

Varieties of Perceptual Independence

F. Gregory Ashby
Ohio State University

James T. Townsend
Purdue University

Several varieties of perceptual independence are investigated. These include sampling independence, dimensional orthogonality, stimulus separability and integrality, and performance parity. A general multivariate perceptual theory is developed, and a precise definition of perceptual independence is offered. Each of these related concepts is then examined within the framework of this theory, and their theoretical interrelationships are explicated. It is shown that none of the concepts are equivalent to perceptual independence but that if separability holds, then sampling independence is equivalent to perceptual independence. Several simple tests of separability are suggested that can be applied to the same data as sampling independence. Dimensional orthogonality is shown to test for independence only if some strong distributional assumptions are made about the perceptual effects of stimuli. Reaction time and information-based performance parity criteria are examined. The potential for empirically testing each of these concepts is discussed.

The stimuli of perception are many dimensional. Tones may vary in frequency, intensity, and duration. Individual characters in text may vary in size, orientation, shape, and the number of line segments they contain (among other things). Even the fruit that we eat may vary in size, shape, texture, sugar content, and the wavelength of light it reflects. A fundamentally important problem is to determine how these dimensions are combined in perceptual processing. Because of this, a notion central to almost all theories of perception, whether they are aimed at the visual, the auditory, or some other modality, is that of perceptual independence. At a macroscopic verbal level, there is agreement among most theoreticians about the meaning of this term. The components *A* and *B* of the two-dimensional stimulus *AB* are said to be independently perceived if the perception of each is in no way contingent on or interacts with the perception of the other, or somewhat more rigorously, if the probability of simultaneously perceiving both components *A* and *B* is equal to the probability of perceiving component *A* times the probability of perceiving component *B* (e.g., Garner & Morton, 1969; Green & Birdsall, 1978; Wandmacher, 1976).

Unfortunately, perceptions are not usually directly observable; instead they first pass through some decision process that uses the perceptions to select a response appropriate to the general experimental milieu. Decision or judgment processes therefore fundamentally alter direct perceptions (i.e., the early stages of perception), thereby making the above definition of perceptual independence difficult to test.

This article was begun while the first author was visiting Harvard University under Grant 606 7517/2 from the National Science Foundation. He is grateful to W. K. Estes for the opportunity the year provided. The second author received support from National Science Foundation Grant 8319377. Both authors would like to thank J. R. Busemeyer, G. R. Lockhead, R. C. MacCallum, N. A. Perrin, J. R. Pomerantz, and an anonymous reviewer for their helpful comments on the manuscript.

Correspondence concerning this article should be addressed to F. Gregory Ashby, who is now at the Department of Psychology, University of California, Santa Barbara, California 93106.

As a consequence, many different independence criteria have been suggested, and several terms have come to be used interchangeably with *independence*. For example, *zero correlation* is often assumed to signal independence (e.g., Garner & Morton, 1969), but as is well known in statistics, uncorrelated variables need not be independent. Similarly, the concepts of *dimensional orthogonality* and *independence* are often assumed equivalent (e.g., Garner & Morton, 1969; Green & Birdsall, 1978; MacCallum, 1976; Regan, 1982; Tanner, 1956; Tucker, 1972; for an exception see Green, Weber, & Duncan, 1977), particularly in the areas of multidimensional scaling and signal detection theory. In fact, the phrase *orthogonal input channels* is often used to connote perceptual independence. We argue below, however, that orthogonality of perceptual dimensions does not imply perceptual independence, at least as it is commonly interpreted. Still other terms that appear to be related to perceptual independence are *stimulus separability* (Shepard, 1964), *performance parity* (Garner & Morton, 1969), and from the letter-recognition literature, *feature sampling independence* (Townsend, Hu, & Ashby, 1980).

The many different tests and meanings of perceptual independence have led to confusion in the literature. We feel this is partly because these concepts have often been studied atheoretically, or at least informally. For example, many of the definitions in the literature are operational. We believe that the ambiguity can be eliminated only by studying perceptual independence and its related concepts within a rigorous theoretical framework. We therefore develop a formal theory within which the various notions, terms, and mechanisms may be interrelated. There are a number of benefits that ensue from such a theory. The following three stand out.

First, several conditions previously associated with perceptual independence turn out to be distinct when placed within a global theoretical framework. Second, a number of the theorems we present go well beyond the taxonomy in pointing to interesting empirical consequences. For instance, new tests of independence are derived and conditions are linked together to provide considerably stronger tests of perceptual independence and its related properties than have previously been available. Finally, we believe

the present theory suggests a true process interpretation of what have usually been static, measurement-focused concepts.

Within the context of this theory, we will offer precise definitions of perceptual independence, sampling independence, dimensional orthogonality, stimulus separability, and performance parity, and we will present several results that detail exact relationships between these concepts. Table 1 indicates which of the definitions and theorems of the article are relevant to each concept. We will show that none alone provides a perfect test of perceptual independence, that is, one that states conditions that are necessary and sufficient for it to hold. However, we demonstrate that if certain conditions are conjoined, perfect tests are possible. Even in other cases, strong, if not perfect, tests may be feasible.

