ASHBY, F. G., & MADDOX, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, 37, 372-400.
- ASHBY, F. G., & PERRIN, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review*, 95, 124-150.
- BADCOCK, D. (1984). How do we discriminate spatial phase? *Vision Research*, 24, 613-623.
- BENNETT, P. J. (1989). An investigation of spatial phase-encoding mechanisms in human adults and infants. *Dissertation Abstracts International*, 50, 2182-B.
- BENNETT, P. J. (1993). The harmonic bandwidth of phase-reversal discrimination. *Perception & Psychophysics*, 53, 292-304.
- BENNETT, P. J., & BANKS, M. S. (1987). Sensitivity loss in odd-symmetric mechanisms underlies phase anomalies in peripheral vision. *Nature (London)*, 326, 873-876.
- BENNETT, P. J., & BANKS, M. S. (1991). The effects of contrast, spatial scale, and orientation on foveal and peripheral phase discrimination. *Vision Research*, 31, 1759-1786.
- BRADDICK, O. J., ATKINSON, J., & WATTAM-BELL, J. R. (1986). Development of discrimination of spatial phase in infancy. *Vision Research*, 26, 1223-1240.
- BURR, D. C., MORRONE, M. C., & SPINELLI, D. (1989). Evidence for edge and bar detectors in human vision. *Vision Research*, 29, 419-431.
- CAELLI, T., RENTSCHLER, I., & SCHEIDLER, W. (1987). Visual pattern recognition in humans. *Biological Cybernetics*, 57, 233-240.
- ENNIS, D. M., MULLEN, K., FRUTERS, J. E. R., & TINDALL, J. (1989). Decision conflicts: Within-trial resampling in Richardson's method of triads. *British Journal of Mathematical & Statistical Psychology*, 42, 265-269.
- FIELD, D. J., & NACHMIAS, J. (1984). Phase reversal discrimination. *Vision Research*, 24, 333-340.
- HOWELL, D. C. (1992). *Statistical methods for psychology* (3rd ed.). Boston: TWS-Kent.
- LUCE, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 103-189). New York: Wiley.
- MEDIN, D. L., & SCHAFFER, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85, 207-238.
- MORGAN, M. J., ROSS, J., & HAYES, A. (1991). The relative importance of local phase and local amplitude in patchwise image reconstruction. *Biological Cybernetics*, 65, 113-119.
- NELDER, J. A., & MEAD, R. (1965). A simplex method for function minimization. *Computer Journal*, 7, 308-313.
- NOSOFKY, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115, 39-57.

- NOSOFSKY, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 13, 87-108.
- NOSOFSKY, R. M. (1992). Exemplar-based approach to relating categorization, identification, and recognition. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 363-394). Hillsdale, NJ: Erlbaum.
- OPPENHEIM, A. V., & LIM, J. S. (1981). The importance of phase in signals. *Proceedings of the IEEE*, 69, 529-541.
- PIOTROWSKI, L. N., & CAMPBELL, F. W. (1982). A demonstration of the visual importance and flexibility of spatial frequency amplitude and phase. *Perception*, 11, 337-346.
- RAMSAY, J. O. (1991). *Multiscale manual (extended version)*. Unpublished manuscript.
- REED, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, 3, 382-407.
- RENTSCHLER, I., & CAELLI, T. (1990). Visual representations in the brain: Inferences from psychophysical research. In H. Haken & M. Stadler (Eds.), *Synergetics of cognition* (pp. 233-248). Heidelberg: Springer-Verlag.
- RENTSCHLER, I., & TREUTWEIN, B. (1985). Loss of spatial phase relationships in extrafoveal vision. *Nature (London)*, 313, 308-310.
- ROMNEY, A. K., BREWER, D. D., & BATCHELDER, W. H. (1993). Predicting clustering from semantic structure. *Psychological Science*, 4, 28-34.
- SCHWARZ, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464.
- SHAPLEY, R., CAELLI, T., GROSSBERG, S., MORGAN, M., & RENTSCHLER, I. (1990). Computational theories of visual perception. In L. Spillman & J. S. Werner (Eds.), *Visual perception: The neurophysiological foundations* (pp. 417-447). San Diego, CA: Academic Press.
- SHEPARD, R. (1987). Toward a universal law of generalization for psychological science. *Science*, 237, 1317-1323.
- SYSTAT: *Statistics, Version 5.2 Edition*. (1992). Evanston, IL: SYSTAT, Inc.
- WELLER, S. C., & ROMNEY, A. K. (1990). Metric scaling: Correspondence analysis. Newbury Park, CA: Sage.
- WEKLER, K. N., & ROMNEY, A. K. (1972). Individual variations in cognitive structures. In A. K. Romney, R. N. Shepard, & S. B. Nerlove (Eds.), *Multidimensional scaling* (Vol. 2, pp. 73-92). New York: Seminar Press.
- WILSON, H. R., LEVI, D., MAFFEI, L., ROVAMO, J., & DEVALOIS, R. (1990). The perception of form: Retina to striate cortex. In L. Spillman & J. S. Werner (Eds.), *Visual perception: The neurological foundations* (pp. 231-272). New York: Academic Press.

## NOTES

1. Any combination of amplitude ( $A$ ) and phase ( $\phi$ ) can be expressed as a sum of cosine and sine components using the trigonometric identity

$$A \cos(x - \phi) = A \{\cos(\phi)\cos(x) + \sin(\phi)\sin(x)\}.$$

If  $A$  represents the threshold amplitude needed to discriminate a 180° phase shift, then the  $x$  and  $y$  coordinates in Figure 2 correspond to  $A \cos(\phi)$  and  $A \sin(\phi)$ , respectively.

2. Despite the model's poor fit to the data, the correlation between predicted and observed values was 0.96. This is consistent with Caelli et al.'s finding of a high correlation between the MDC model and their classification data.

3. The number of unique combinations of  $r$  stimuli chosen from a set of  $n$  is given by  $n!/r!(n-r)!$ . Since the total sample has 20 stimuli and each triad consists of 3, the number of combinations is 1,140.

4. The Bayesian information criterion is somewhat more conservative than Akaike's information criterion (Akaike, 1974) in that it prefers models of lower dimensionality.

5. Adjusted  $R^2 = R^2 - (p-1/n-p)(1-R^2)$ , where  $n$  is the number of cases and  $p$  is the number of predictors, including the constant (see *Systat: Statistics*, p. 176, for details).

(Manuscript received September 8, 1992;  
revision accepted for publication November 16, 1993)