Estimating the parameters of multidimensional signal detection theory from simultaneous ratings on separate stimulus components

F. GREGORY ASHBY
University of California at Santa Barbara, Santa Barbara, California

In the concurrent rating task, stimuli are constructed by factorially combining several levels of separate physical components. Subjects give simultaneous judgments of their confidence that each component is presented at the highest possible level. The concurrent rating task has great potential for testing hypotheses and estimating parameters of multidimensional signal detection theory. In this article, the relationship between the parameters of the distribution of rating responses and the underlying perceptual distributions is detailed. The product moment correlation between ratings is shown to seriously underestimate the degree of perceptual dependence. More accurate estimators of the correlation between perceptual effects are suggested. The relation between the interrating correlation and measures of perceptual dependence developed for complete identification tasks is discussed. A powerful test of perceptual separability is then developed. Finally, these techniques are applied to some representative data.

Multidimensional generalizations of signal detection theory have recently been employed to account for a wide variety of psychophysical and perceptual phenomena (e.g., Ashby & Gott, 1988; Ashby & Perrin, 1988; Ashby & Townsend, 1986; Graham, Kramer, & Yager, 1987; Hirsch, Hylton, & Graham, 1982; Olzak, 1986; Olzak & Wickens, 1983). Because of the difficulties in parameter estimation, earlier applications made many simplifying distributional assumptions (e.g., Green & Birdsall, 1978; Tanner, 1956). However, several recent studies have reported evidence indicating that the model is required in a full generality (Ashby & Gott, 1988; Ashby & Townsend, 1986; Hirsch et al., 1982; Olzak, 1986).

Hirsch et al. developed an experimental task, which I will refer to as the concurrent rating task, that enabled them to quickly estimate many of the free parameters in the most general multidimensional signal detection model (see, also, Olzak, 1986). In the concurrent rating task, two or more levels of two or more (unidimensional) stimulus components are factoredly combined to create a set of compound stimuli. For example, an $n \times m$ concurrent rating task has $n$ separate components, each with $m$ levels. On each experimental trial, the subject gives $n$ rating responses, one for each stimulus component. These ratings reflect the subject's confidence that the associated component was presented at the highest possible level. Both Hirsch et al. (1982) and Olzak (1986) chose sine-wave gratings of a certain frequency as their stimulus components, and the levels varied the contrast of the gratings. At the lowest level, the grating was always presented at zero contrast. Thus, the subjects were told to respond according to their confidence that a grating of each relevant frequency was presented.

In this article, I will show that the methods that have been used to determine the degree of interaction that occurs when separate components are processed have serious flaws, and I will suggest superior alternative methods. Finally, these methods will be applied to data from one of Olzak's (1986) subjects (reported by Olzak & Wickens, 1983).

**MULTIDIMENSIONAL SIGNAL DETECTION THEORY**

In multidimensional signal detection theory, the perceptual effect of each stimulus component is represented by a random vector. Thus, over many trials, the perceptual effects of a stimulus constructed from several separate components are represented by a multivariate probability distribution. Because both Hirsch et al. (1982) and Olzak (1986) employed only two stimulus components, they assumed that the perceptual distributions were bivariate. I will also assume bivariate perceptual distributions, but each of my results generalizes easily to the $n$-dimensional case.

The primary focus of this article is on possible interactions in the perceptual processing of the two stimulus components. There are two logically unrelated forms of perceptual interaction (Ashby & Townsend, 1986). First, on a particular trial, the perceptual effect of one component could be statistically dependent on the perceptual effect of the other. If no such statistical dependence occurs, then perceptual independence results. Second, over the course of many trials, the perceptual effect of one component

I would like to thank Nancy Perrin for her helpful comments on an earlier draft of this article. Requests for reprints should be sent to F. Gregory Ashby, Department of Psychology, University of California, Santa Barbara, CA 93106.
could depend on the level of the other component. If no such dependence occurs, perceptual separability results. 

Consider a stimulus ensemble constructed by factorially combining two levels of two components, A and B. Then the four stimuli are \( A_1B_1, A_1B_2, A_2B_1, \) and \( A_2B_2 \). Suppose \( x \) is the perceptual dimension associated with component \( A \) and \( y \) is the dimension associated with component \( B \), and denote the perceptual distribution (i.e., probability density function) corresponding to stimulus \( A_2B_2 \) by \( f_{A_2B_2}(x,y) \). If the perceptual distributions are multivariate normal, then \( f_{A_2B_2}(x,y) \) is completely determined by five parameters: the mean perceptual effect associated with each component, denoted by \( \mu_{A_2B_2}(x) \) and \( \mu_{A_2B_2}(y) \); the variance of the perceptual effects associated with each component, denoted by \( \sigma_{A_2B_2}^2(x) \) and \( \sigma_{A_2B_2}^2(y) \); and the correlation between the perceptual effects associated with each component, denoted by \( \rho_{A_2B_2}(x,y) \).

According to Ashby and Townsend (1986), when the perceptual distributions are multivariate normal, perceptual independence occurs if and only if \( \rho_{A_2B_2}(x,y) = 0 \). The degree to which \( \rho \) deviates from zero is a measure of the amount of perceptual dependence. On the other hand, perceptual separability of component \( A \) from component \( B \) holds if the marginal distribution associated with each level of component \( A \) does not depend on the level of component \( B \); that is, if

\[
\int_{-\infty}^{\infty} f_{A_2B_2}(x,y) dy = \int_{-\infty}^{\infty} f_{A_1B_2}(x,y) dy = f_{A_2B_2}(x) \tag{1}
\]

for \( p = 1 \) and 2 and for all values of \( x \). The reader is referred to Ashby and Townsend (1986) for a more thorough discussion and justification of these definitions.