A critical attribute of a theory of perceptual independence is that it have a separate structure devoted to both perceptual and decisional processes. This feature has been missing from most preceding treatments. Perceptual independence is, by definition, a perceptual phenomenon, and we shall therefore wish to define it as an attribute of the perceptual system.¹ On the other hand, we shall see that several observable tests of independence make quite strong assumptions about decision processes, and so an accurate model of both of these subsystems is needed.

One of the most successful psychophysical models at separating perceptual from decisional effects is signal detection theory (Green & Swets, 1966; Peterson, Birdsall, & Fox, 1954; Tanner & Swets, 1954). The theory of perceptual behavior we develop may be viewed as a substantial generalization of signal detection theory, which can account for experiments with stimuli composed of two or more components and with any of a wide variety of response instructions. This will serve several purposes. First, it introduces a general and powerful perceptual theory. Second, it suggests a natural yet rigorous definition of perceptual independence and allows us to examine the various tests of independence that have been suggested and that are being used in the literature (such as dimensional orthogonality). It also allows us to explicate the theoretical interrelationships among many independence-related concepts and to explore their potential for empirical testability. Finally, and very importantly, it allows us to unify the detection and recognition literature with the perceptual independence literature. Historically, with the exception of the orthogonality concept, these have been largely unrelated.

Although one section, late in the article, is devoted to reaction time, most of the results we derive are in terms of accuracy

Table 2
Important Variables

Variable	Definition
Stimulus X, Y $A_i B_j$	Physical dimensions Compound stimulus constructed from component A at level i and component B at level j .
Perceptual x, y $f_{A_i B_j}(x, y)$ $g_{A_i B_j}(x), g_{A_i B_j}(y)$	Perceptual dimensions Distribution of perceptual effects elicited by presentation of stimulus $A_i B_j$ Marginal distributions of the perceptual effects on each perceptual dimension elicited by presentation of stimulus $A_i B_j$
Decisional $r(x, y)$ $R_r(x, y)$	Within-trial response function that selects an appropriate response on the basis of the perceptual effect generated on each trial Global response function that describes performance over the course of many trials (e.g., $P(\text{correct}), d', \text{etc.}$)
Response a, b_j	Response associated with stimulus $A_i B_j$ in a complete identification experiment

measures. This is primarily because the derivation of reaction time predictions requires an even richer theoretical structure than the one developed here. Extra processing assumptions are needed (e.g., whether processing is serial or parallel, self-terminating or exhaustive, etc.), and these make the theory more controversial. Our goal here is to avoid controversy by interrelating the many independence-related concepts through as general and innocuous a theory as possible. Once this goal is accomplished, the deeper structure necessary to generate reaction time predictions can be added. Consistent qualitative agreement between the reaction time and the accuracy tests would validate these new assumptions.

We begin our development, in the second section, by introducing the general recognition theory as well as some important primitive concepts and the requisite notation. In the third section we present and then illustrate a precise definition of perceptual independence within the context of the theory. The next four sections examine a number of specific notions that are thought to be closely related to perceptual independence. These include sampling independence, dimensional orthogonality, stimulus separability, and performance parity, respectively. Definitions and theorems will establish their positions in the theory and give them empirical reference. Finally, we close with some conclusions and general comments.

The General Recognition Theory

In this section, we develop the general theory and the primitive concepts necessary for our investigations. For future reference, Table 2 contains summary definitions of the key variables.

¹ Although we emphasize the perceptual side of the system, we do not wish to categorize perceptual independence as necessarily "sensory," in the strict sense of the term. Certain higher order or top-down influences are not necessarily ruled out.

Table 1
Important Terms and the Definitions and Theorems in Which They Are Discussed

Term	Relevant definitions	Relevant theorems
Perceptual independence	1	1, 3, 4
Sampling independence	2	1
Dimensional orthogonality	(see Orthogonality section)	2, 3
Perceptual separability	3	4, 5, 6
Decisional separability	4	1, 4, 5, 6
Temporal separability	7	7
Performance parity	5, 6	—

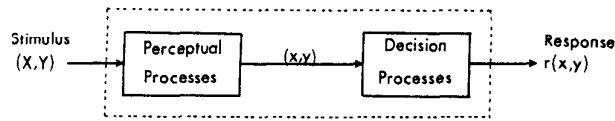


Figure 1. Schematic of the events occurring within a single trial.

In the experiments we will consider, the stimuli typically will be physically definable in terms of several separate stimulus components. Each of these may vary along some physical dimension that we denote as X , Y , and so forth. On any particular trial the experimenter will present the subject with some stimulus containing one or more components. The perceptual effects of these components will be denoted as x and y .

More explicit definitions of stimulus and stimulus component will depend on the processing level of interest (Taylor, 1976). For example, in the visual domain, if the interest is in the perceptual processing of separate line segments of individual alphanumeric characters, then a stimulus will contain a single character and the stimulus components will be the line segments making up that character. On the other hand, if the interest is in the perceptual processing of separate alphanumeric characters, then a stimulus will contain several characters and the stimulus components will be the characters making up the display.

If the components have natural unidimensional physical representations, such as might occur with line length or tone intensity, then in many cases we might expect the perceptual effects x and y to be unidimensional variables. Frequently, however, stimulus components will be more complex, with the result that x and y will themselves be multidimensional vectors. This would happen, for example, with stimuli created by factorially combining several levels of a geometric form with several levels of wavelength (so that the geometric forms reflect light of different wavelengths). To simplify the arguments, both mathematically and logically, we will treat the perceptual effects x and y as unidimensional variables throughout this article. None of the results we present nor the arguments we make are affected by this simplification.

From the perceptual effects of a particular stimulus, the subject's decision process must assign a response. We will denote this within-trial response function by $r(x, y)$. Of course, the nature of this function will depend on the type of experimental paradigm in use. For instance, in a yes-no detection task, r will assign either a yes or a no to every possible pair (x, y) . Figure 1 contains a schematic illustrating the use of this notation to describe events on a single experimental trial.

To account for observed variability in subject performance, we assume that presentation of the same stimulus (X, Y) does not always produce the same perceptual effect (x, y) or, in other words, that the perceptual effects x and y are random with cumulative distribution (or frequency) function $F(x, y)$ and joint density (or relative frequency) function (when it exists) $f(x, y)$. Figure 2 shows an example in which the stimulus ensemble contains two stimuli, S_1 and S_2 , each of which is constructed from the same two physical components. Figure 2a shows the distribution of the perceptual effects when the two stimuli are presented. Note that presentation of either stimulus could generate a perceptual effect anywhere in the space. The plane cutting through the two joint density functions describes the equal prob-

ability contours of the two distributions. Figure 2b is a view of this plane from above. Every perceptual effect associated with any point on either circle is equally likely to occur. Points inside the circle are more likely to occur and points outside the circle are less likely. By varying the height of the plane that cuts through the densities, a continuum of equal probability contours can be generated for each distribution. For many common distributions, however, including the examples we consider below, all contours from the same distribution have the same shape. Thus it suffices to examine just one. Such a contour of equal probability contains enough information about the distribution of perceptual effects to answer most of the questions we shall be interested in, and so we will have occasion to refer to them many times in this article.

In addition to a formal notation to describe events that occur within trials, we shall also have need to describe things on a more global level. On numerous occasions we will be interested in an average overall response measure such as probability correct or the signal detectability measure d' . Such a global response mea-

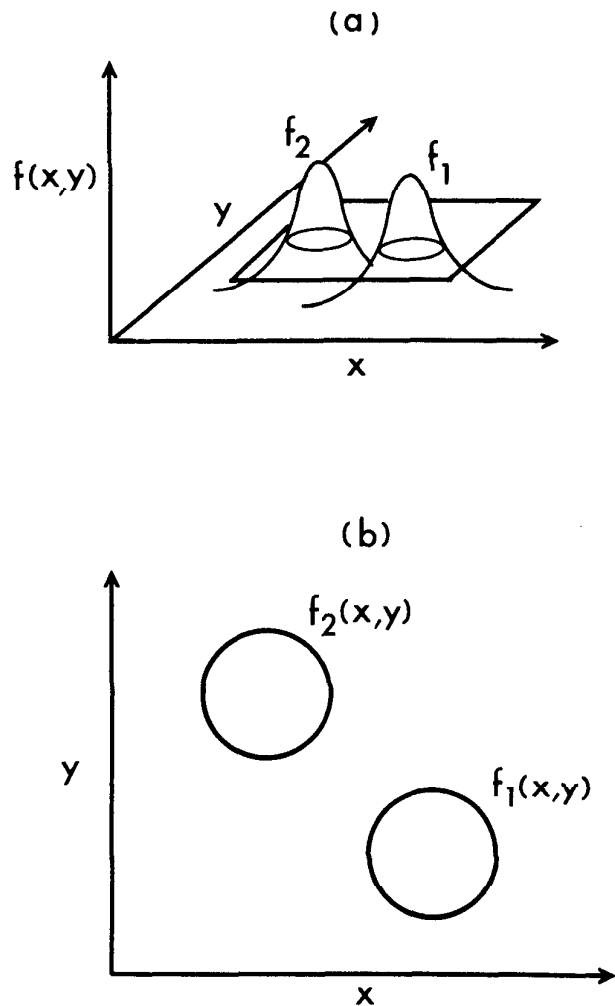


Figure 2. (a) Distribution of perceptual effects in an experiment with two stimuli, S_1 and S_2 , that are each composed of two stimulus components. (b) Contours of equal probability associated with Figure 2a.

sure clearly may depend on the distribution of the perceptual effects $F(x, y)$ and on the within-trial response function $r(x, y)$. To keep notation simple, however, we will denote it simply by $R_i(x, y)$.

As we will see, one of the more important experimental paradigms in this area is the complete identification experiment in which stimuli are constructed by factorially combining several levels of two or more stimulus components. With two levels of two components A and B , the four stimuli are therefore A_1B_1 , A_1B_2 , A_2B_1 , and A_2B_2 . The subject's task is to identify the stimulus on each trial. We will use the notation $a_i b_j$ to denote the response associated with stimulus $A_i B_j$. The four responses are therefore $a_1 b_1$, $a_1 b_2$, $a_2 b_1$, and $a_2 b_2$.

As a concrete example, suppose we are interested in the perceptual processing of the wavelength and size of geometric figures such as squares. First, (at least) two sizes (e.g., the length of a side) and two wavelengths are selected that are each perhaps one or two just noticeable differences (jnd) apart. Call the two sizes A_1 and A_2 and the two wavelengths B_1 and B_2 . Four stimuli are now constructed by factorial combination. For example, stimulus $A_1 B_2$ might be a square of width A_1 cm cut from a colored paper reflecting a wavelength of B_2 nm.