In this article, two experimental paradigms play a central role, the concurrent rating task and the complete identification experiment. The following definition makes them explicit:

DEFINITION. Consider a stimulus ensemble constructed by factorially combining the different levels of \( n \) separate components. Suppose component \( i \) has \( m_i \) levels (so that the ensemble contains \( \prod_{i=1}^{n} m_i \) stimuli). On each experimental trial, one stimulus is selected from the ensemble and is presented to the subject. In a concurrent rating task, the subject gives \( n \) separate rating responses, one for each component. The \( i \)th rating, a number from 1 to \( k_i+1 \), reflects the subject's confidence that the \( i \)th component was presented at the highest possible level (i.e., at level \( m_i \)). If all components have \( m \) levels (i.e., \( m_1 = m_2 = \ldots = m \)) then the experiment will be called an \( n \times m \) concurrent rating task. In a complete identification experiment, the subject gives one response, which uniquely identifies the presented stimulus.

Note that one way to uniquely identify the stimulus is to name the level of each of its components. With \( m_i \) levels, a subject can name the level of a component by responding with a number from 1 to \( m_i \). Therefore, a complete identification experiment is functionally equivalent to a concurrent rating task, where the number of alternative ratings on each component equals the number of component levels (i.e., where \( k_i+1 = m_i \)). In the typical concurrent rating task, however, the number of alternative ratings exceeds the number of component levels \( (k+1 > m_i) \). On the other hand, it is possible for the number of alternative ratings to be less than the number of component levels \( (k+1 < m_i) \). In this case, the task becomes one of categorization, because the number of response alternatives is less than the number of stimulus alternatives. Thus, the concurrent rating task is an exceedingly general experimental paradigm.

Ashby and Townsend (1986) developed several tests of perceptual independence and perceptual separability that can be applied to data from a complete identification experiment. However, because of its increased generality, the concurrent rating task has even more potential than the complete identification experiment for testing hypotheses about the underlying perceptual distributions. Several such tests have already been proposed.

Both Hirsch et al. (1982) and Olzak (1986) computed the Pearson product moment correlation of the subject's ratings on each dimension and used the statistic as an estimate of \( \rho_{A_2B_2}(x,y) \). I will show below, however, that although the interrating correlation is related to the correlation between the underlying perceptual effects, it seriously underestimates \( \rho_{A_2B_2}(x,y) \). As an alternative, a more accurate estimator will be proposed. Olzak (1986) tested for perceptual separability by checking whether the means of the four perceptual distributions formed a rectangle. A rectangular configuration of the means is a necessary but not sufficient condition for perceptual separability (i.e., Equation 1) to hold. Thus, the means could form a rectangle even though perceptual separability fails. A stronger test of perceptual separability is suggested below.

THE CONCURRENT RATING TASK

All techniques for analyzing data from the concurrent rating task depend on the following critical assumption:

DECISIONAL SEPARABILITY. The subject's rating about one component does not depend on the perceived value of the other component.

Decisional separability should occur if the subject is able to follow the experimenter's instructions. If it does occur, we can assume that with \( k+1 \) ratings on each component, the subject selects \( k \) criteria on each dimension \( (X_1, X_2, \ldots, X_k) \) on dimension \( x \) and \( Y_1, Y_2, \ldots, Y_l \) on dimension \( y \) and, for example, responds with confidence \( R_x = i \) on dimension \( x \) if \( X_i - 1 < x \leq X_i \) and with confidence \( R_y = j \) on dimension \( y \) if \( Y_j - 1 < y \leq Y_j \). Variability in the perceptual effects \( x \) and \( y \) will cause variability in the ratings \( i \) and \( j \). Of critical interest is the relationship between the parameters of the perceptual distribution,
for all values of \( x \) and \( y \). Theorem 1 makes it obvious that if this condition holds, then the ratings \( R_x \) and \( R_y \) will be uncorrelated.

Of equal or greater interest, however, is the reverse implication. Do uncorrelated ratings imply perceptual independence? To answer this question, consider first the case in which the subject is asked to use a 2-point rating scale on each dimension. In this case, the covariance expression from Theorem 1 reduces to

\[
\text{cov}_{A_p B_q}(R_x, R_y) = F_{A_p B_q}(X_1, Y_1) - F_{A_p B_q}(X_1) F_{A_p B_q}(Y_1)
\]

Therefore, uncorrelated ratings for all different criterion placements implies perceptual independence. On the other hand, if more than a 2-point rating scale is used, it is easy to find criterion placements that cause the ratings to be uncorrelated, but for which

\[
F_{A_p B_q}(X_i, Y_j) \neq F_{A_p B_q}(X_i) F_{A_p B_q}(Y_j)
\]

for all values of \( i \) and \( j \), and thus, in general, uncorrelated ratings do not imply perceptual independence. However, if uncorrelated ratings are found for all possible criterion placements, then perceptual independence is implied. To see this, note that if, on each dimension, the subject places all but one of the criteria either above the largest possible perceptual effect or below the smallest possible perceptual effect, then a \( k+1 \) point rating scale is effectively reduced to a 2-point scale, and we have already seen that with a 2-point scale, uncorrelated ratings imply perceptual independence. We have therefore proved the following:

**Theorem 2.** If decisional separability holds in the \( 2 \times m \) concurrent rating task, then perceptual independence guarantees uncorrelated ratings. If uncorrelated ratings are found for all possible criterion placements, then perceptual independence is implied.