When the stimuli are constructed from two components, we will assume that the perceptual space and also the contours of equal probability are two-dimensional, as in Figure 3. For our wavelength-size example, this assumption is almost surely wrong; more dimensions will be needed to represent the perceptual effects. For example, three dimensions may be required to represent the perceptual effects of wavelength, which are known as hue. However, for our purposes, the important factor is whether any hue dimensions interact with any perceived size dimensions. Interactions between hue dimensions will not affect our conclusions about the perceptual independence of hue and perceived size. Therefore, as previously mentioned, without loss of generality we can consider the simplest case, in which the perceptual effect of each component is unidimensional.

For convenience, suppose x is the perceptual dimension associated with component A and y is the dimension associated with B , and denote the perceptual distribution (i.e., probability density function) corresponding to stimulus $A_i B_j$ by $f_{A_i B_j}(x, y)$.

The presentation of any stimulus induces a perceptual effect (x, y) . The subject's response function $r(x, y)$ uses this perceptual effect to select one of the four responses $a_1 b_1$, $a_1 b_2$, $a_2 b_1$, or $a_2 b_2$. An optimal response function² would select the response associated with the stimulus most likely to have produced that perceptual effect. This effectively divides up the (x, y) space into four regions, one associated with each response alternative. The response is determined by the region into which the perceptual effect (x, y) falls. Thus, a subject responding optimally never guesses (unless, for a given perceptual sample, there happen to be two or more responses equally likely to be correct). The dotted lines $y = y_0$ and $x = x_0$ in Figure 3 are the optimal decision bounds for the illustrated contours. In this special case, where the bounds are parallel to the coordinate axes, we call the intersection points x_0 and y_0 response or decision criteria. In effect, the subject selects x_0 as the criterion for deciding the level of component A and y_0 as the criterion for B . Thus, for instance, if a perceptual sample falls in the upper right-hand quadrant in Figure 3, the effect associated with component A exceeds the

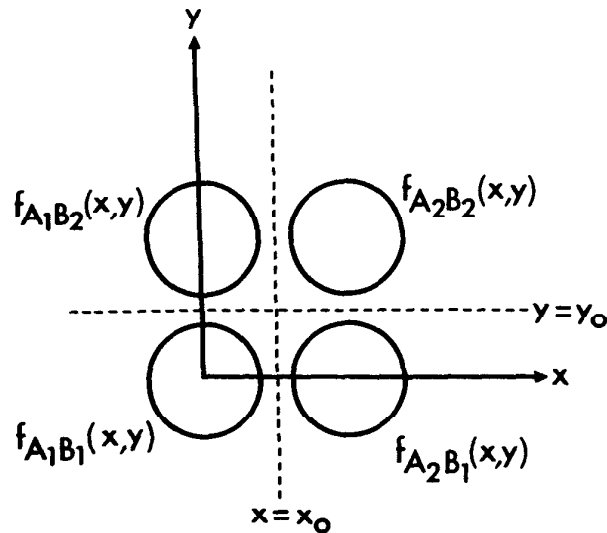


Figure 3. Contours of equal probability and decision bounds from the complete identification experiment with stimuli $A_1 B_1$, $A_1 B_2$, $A_2 B_1$, and $A_2 B_2$.

criterion associated with A (i.e., $x > x_0$) and the effect associated with B exceeds the B criterion (i.e., $y > y_0$) and so the subject responds $a_2 b_2$.

In its most general form, the general recognition theory makes no distributional assumptions. However, if we assume that the perceptual distributions are multivariate normal, the resulting special case, the *general Gaussian recognition model*,³ can be viewed as a multivariate extension of the (Case 1) model of Thurstone's law of categorical judgment (Thurstone, 1927; see also Hefner, 1958; Torgerson, 1958; Zinnes & MacKay, 1983) or, alternatively, as a multidimensional generalization of signal detection theory (see, e.g., Green & Swets, 1966; Townsend & Landon, 1983) applied to the problem of stimulus recognition. One of the first such applications was by Tanner (1956; see Wandell, 1982, for a later application), who assumed that the perceptual effects of noise and each of two auditory signals could be represented as bivariate normal distributions.

Every multivariate distribution has associated with it a covariance matrix that has a row and column assigned to each dimension over which the distribution is defined (e.g., Morrison, 1976). The entries on the main diagonal of the matrix are variances and the off-diagonal entries are covariances. Tanner assumed that the covariance matrix associated with each of his stimuli equaled the identity. In other words he assumed that the variance on each dimension was 1.0 and that each covariance

² There are many performance measures that a subject responding optimally could be maximizing. Many of these lead to the same decision rule. For our purposes a convenient measure is probability correct. Thus, throughout the article we assume that an optimal decision rule is one that maximizes response accuracy.

³ As is standard in the statistical literature, we use the terms *Gaussian* and *normal* interchangeably.

was zero. Under these assumptions, the ideal response function is a minimum distance classifier that computes the distance from a given sample to the mean of each perceptual distribution and delivers the response associated with the nearest mean (e.g., Ashby & Gott, 1985). Such a device performs better as the distance between the means of the various perceptual distributions increases and, in fact, the probability of a correct recognition increases monotonically with intermean distance.

Figure 3 illustrates the contours of equal probability postulated by this very simple generalization of Thurstone's Case 5 model for the complete identification experiment with the four stimuli A_1B_1 , A_1B_2 , A_2B_1 , and A_2B_2 . Because of its important historical and contemporary role, we will retain the multivariate normal assumption at this level. In what follows, however, we will have need of this model in its full generality and thus, unlike Tanner, we shall make no assumptions about the covariance matrices associated with each perceptual distribution. The contours of equal probability need not, therefore, always be circles. If the variances are unequal the contours are ellipses, and if the variates are correlated the contours are rotated.

Besides containing signal detection theory and Thurstone's law of categorical judgment as special cases, Ashby and Perrin (1985) showed that another important class of models that are a special case of the general Gaussian recognition model are the multidimensional scaling models that employ a Euclidean distance metric (e.g., Davison, 1983; Kruskal, 1964a, 1964b; Shepard, 1962a, 1962b; Torgerson, 1958; Tucker, 1972). This relationship will turn out to be important in our discussion of dimensional orthogonality because some multidimensional scaling models use orthogonality as a measure of perceptual independence.

Multidimensional scaling models are based on the assumption that stimulus similarity is inversely related to psychological distance. Empirical similarity judgments are used to construct a psychological space in which the perceptual effect of each stimulus is a point, with the property that points corresponding to similar stimuli are close together.

In the general Gaussian recognition model, the probability of confusing two stimuli is a function of the amount of overlap of the two perceptual distributions and is only indirectly related to the distance between the means of the two distributions. Therefore, under the assumption that confusability and similarity covary (as, e.g., in Luce, 1963; Shepard, 1964; Tversky & Gati, 1982), the general Gaussian recognition model suggests that similarity is functionally related to distributional overlap rather than to intermean distance (Ashby, 1982).

In the special case when all variates are uncorrelated and all variances are equal, the intermean distance is monotonic with distributional overlap. Increasing the distance between the means of two distributions will then decrease their overlap and vice versa. Ashby and Perrin (1985) show that under certain conditions the dissimilarity measure predicted by this general Gaussian recognition model will exactly equal the Euclidean distance between perceptual means.⁴ Thus, any predictions made about similarity ratings by Euclidean multidimensional scaling models will be equivalent to the predictions of some general Gaussian recognition model.

We are now in a position to examine in some detail the varieties of perceptual independence.

Perceptual Independence

Two arbitrary random variables W and Z are said to be *statistically independent* if the probability distribution of W does not depend on the value of Z (and vice versa) or, more formally, if and only if

$$f_{W|Z}(w|z) = f_W(w)$$

and

$$f_{Z|W}(z|w) = f_Z(z),$$

where, for example, $f_{W|Z}(w|z)$ is the conditional distribution (i.e., density) of W given Z . Because the joint density function of W and Z can always be written as

$$f_{WZ}(w, z) = f_{W|Z}(w|z)f_Z(z) = f_W(w)f_{Z|W}(z|w),$$

an equivalent definition is that the random variables W and Z are statistically independent if and only if

$$f_{WZ}(w, z) = f_W(w)f_Z(z). \quad (1)$$

When we say that the components of the stimulus (X, Y) are perceived independently, we mean that the perceptual effects of components X and Y do not interact, or more specifically that they are statistically independent. This leads to our first definition.

Definition 1: Perceptual independence of components A and B holds in stimulus A_iB_j if and only if the perceptual effects of A and B are statistically independent, that is, if and only if

$$f_{A_iB_j}(x, y) = g_{A_iB_j}(x)g_{A_iB_j}(y)$$

for all x and y and where, for example, $g_{A_iB_j}(x)$ is the marginal distribution on dimension x when the stimulus is A_iB_j .

This definition is a direct application of Equation 1. In our size-wavelength example, size and wavelength are perceived independently if the random perceptual effects associated with the size of the square are statistically independent of the perceptual effects associated with its wavelength. Note that this is a global level definition in the sense that it makes the perceptual independence of components A and B in stimulus A_iB_j a property of all A_iB_j trials.

It is extremely important to note, however, that the definition refers to only a single stimulus A_iB_j . Thus, perceptual independence may occur on trials when, say, A_1B_2 is the presented stimulus even if a perceptual dependence occurs on other trials (i.e., on A_1B_1 , A_2B_2 , or A_2B_1 trials). For example, when B is at level 2, the components of stimulus A_iB_2 might be perceived independently when A is at level 1 but not when A is at level 2. This might occur, for instance, if A and B are tones that differ in frequency. In this case, we expect them to be perceived independently when their frequency difference is large but not when it is small.

⁴ Besides assuming that each perceptual distribution has the same covariance matrix, these conditions include detailing a specific inverse relationship between similarity and dissimilarity and assuming that perceived similarity does not depend on stimulus context. Under these assumptions the general Gaussian recognition model makes predictions identical to Euclidean multidimensional scaling models, even those with weighted oblique dimensions (e.g., Tucker, 1972).

On a different issue, note that the definition is mute concerning whether the (marginal) perceptual distribution of, say, component A at level i , $g_{A_i}(x)$, is affected by the level of component B . In other words $g_{A_i}(x)$, may not necessarily equal $g_{A_i B_j}(x)$. It is very important: that this kind of notion be kept distinct from perceptual independence. The possible effect of the perception of one component on the level of the other will arise again in our discussion of separability and integrality.