The special case in which the number of ratings on each dimension equals the number of levels of the component associated with that dimension is of particular interest, because, as previously noted, it makes the concurrent rating task functionally equivalent to a complete identification experiment. It is therefore of interest to compare the notion of uncorrelated ratings with measures of perceptual independence that have been developed for the identification task.

Call the response associated with stimulus \( A_p B_q \) in an identification task \( a_p b_q \), and suppose there are two levels of each component. For these conditions, Ashby and Townsend (1986) proposed testing for perceptual independence by examining the concept of sampling independence, which states that

\[
P(a_2 b_1 | A_p B_q) = [P(a_1 b_1 | A_p B_q) + P(a_2 b_1 | A_p B_q)]
\]

\[
\cdot [P(a_2 b_1 | A_p B_q) + P(a_2 b_1 | A_p B_q)].
\]

Ashby and Townsend (1986) showed that if decisional separability holds, then perceptual independence implies

**TESTING FOR PERCEPTUAL INDEPENDENCE**

As stated previously, the main focus of this article is on the correlation (or covariance) between ratings and its relation to the correlation between perceptual effects and therefore to the important concept of perceptual independence. By definition (Ashby & Townsend, 1986), the components of the stimulus \( A_p B_q \) are perceived independently if and only if

\[
F_{A_p B_q}(x, y) = F_{A_p B_q}(x) F_{A_p B_q}(y)
\]
sampling independence, and if sampling independence holds for all possible criterion placements, then perceptual independence is implied. This result is in exactly the same form as Theorem 2, thus suggesting a possible close relationship between sampling independence and uncorrelated ratings. The next result confirms this intuition.

**Theorem 3.** Suppose four stimuli, constructed by factorially combining two levels of two separate components, are used in both a 2 × 2 concurrent rating task with 2-point rating scales on each dimension and in a complete identification experiment. If decisional separability holds, and the subject employs the same response criteria in both tasks, then uncorrelated ratings in the concurrent rating task will occur if and only if sampling independence occurs in the complete identification task.

Thus, under the appropriate conditions, sampling independence and uncorrelated ratings are equivalent concepts. As such, the results established by Ashby and Townsend (1986) that relate sampling independence to perceptual independence are also relevant to the concurrent rating task.

We have now established that when perceptual independence occurs, an excellent method of verifying this independence is to correlate ratings in the concurrent rating task. However, if a perceptual dependence occurs, we might naturally ask (1) whether this ensures correlated ratings, and (2) how accurately the correlation between ratings, \( \epsilon_{A_0B_0}(R_x, R_y) \), estimates the correlation between percepts, \( \epsilon_{A_0B_0}(x, y) \).

If the perceptual effects are correlated, then there exists at least one \( x = x_0 \) and \( y = y_0 \), for which

\[
F_{A_0B_0}(x_0, y_0) \neq F_{A_0B_0}(x_0)F_{A_0B_0}(y_0).
\]

Of course, for certain criterion placements, uncorrelated ratings may still result. However, there will exist at least some placements that ensure correlated ratings. For example, if \( X_0 = x_0 \) and \( Y_0 = y_0 \) and all other criteria are greater than the largest possible perceptual effect, then the rating covariance becomes

\[
\text{cov}_{A_0B_0}(R_x, R_y) = F_{A_0B_0}(x_0, y_0) - F_{A_0B_0}(x_0)F_{A_0B_0}(y_0) \neq 0.
\]

Therefore, a perceptual dependence will, eventually, cause correlated ratings.

We finally turn to the numerical correspondence between \( \epsilon_{A_0B_0}(R_x, R_y) \) and \( \epsilon_{A_0B_0}(x, y) \). Analytically, this is a difficult problem. Therefore, to solve it, I turned to computer simulation. A total of 75 simulations were performed. At the beginning of each simulation, sets of criteria were defined on each of two dimensions. Next a bivariate distribution was sampled for 1,000 trials. For each sample, the criteria were used to create a rating on each dimension. At the end of the simulation, the correlation between ratings was estimated and compared with the correlation between perceptual effects.

In all simulations, the perceptual distributions were bivariate logistic. The logistic distribution was chosen because it is very similar in shape to the normal distribution but has a closed-form expression for its cumulative distribution function, and therefore is easy to simulate. The perceptual means on each dimension were arbitrarily set to 10.0 and the perceptual variances were set to 1.0. About half the simulations involved a 4-point rating scale, and half involved a 6-point scale. Although the product moment correlation between ratings was not as good a predictor of the perceptual correlation when a 4-point scale was used, overall the same pattern of results was obtained with both scales. Results of some of the simulations with the 6-point rating scale are given in Table 1. An asterisk indicates that the null hypothesis that the rating correlation equals the perceptual correlation was rejected at \( \alpha = .01 \). The column marked \( r_{R_x, R_y} \) will be discussed later in this article.

Note that a very consistent pattern emerges from Table 1. When the criteria are centered around the perceptual mean, the product moment correlation between ratings very accurately estimates the perceptual correlation. However, as the criteria become centered either

<table>
<thead>
<tr>
<th>Simulation Number</th>
<th>Criteria on Each Dimension</th>
<th>Perceptual Correlation ( \rho(x, y) )</th>
<th>Rating Correlation ( r(R_x, R_y) )</th>
<th>Tetrachoric Correlation ( r_{R_x, R_y} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8, 9, 10, 11, 12</td>
<td>.7</td>
<td>.69</td>
<td>.67</td>
</tr>
<tr>
<td>2</td>
<td>.3</td>
<td>.27</td>
<td>.27</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-.3</td>
<td>-.25</td>
<td>-.21</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-.7</td>
<td>-.70</td>
<td>-.69</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>9, 10, 11, 12, 13</td>
<td>.7</td>
<td>.58*</td>
<td>.66</td>
</tr>
<tr>
<td>6</td>
<td>.3</td>
<td>.24</td>
<td>.28</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>.3</td>
<td>-.30</td>
<td>-.31</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>-.7</td>
<td>-.58*</td>
<td>-.65</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>7, 8, 9, 10, 11</td>
<td>.7</td>
<td>.68</td>
<td>.72</td>
</tr>
<tr>
<td>10</td>
<td>.3</td>
<td>.25</td>
<td>.27</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>-.3</td>
<td>-.28</td>
<td>-.32</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>-.7</td>
<td>-.60*</td>
<td>-.66</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>10, 11, 12, 13, 14</td>
<td>.7</td>
<td>.38*</td>
<td>.72</td>
</tr>
<tr>
<td>14</td>
<td>.3</td>
<td>.19*</td>
<td>.32</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>-.3</td>
<td>-.08*</td>
<td>-.25</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>-.7</td>
<td>-.29*</td>
<td>-.65</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>6, 7, 8, 9, 10</td>
<td>.7</td>
<td>.46*</td>
<td>.66</td>
</tr>
<tr>
<td>18</td>
<td>.3</td>
<td>.25</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>-.3</td>
<td>-.11*</td>
<td>-.30</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>-.7</td>
<td>-.29*</td>
<td>-.68</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>11, 12, 13, 14, 15</td>
<td>.7</td>
<td>.11*</td>
<td>.78</td>
</tr>
<tr>
<td>22</td>
<td>.3</td>
<td>.04*</td>
<td>.37</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>-.3</td>
<td>.02*</td>
<td>-.01*</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>-.7</td>
<td>-.03*</td>
<td>-.49*</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>5, 6, 7, 8, 9</td>
<td>.7</td>
<td>.22*</td>
<td>.82*</td>
</tr>
<tr>
<td>26</td>
<td>.3</td>
<td>.11*</td>
<td>.50*</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>-.3</td>
<td>-.02*</td>
<td>-.30</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>-.7</td>
<td>-.03*</td>
<td>-.72</td>
<td></td>
</tr>
</tbody>
</table>

Note—In all cases, the perceptual mean on each dimension was 10.0, the variance was 1.0, and the sample size was 1,000. An asterisk in the rating correlation column indicates significance at the \( \alpha = .01 \) level. An asterisk in the tetrachoric correlation column indicates that the estimate is more than 2.5 standard errors from the perceptual correlation.
above or below the perceptual mean, the correlation between the ratings begins to diverge from the perceptual correlation. In fact, the ratings tend to become uncorrelated, no matter what the perceptual correlation. This is a serious problem for the \( n \times m \) concurrent rating task. To satisfy the experimenter’s instructions, the subject must spread the criteria out far enough to cover the \( m \) levels of each component. For example, when \( m = 2 \), we expect the criteria to be centered on each dimension below the upper mean and above the lower mean. In this case, the product moment correlation between ratings will severely underestimate the degree of perceptual correlation.

Other simulations were performed in which the criteria on one dimension were centered above the relevant perceptual mean and the criteria on the other dimension were centered below the relevant perceptual mean. The same pattern of results occurred. When the criteria were centered on either side of the mean, the rating correlation failed to accurately estimate the perceptual correlation. I also performed several simulations in which the perceptual effects were uncorrelated. In each of these, the rating correlation was very close to zero. Thus, the problem with using the product moment correlation between ratings as an estimate of the correlation between perceptual effects is restricted to instances in which a perceptual dependence exists and in which the response criteria are centered either above or below the perceptual means.

Given the obvious deficiencies in \( r_{Ap,Bp}(R_x,R_y) \) as an estimator of \( Q_{Ap,Bp}(x,y) \), it is natural to ask whether a better estimator can be found. As it turns out, this problem has been fairly extensively studied in the statistical literature. With a 2-point rating scale and when the perceptual distributions are bivariate normal, the maximum likelihood estimator of \( Q_{Ap,Bp}(x,y) \) is the tetrachoric \( r \), first suggested by Pearson (1901). The tetrachoric \( r \) is obtained by solving the following expression for the cumulative normal distribution function at the criteria, for \( r \):

\[
P^*(R_x=1, R_y=1) = [2\pi(1-r^2)]^{-1} \int_{-\infty}^{\hat{R}_x} \int_{-\infty}^{\hat{R}_y} \exp \left( -\frac{x^2-2rxy+y^2}{2(1-r^2)} \right) dx \, dy,
\]

(2)

where \( P^*(R_x=1, R_y=1) \) is the observed proportion of trials on which both ratings equal 1. The criteria can be estimated from the marginals:

\[
P^*(R_x=1) = (2\pi)^{-1/2} \int_{-\infty}^{\hat{R}_x} \exp(-x^2/2) dx
\]

and

\[
P^*(R_y=1) = (2\pi)^{-1/2} \int_{-\infty}^{\hat{R}_y} \exp(-y^2/2) dy.
\]