In the special Gaussian case, it is well known that Definition 1 reduces to a condition on the covariance matrix (i.e., with variances on the diagonal and covariances elsewhere) associated with each perceptual distribution.

Lemma 1: In the special Gaussian case, perceptual independence of components A and B holds in stimulus $A_i B_j$ if and only if the perceptual effects of components A and B are uncorrelated on $A_i B_j$ trials.

As noted earlier, the contours of equal probability of a bivariate normal distribution are always ellipses or circles (as in Figure 3). The major and minor axes of the ellipse agree with the x and y axes of the perceptual space only when the two variates are uncorrelated (i.e., the covariance matrix is diagonal), which, according to Lemma 1, implies a statistical independence in the special Gaussian case.

A major difficulty involved in using Definition 1 to empirically test for perceptual independence is, of course, that in all current experimental paradigms the distribution of perceptual effects $f_{A_i B_j}(x, y)$ is unobservable. Both the within-trial and the global response functions that are observable represent a loss of information. For example, the within-trial response function r assigns the same response to many different points in the perceptual space, thus precluding the recovery of the distribution of perceptual effects from a subject's responses, and so a direct test of perceptual independence is not presently feasible.

Nevertheless, we believe that Definition 1 successfully captures the essence of what is meant in the literature by perceptual independence. Many conditions can be found in this literature that appear related to perceptual independence. We take these as attempts to define observable conditions from which an underlying perceptual independence might be inferred. Using Definition 1 and the general recognition theory, we will be able to determine the efficacy of some of these attempts. We begin with an examination of the concept of sampling independence.

Sampling Independence

A current popular theory of pattern recognition stresses the concept of mental feature representations that are compared with an analogous set of features stored in memory. This approach has seemed particularly viable in visual perception, no doubt in large part because of some important neurophysiological discoveries of the late 1950s and the early 1960s (e.g., Hubel & Wiesel, 1959, 1962; Lettvin, Maturana, McCulloch, & Pitts, 1959). Provided one is prepared to hypothesize which parts of a set of stimuli constitute the features, it becomes possible to test specific axioms on which feature-analytic models of pattern recognition are based (e.g., Townsend & Ashby, 1982; Wandmacher, 1976). Of particular interest here are tests of whether

basic features are perceived or sampled independently of one another.

In an attempt to test the assumption of feature sampling independence directly, Townsend, Hu, and Ashby (1980, 1981) conducted a complete identification experiment in which all possible combinations of a horizontal and a vertical line segment were tachistoscopically presented to each subject. For convenience, call the horizontal feature A and the vertical feature B and let the subscript 2 (1) denote trials on which the corresponding feature was (was not) presented. Then the four stimuli used were $A_1 B_1$, $A_1 B_2$, $A_2 B_1$, and $A_2 B_2$. Now consider the $A_2 B_2$ trials when both features A and B were presented. Feature sampling independence states that the probability that both features are sampled is equal to the probability that feature A is sampled times the probability that feature B is sampled. Let a and b denote the events that features A and B , respectively, are reported, and assume that a feature is reported whenever it is sampled. Then sampling independence is equivalent to

$$P(a_2 b_2 | A_2 B_2) = P(A \text{ is sampled} | A_2 B_2) \\ \times P(B \text{ is sampled} | A_2 B_2).$$

Now when A and B are presented, feature A is sampled when A alone is reported or when both A and B are reported. Therefore, sampling independence can be written as

$$P(a_2 b_2 | A_2 B_2) = [P(a_2 b_1 | A_2 B_2) + P(a_2 b_2 | A_2 B_2)] \\ \times [P(a_1 b_2 | A_2 B_2) + P(a_2 b_2 | A_2 B_2)]. \quad (2)$$

Stimuli of the sort used by Townsend et al., in which the two levels of each component are presence and absence, appear to differ fundamentally from the continuous valued components we have so far considered. Physical components that are either present or absent have been called *features* (Garner, 1978; Gibson, 1969), whereas components that can take on any value in a continuum have been called *dimensions* (e.g., frequency or intensity of a tone). How can features be represented in the general recognition theory? First of all, because a stimulus component is binary valued does not mean that the resulting perceptual effect is binary valued. Perceptual noise may generate a continuous valued perceptual effect. In fact, Ashby and Perrin (1985) successfully accounted for the results of the Townsend et al. experiment by making this very assumption. Second, even if the perceptual effects are binary valued, the general recognition theory is still an excellent candidate for describing the perceptual and decisional processes involved. This is because, in its most general form, it carries no distributional assumptions. Thus, instead of employing multivariate normal distributions as in the general Gaussian recognition model, a variant of the general recognition theory could be constructed from a discrete multivariate distribution (i.e., one in which the marginals are binary valued Bernoulli distributions).

Note that in this discussion we emphasize the difference between stimulus features and dimensions and perceptual features and dimensions. One of the great problems in perception over the last several decades has been to develop a methodology for psychophysically identifying perceptual features. The general recognition theory suggests a solution to this problem that involves identifying features with the eigenvectors of the covariance

matrix associated with each perceptual distribution. A future article will elaborate and test the method.