Pearson’s (1901) idea was to expand the right side of Equation 2 into a series in terms of \( r \), called the tetrachoric series. Unfortunately, the analytic solution for the tetrachoric \( r \) is long and complicated. Consequently, several methods for approximating the solution have been developed. Kirk (1973) used Gaussian quadrature, supplemented by Newton-Raphson iteration, to evaluate an equivalent form of the Equation 2 integral. However, a simpler approximation is based on the so-called cosine-pi formula (e.g., Guilford & Fruchter, 1973). Let

\[
a = \hat{P}_{Ap,Bp}(R_x=2, R_y=2)
\]

\[
b = \hat{P}_{Ap,Bp}(R_x=1, R_y=2)
\]

\[
c = \hat{P}_{Ap,Bp}(R_x=2, R_y=1)
\]

and

\[
d = \hat{P}_{Ap,Bp}(R_x=1, R_y=1).
\]

In terms of radians, the cosine-pi approximation to the tetrachoric \( r \) is

\[
r_{tel}(R_x,R_y) = \cos \left( \frac{\pi}{1 + \sqrt{ad/bc}} \right).
\]

(3)

The cosine-pi formula gives very close approximations to the tetrachoric \( r \), but only when the criterion on each dimension is placed at the perceptual median. In this respect, Equation 3 suffers the same shortcomings as the product moment correlation between ratings. The estimate becomes very bad as the criteria are moved either above or below the median.

When more than a 2-point rating scale is used, the tetrachoric \( r \) becomes a polytrophic \( r \). Accurate approximations to the polytrophic \( r \) have been developed (Hamdan, 1971; Lancaster & Hamdan, 1964; Olsson, 1979a; Pearson & Pearson, 1922; Ritchie-Scott, 1918), but each is even more complicated than the methods for approximating the tetrachoric \( r \). For the concurrent rating task, the best method for estimating the polytrophic \( r \) may be the maximum likelihood procedure developed by Olsson (1979a). In addition to producing an estimate of \( Q_{Ap,Bp}(x,y) \), this procedure yields estimates of each criterion, a variance-covariance matrix of the estimates, and a measure of goodness-of-fit. It uses Kirk’s (1973) algorithm as well as programs found in the IMSL (1975) library. However, it turns out that even a simple generalization of the cosine-pi approximation to the tetrachoric \( r \) provides much more accurate estimates of \( Q_{Ap,Bp}(x,y) \) than does the product moment correlation between ratings.

Suppose a \( k+1 \) point rating scale is employed in a \( 2 \times m \) concurrent rating task. Consider trials on which stimulus \( Ap,Bp \) is presented. Let \( X_i \) be the criterion on dimension \( x \) that is closest to the median of the \( Ap,Bp \) marginal perceptual distribution on dimension \( x \), and denote the rating given when \( X_i - 1 < x \leq X_i \) by \( L_i \). Let \( Y_j \) and \( J_i \) be the corresponding criterion and rating on dimension \( y \). Then

\[
|P_{Ap,Bp}(R_x \leq L) - .5| \\
\leq |P_{Ap,Bp}(R_x \leq i) - .5| \\
\text{for all } i \neq L
\]

and

\[
|P_{Ap,Bp}(R_y \leq J) - .5| \\
\leq |P_{Ap,Bp}(R_y \leq j) - .5| \\
\text{for all } j \neq J.
\]
and so \( I \) and \( J \) are easy to estimate. Next, define

\[
a_\ast = \hat{p}_{A_\ast B_\ast}(R_x > \hat{I}, R_y > \hat{J})
\]

\[
b_\ast = \hat{p}_{A_\ast B_\ast}(R_x \leq \hat{I}, R_y > \hat{J})
\]

\[
c_\ast = \hat{p}_{A_\ast B_\ast}(R_x > \hat{I}, R_y \leq \hat{J})
\]

and

\[
d_\ast = \hat{p}_{A_\ast B_\ast}(R_x \leq \hat{I}, R_y \leq \hat{J}).
\]

If \( a_\ast, b_\ast, c_\ast, \) and \( d_\ast \) are used in Equation 3 in place of \( a, b, c, \) and \( d \), the resulting estimate should be fairly accurate. The tetrachoric \( r \) was approximated in this fashion for each of the simulations reported earlier. It was found to provide consistently better estimates of \( q_{A_\ast B_\ast}(x, y) \) than the product moment correlation. In addition, I also compared this approximation to the tetrachoric \( r \) with one that also utilized \( a_\ast, b_\ast, c_\ast, \) and \( d_\ast \), but which was based on Kirk’s (1973) algorithm. For the simulations I conducted, Kirk’s method performed about as well as the cosine-pi approximation, and therefore I will limit discussion to the estimates obtained with the simpler cosine-pi formula.

For each simulation reported in Table 1, the (cosine-pi) tetrachoric \( r \) estimate is given in the rightmost column. An asterisk indicates that the estimated value is more than 2.5 standard errors away from the perceptual correlation. Note that \( r_{est} \) is relatively unaffected by criterion placement, at least as long as the lowest criterion on each dimension is below the mean and the highest criterion is above. As either the lowest criterion moves above the mean or the highest criterion moves below, it will become harder to guarantee a criterion close to the median, and so the accuracy of this approximation to the tetrachoric \( r \) will deteriorate. With data where such extreme criterion placement is suspected, a better strategy would be to select a more sophisticated estimate of \( q_{A_\ast B_\ast}(x, y) \) (e.g., Olsson’s, 1979a, maximum likelihood estimate of the polychoric \( r \)). In spite of its limitations, Table 1 makes it apparent that this approximation to the tetrachoric \( r \) provides clearly superior estimates of the degree of underlying perceptual independence than does the product moment correlation between ratings. Interestingly, similar statistical results have recently been found within the domain of factor analysis (Muthen & Kaplan, 1985; Olsson, 1979b).