Because sampling independence was originally proposed to test some popular assumptions of feature-analytic models of pattern recognition, the two levels of components A and B in Equation 2 were assumed to designate the presence or absence of a corresponding feature. Nothing in the derivation of Equation 2, though, prevents, say, A_1 and A_2 from referring to any arbitrary levels of component A . For example, the levels of A might refer to different line lengths, tone intensities, or letter sizes. Similarly, the condition is not logically restricted to A_2B_2 trials, and so we generalize the Townsend et al. (1980, 1981) definition of sampling independence to the following.

Definition 2: Consider the complete identification experiment with stimuli A_1B_1 , A_1B_2 , A_2B_1 , and A_2B_2 (and responses a_1b_1 , a_1b_2 , a_2b_1 , and a_2b_2). Sampling independence of components A and B in stimulus A_iB_j occurs if and only if⁵

$$P(a_2b_2|A_iB_j) = [P(a_2b_1|A_iB_j) + P(a_2b_2|A_iB_j)] \\ \times [P(a_1b_2|A_iB_j) + P(a_2b_2|A_iB_j)].$$

The following result establishes the relationship between sampling independence and perceptual independence. It will be noted that the second part of the theorem requires sampling independence to hold for differing decision criteria. By this we mean any experimental manipulation that induces the subject to change response criteria on the different components. For example, in Figure 3 a different criterion on component A will result in a different value of x_0 . An increase in x_0 indicates a more conservative response strategy. The observer requires stronger evidence before signaling the presence of component A (or level 2 of component A). One popular method of inducing such criterion changes is to manipulate response payoff (e.g., Green & Swets, 1966; Townsend & Ashby, 1982).

The proof of this and all other results of this article is given in the Appendix.

Theorem 1: Consider the complete identification experiment with stimuli A_1B_1 , A_1B_2 , A_2B_1 , and A_2B_2 :

1. Sampling independence of components A and B in stimulus A_iB_j occurs if perceptual independence holds on components A and B and if the decision bounds are parallel to the coordinate axes.
2. Perceptual independence of components A and B in stimulus A_iB_j occurs if sampling independence holds on components A and B for differing decision criteria and if the decision bounds are parallel to the coordinate axes.
3. If the decision bounds are not parallel to the coordinate axes, then sampling independence is logically unrelated to perceptual independence.

We discuss each part of the theorem in turn. First, Part 1 indicates that if perceptual independence occurs, then sampling independence is logically implied if the subject adopts decision bounds that are parallel to the coordinate axes. Figure 3 contains an example in which this condition holds along with perceptual independence (the latter because of the circular equal probability contours). Data collected from a subject described by the Figure 3 model will therefore exhibit sampling independence. It is im-

portant to note, however, that although the contours in this figure were generated from normal distributions, no distributional assumptions are needed to prove any parts of Theorem 1. Thus, if perceptual independence holds and if the decision bounds are parallel to the coordinate axes, sampling independence will result, no matter what the distribution of perceptual effects (and even if the subject is responding nonoptimally).

Part 2 of Theorem 1 indicates that a finding of sampling independence across conditions with differing decision criteria implies perceptual independence, again, however, only if the decision bounds are parallel to the coordinate axes. The stipulation that sampling independence must hold for differing decision criteria is necessary to ensure that it is a global property of the identification process and not some local aberrance caused by certain specific criterion placements. It is therefore similar to the requirement of perceptual independence that the equality in Definition 1 must hold for all values of x and y .

Parts 1 and 2 of Theorem 1 suggest that sampling independence could form the basis of a useful test of perceptual independence. Part 3 indicates that, by itself, however, it is not a perfect test, because in the absence of decision bounds that are parallel to the coordinate axes, sampling independence may hold even if perceptual independence does not, and vice versa.

Taken as a whole, Theorem 1 underscores the importance of finding a method to empirically test for decision bounds that are parallel to the coordinate axes. Such a method, used in conjunction with sampling independence, would form a powerful test of perceptual independence. Later in this article we will suggest several such tests and then apply them, together with sampling independence, to data from the complete identification experiment of Townsend et al. (1980).

Before concluding this section, it is of interest to note that Theorem 1 places no perceptual independence constraints on the perceptual distributions associated with the alternative stimuli (those other than A_iB_j). In fact, it places no constraints whatsoever. Thus, the perceptual independence of a pair of components in one stimulus configuration logically implies nothing about whether they will be perceived independently when in another configuration.

Orthogonality

As noted earlier, a concept that is often closely associated with independence is dimensional orthogonality. Dimensional orthogonality is not the same as statistical orthogonality. Two random variables W and Z are said to be statistically orthogonal if the expected value of their crossproduct is zero, that is, if

$$E(WZ) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} wzf_{WZ}(w, z)dw dz = 0.$$

A concept that seems related to statistical orthogonality is correlation. The random variables W and Z are uncorrelated if $E(WZ) = E(W) \times E(Z)$. The notions of correlation and co-

⁵ Each of the other conditional response probabilities could be decomposed in a similar fashion. For example, under conditions of sampling independence, $P(a_1b_2|A_iB_j) = [P(a_1b_1|A_iB_j) + P(a_1b_2|A_iB_j)] \times [P(a_1b_2|A_iB_j) + P(a_2b_2|A_iB_j)]$.

variance are related by a simple translation,⁶ and so zero correlation implies and is implied by zero covariance. Statistical independence implies zero correlation, but zero correlation does not imply statistical independence. As we saw in Lemma 1, however, an exception occurs when the random variables are (jointly) normally distributed. In this case, zero correlation implies statistical independence, and the two terms are therefore interchangeable. In general, however, this equivalence does not hold. Two random variables may be uncorrelated but not statistically independent.