For purposes of hypothesis testing, an estimate of the standard error of \( r_{est} \) is needed. Unfortunately, the general expression for estimating \( \sigma_{r_{est}} \) is too complicated to be practical. However, the expression simplifies greatly under the null hypothesis, \( H_0: q_{A_\ast B_\ast}(x, y) = 0 \). In this case (Guilford & Fruchter, 1973),

\[
\hat{\sigma}_{r_{est}} = \frac{\sqrt{(a_\ast + b_\ast)(a_\ast + c_\ast)(c_\ast + d_\ast)(b_\ast + d_\ast)}}{f_{A_\ast B_\ast}(X_\ast)f_{A_\ast B_\ast}(Y_\ast)\sqrt{n}},
\]

where \( n \) is the sample size. In the simulations reported in Table 1, the modal standard error was .04, although some of the standard errors were as large as .07. Therefore, fairly large sample sizes are needed to obtain accurate estimates of \( q_{A_\ast B_\ast}(x, y) \) using this technique. Again, if smaller standard errors are desired, more sophisticated techniques are available.

### TESTING FOR PERCEPTUAL SEPARABILITY

In addition to providing relatively simple yet accurate estimates of the degree of perceptual dependence, the concurrent rating task provides powerful tests of perceptual separability. As mentioned earlier, one consequence of perceptual separability is that the perceptual means, and also the means of the ratings, form a rectangle. This is a weak test, however, because even if the perceptual means form a rectangle, perceptual separability might fail (see, e.g., Ashby & Townsend, 1986). Luckily, a stronger test is possible.

From Equation 1 it is clear that perceptual separability holds, for example, only if

\[
F_{A_\ast B_1}(x) = F_{A_\ast B_2}(x)
\]

for all values of \( x \). Therefore, it must be true that if perceptual separability holds, then

\[
F_{A_\ast B_1}(X_i) = F_{A_\ast B_2}(X_i)
\]

for each criterion placement \( X_i \). Because of the equality

\[
F_{A_\ast B_\ast}(X) = P_{A_\ast B_\ast}(R_x \leq i),
\]

a strong consequence of perceptual separability is, therefore, that

\[
P_{A_\ast B_1}(R_x \leq i) = P_{A_\ast B_2}(R_x \leq i)
\]

for all values of \( i \) and \( p \)

\[\text{(4)}\]

and

\[
P_{A_1 B_\ast}(R_y \leq j) = P_{A_2 B_\ast}(R_y \leq j)
\]

for all values of \( j \) and \( q \).

\[\text{(5)}\]

With 2-point rating scales, these equalities reduce to the condition Ashby and Townsend (1986) called *marginal response invariance*, and so I will use the same terminology when referring to Equations 4 and 5. A test of marginal response invariance is stronger than checking for a rectangular configuration among the mean ratings, because marginal response invariance implies such a configuration, but a rectangular configuration among the mean ratings does not imply marginal response invariance.

Note that each term in Equations 4 and 5 is a marginal cumulative probability distribution function. For example, Equation 4 states that the distribution of ratings on component \( A \) does not depend on the level of component \( B \). This is a natural extension of the Equation 1 definition of perceptual separability, which states that the distribution of perceptual effects associated with component \( A \) does not depend on the level of component \( B \). Because Equations 4 and 5 each assert the equality of two cumula-
tive distribution functions, the null hypothesis of marginal response invariance is easily tested using the Kolmogorov-Smirnov statistic.

AN EMPIRICAL APPLICATION

As a demonstration of their empirical utility, I applied the techniques developed above to the data of 1 subject employed by Olzak (1986) and reported by Olzak and Wickens (1983). Olzak's experiment involved a 2 × 2 concurrent rating task with a 6-point rating scale. Her components were sine-wave gratings differing in spatial frequency (3 and 18 cycles/degree; cpd) and the two levels of each component were absent and present. Each of the four stimuli were presented 500 times.

Before beginning, it is important to recall that each of the techniques discussed in this article assumes decisional separability. For example, if decisional separability fails, there is no guarantee that the approximation to the tetrachoric $r$ developed above will provide a reasonable estimate of the perceptual correlation. Similarly, a violation of marginal response invariance could be caused by a failure of either perceptual separability or decisional separability (or both), and so caution must be exercised before concluding that the violation is due to a failure of perceptual separability. On the other hand, if marginal response invariance is satisfied, then perceptual separability and decisional separability are both strongly indicated. Therefore, when applying the techniques, it makes sense to test marginal response invariance first. If it is satisfied, then one's confidence in $r_{st}$ as an estimator of $\rho(x,y)$ is bolstered. If marginal response invariance fails, then the $r_{st}$ estimates of $\rho(x,y)$ may still be valid, but they should be interpreted with caution.

With this strategy in mind, the marginal cumulative rating distribution functions specified by marginal response invariance were estimated. These are shown in Figures 1 and 2. Figure 1 shows estimates of the marginals on the low-spatial-frequency (3 cpd) dimension, and Figure 2 shows estimates of the marginals on the high-spatial-frequency (18 cpd) dimension. The solid lines indicate that the relevant component (low frequency in Figure 1 and high frequency in Figure 2) was absent; the dashed lines indicate that it was present. Finally, plus indicates that the irrelevant component (high frequency in Figure 1 and low frequency in Figure 2) was absent; the circle indicates that it was present.

If perceptual separability and decisional separability both hold, the marginal cumulative rating distributions should not depend on whether the irrelevant component was absent or present. Thus, in Figures 1 and 2, the two solid lines should coincide, as should the two dashed lines. Visual inspection indicates that this condition is satisfied in Figure 2 but not in Figure 1. Kolmogorov-Smirnov tests support this hypothesis. For the Figure 1 data, the null hypothesis that the presence of the high-frequency component had no effect on the perception of the low-frequency component can be rejected with $p < .01$, both on low-frequency component-absent trials and on low-frequency component-present trials. In fact, Figure 1 indicates that the presence of the high-frequency component made the subject less confident that the low-frequency component was present. On the other hand, for the Figure 2 data, the null hypothesis that the presence of the low-frequency component had no effect on the perception of the high-frequency component could not be rejected, even with $p = .20$, either on high-frequency
component-absent trials or on high-frequency component-present trials.

The data therefore support the conclusion that both perceptual and decisional separability hold on the high-frequency dimension, but that one or both forms of separability fail on the low-frequency dimension. Although it is logically possible for decisional separability to hold on one dimension but not another because it is a response strategy, it seems likely that a subject who employs decisional separability on one dimension will employ it on others. A plausible interpretation of these results therefore is that decisional separability holds on both dimensions but that there exists an asymmetric perceptual interaction. The high-frequency channel affects the low-frequency channel, but the low-frequency channel has no effect on the high. This is the same conclusion that Olzak (1986) made after examining the estimates of the perceptual means, and it has also been supported by several other studies (e.g., Furchner, Thomas, & Cambell, 1977; Hirsch et al., 1982).

Next, to test for perceptual independence, several correlations were computed. Table 2 contains estimates of the product moment correlation between ratings, the approximation to the tetrachoric r described above, and its standard error. Note that, as in the simulations, the product moment correlation is of a consistently smaller magnitude than the tetrachoric r. This discrepancy is especially pronounced for the low-frequency component-only stimulus. In this case, the tetrachoric r estimate is more than twice the magnitude of the product moment correlation. In fact, on the basis of this product moment correlation, Olzak (1986) concluded that the underlying correlation between percepts was not significantly different from zero. However, Table 2 indicates that the estimate based on the tetrachoric r is more than four standard deviations from zero. Finally, note that in all four cases r_{het} is more than one standard deviation from zero, in three cases it is more than two standard deviations from zero, and in two it is more than four standard deviations from zero. It therefore appears that the strategy of using the product moment correlation between ratings as an estimate of \( q_{AB}(x,y) \) is seriously biased in favor of perceptual independence.

### REFERENCES


### NOTES

1. For simplicity, I assume the task requires the subject to use a \( k+1 \) point rating scale on each dimension. The results of this article easily generalize to tasks in which subjects use a \( k+1 \) point rating scale on dimension \( x \) and a \( k+1 \) point scale on dimension \( y \).

2. Although confidence ratings are typically assumed to lie only on an ordinal scale, the results derived in this article explicitly assume that the subjects employ the scale \( R = 1, 2, \ldots, k+1 \). Therefore, in a sense, the rating scales are absolute, since the only allowable transformation is the identity. As such, means, variances, and correlations are meaningful statistics.

3. Strictly speaking, these should be step functions. The estimates were presented in this fashion to facilitate visual comparison.

### APPENDIX

To prove Theorem 1, it is necessary to know the relation between the mean ratings and the underlying perceptual distribu-
lemmas. Lemma 1 details this relation. Although the relation between the rating variances and the perceptual distributions is not needed to prove Theorem 1, it is included in Lemma 1 for completeness.

**Lemma 1.** If decisional separability holds and the subject employs \(k+1\) ratings on each dimension in the \(2 \times m\) concurrent rating task, then the mean ratings on each dimension on trials when stimulus \(A_p B_q\) is presented are given by

\[
\mu_{A_p B_q}(R_x) = (k+1) - \sum_{i=1}^{k} F_{A_p B_q}(X_i)
\]

and

\[
\mu_{A_p B_q}(R_y) = (k+1) - \sum_{j=1}^{k} F_{A_p B_q}(Y_j)
\]

and the rating variances are given by

\[
\sigma^2_{A_p B_q}(R_x) = \sum_{i=1}^{k} (2(k-i) + 1) F_{A_p B_q}(X_i) - \left[ \sum_{j=1}^{k} F_{A_p B_q}(X_j) \right]^2
\]

and

\[
\sigma^2_{A_p B_q}(R_y) = \sum_{j=1}^{k} (2(k-j) + 1) F_{A_p B_q}(Y_j) - \left[ \sum_{j=1}^{k} F_{A_p B_q}(Y_j) \right]^2.
\]

**Proof of Lemma 1.** Only the expression for the mean ratings will be derived. A proof of the variance results is available from the author upon request.

To begin, note that with criteria at \(X_1, X_2, \ldots, X_k\)

\[
P_{A_p B_q}(R_x \leq i) = P_{A_p B_q}(x \leq X_i) = F_{A_p B_q}(X_i).
\]

Therefore,

\[
\mu_{A_p B_q}(R_x) = \sum_{i=1}^{k+1} i P_{A_p B_q}(R_x = i) = \sum_{i=1}^{k+1} i F_{A_p B_q}(X_i)
\]  \hspace{1cm} (A1)

\[
= \sum_{i=1}^{k+1} i P_{A_p B_q}(R_x = i) \sum_{m=1}^{i} 1
\]

\[
= \sum_{m=1}^{k+1} \sum_{i=m}^{k+1} \sum_{i=1}^{i} 1
\]

\[
= \sum_{m=1}^{k+1} i P_{A_p B_q}(R_x = i)
\]

\[
= \sum_{m=1}^{k+1} i P_{A_p B_q}(R_x = i) = \sum_{m=1}^{k+1} i F_{A_p B_q}(X_m)
\]  \hspace{1cm} (A2)

\[
= \sum_{m=1}^{k+1} [1 - P_{A_p B_q}(R_x < i)]
\]

\[
= 1 + \sum_{i=1}^{k} [1 - P_{A_p B_q}(R_x \leq i)]
\]

\[
= (k+1) - \sum_{i=1}^{k} F_{A_p B_q}(X_i). \Box
\]

A similar derivation yields the expression for \(\mu_{A_p B_q}(R_y)\).

**Proof of Theorem 1.** As shorthand, let \(\mu_{R_x}\) denote \(\mu_{A_p B_q}(R_x)\). Then

\[
\text{cov}_{A_p B_q}(R_x, R_y)
\]

\[
= \sum_{j=1}^{k+1} \sum_{i=1}^{k+1} ij P_{A_p B_q}(R_x = i, R_y = j) - \mu_{R_x} \mu_{R_y}
\]

\[
= \sum_{j=1}^{k+1} \sum_{i=1}^{k+1} \sum_{R_x \geq i, R_y \geq j} P_{A_p B_q}(R_x, R_y) - \mu_{R_x} \mu_{R_y}
\]

\[
= \sum_{j=1}^{k+1} \sum_{i=1}^{k+1} [1 - P_{A_p B_q}(R_x < i) - P_{A_p B_q}(R_y < j)]
\]

\[
+ P_{A_p B_q}(R_y < j) - \mu_{R_x} \mu_{R_y}
\]

\[
= (k+1)^2 - (k+1) \sum_{i=1}^{k} P_{A_p B_q}(R_x \leq i)
\]

\[
- (k+1) \sum_{j=1}^{k} P_{A_p B_q}(R_y \leq j)
\]

\[
+ \sum_{j=1}^{k} \sum_{i=1}^{k} P_{A_p B_q}(R_x \leq i, R_y \leq j) - \mu_{R_x} \mu_{R_y}
\]

\[
= (k+1)^2 - (k+1) \sum_{j=1}^{k} F_{A_p B_q}(X_j) - (k+1) \sum_{j=1}^{k} F_{A_p B_q}(Y_j)
\]

\[
+ \sum_{j=1}^{k} \sum_{i=1}^{k} F_{A_p B_q}(X_i, Y_j) - \mu_{R_x} \mu_{R_y}
\]

\[
= (k+1)^2 - (k+1) \sum_{j=1}^{k} F_{A_p B_q}(X_j) - (k+1) \sum_{j=1}^{k} F_{A_p B_q}(Y_j)
\]

\[
+ \sum_{j=1}^{k} \sum_{i=1}^{k} F_{A_p B_q}(X_i, Y_j) - \mu_{R_x} \mu_{R_y}
\]

\[
= (k+1)^2 - (k+1) \sum_{i=1}^{k} F_{A_p B_q}(X_i) - (k+1) \sum_{j=1}^{k} F_{A_p B_q}(Y_j)
\]

\[
+ \sum_{j=1}^{k} \sum_{i=1}^{k} F_{A_p B_q}(X_i, Y_j) - \mu_{R_x} \mu_{R_y}
\]

\[
= (k+1)^2 - (k+1) \sum_{i=1}^{k} F_{A_p B_q}(X_i) - (k+1) \sum_{j=1}^{k} F_{A_p B_q}(Y_j)
\]

\[
+ \sum_{j=1}^{k} \sum_{i=1}^{k} F_{A_p B_q}(X_i, Y_j) - \mu_{R_x} \mu_{R_y}
\]

\[
= \sum_{j=1}^{k+1} \sum_{i=1}^{k+1} i F_{A_p B_q}(X_i, Y_j) \hspace{1cm} \Box
\]

**Proof of Theorem 2.** The proof is sketched in the text immediately preceding the statement of the result. \(\Box\)

**Proof of Theorem 3.** Under the conditions of the theorem, the two tasks are functionally equivalent. With 2-point rating scales on each dimension, note that

\[
p_{11} = P_{A_p B_q}(R_x = 1, R_y = 1) = F_{A_p B_q}(X_1, Y_1)
\]

\[
p_{12} = P_{A_p B_q}(R_x = 1, R_y = 2) = F_{A_p B_q}(X_1, Y_2) - F_{A_p B_q}(X_1, Y_1)
\]

\[
p_{21} = P_{A_p B_q}(R_x = 2, R_y = 1) = F_{A_p B_q}(Y_1) - F_{A_p B_q}(X_1, Y_1)
\]

and

\[
p_{22} = P_{A_p B_q}(R_x = 2, R_y = 2) = 1 - F_{A_p B_q}(X_1, Y_1) + F_{A_p B_q}(X_1, Y_1)
\]

Now uncorrelated ratings occur if and only if

\[
F_{A_p B_q}(X_1, Y_1) = F_{A_p B_q}(X_1) F_{A_p B_q}(Y_1)
\]
Simplifying leads to
\[ p_{22} = p_{11} p_{12} + p_{11} p_{22} + p_{12} p_{22} + p_{12} p_{22} \]
\[ = (p_{11} + p_{22})(p_{12} + p_{22}), \]
which is sampling independence. Therefore, uncorrelated ratings implies sampling independence. Each step in this derivation is reversible, and so it follows that sampling independence implies uncorrelated ratings. □

(Manuscript received July 1, 1987; revision accepted for publication March 4, 1988).