It is also easy to see that statistical orthogonality is only weakly related to the notions of statistical independence and correlation. Random variables may be orthogonal but not independent or uncorrelated. Further, they may be independent (and so uncorrelated) but not orthogonal.⁷ For these reasons, statistical orthogonality has not been used to test perceptual independence, so we will not consider it further.

Perhaps a kind of orthogonality more familiar to psychologists and the kind most usually associated with perceptual independence is dimensional orthogonality, which occurs if the coordinate axes are perpendicular. The idea that orthogonal dimensions somehow indicate perceptual independence appears to have its roots in the separate traditions of signal detection theory (Tanner, 1956) and multidimensional scaling (e.g., Tucker, 1972). Even so, the logical antecedents seem to be the same in both cases. Because Tanner's work has historical precedence, we focus on it.

Tanner (1956) was interested in determining whether a given pair of tones S_A and S_B that differ in frequency are perceived independently when embedded in white noise N . His experiment involved three conditions: a yes-no detection task with signal S_A (stimuli were N and $N + S_A$), a yes-no detection task with signal S_B , and a complete identification experiment with stimuli $N + S_A$ and $N + S_B$.

With respect to testing for perceptual independence, we see that this design suffers from an initial liability. The two tones are never simultaneously presented (i.e., there is no stimulus $N + S_A + S_B$), and so only indirect information is available about whether their perceptual effects will interact. Using this kind of experimental design, perceptual independence can therefore only be tested if some extra assumptions are made.

Tanner assumed that the perceptual effects of noise and of each signal embedded in noise have bivariate normal distributions each with zero covariance. He also assumed that all variances are equal. Contours of equal probability predicted by this model are illustrated in Figure 4.

Fixing the mean of the N distribution at $(0, 0)$, Tanner assumed the means of the $N + S_A$ and the $N + S_B$ perceptual distributions to be $(0, d'_A)$ and $(d'_B, 0)$ respectively, that is, to be fixed on the two perceptual dimensions. His idea was that the degree of perceptual dependence between the two signals could be represented by the angle between these two dimensions. He assumed that oblique dimensions indicate a perceptual dependence and that orthogonal dimensions indicate a perceptual independence.

Tanner proposed estimating the value of the angle between the dimensions by using a standard signal detectability analysis to estimate the three d' values indicated in Figure 4. With these, the angle can be estimated from the Pythagorean relation

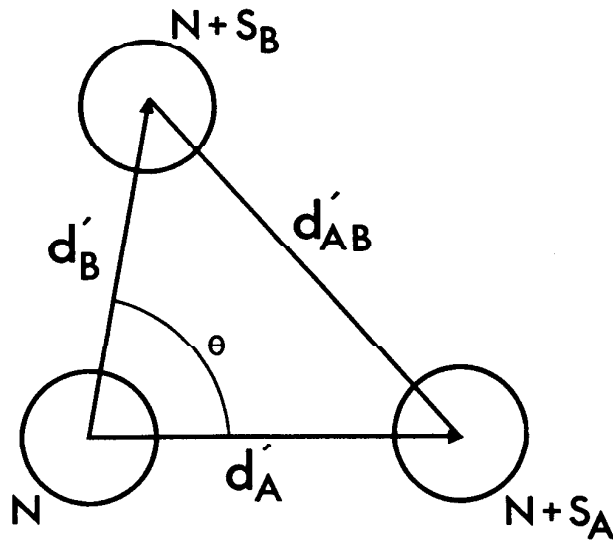


Figure 4. Contours of equal probability from the experiment of Tanner (1956).

$$(d'_{AB})^2 = (d'_A)^2 + (d'_B)^2 - 2d'_A d'_B \cos \theta. \quad (3)$$

Tanner used Equation 3 to estimate θ and found that, for tones varying widely in frequency, orthogonality was approximately satisfied.

Tucker (1972; see also Carroll & Chang, 1972) generalized standard Euclidean multidimensional scaling models in a similar fashion. The result, now referred to as the *general Euclidean scaling model* (Young & Hammer, in press), uses oblique (i.e., nonorthogonal) dimensions to model perceptual dependencies in the same way as Tanner's model.⁸

With their potentially oblique dimensions, these models appear to be quite different from those specified by the general recognition theory. However, as the next result shows, it is always possible to transform (linearly) a space with oblique dimensions to one whose dimensions are orthogonal, so that for every model in the class described by Tanner (1956) there exists an equivalent model with orthogonal dimensions. The transformation does change the variances and covariances of each distribution, but as the next result shows, the angle between the dimensions in

⁶ Specifically, the correlation between two random variables W and Z , ρ_{WZ} , is defined as $\rho_{WZ} = \text{cov}_{WZ} / \sigma_W \sigma_Z$, where cov_{WZ} is the covariance between W and Z and σ_W and σ_Z are the standard deviation of W and Z , respectively.

⁷ However, if they are uncorrelated then the new random variables $W - E(W)$ and $Z - E(Z)$ are orthogonal.

⁸ The general Euclidean scaling model represents the perceptual effects of stimuli as points in a Euclidean space. These points are assumed to have the same coordinates for all subjects, but the angle between dimensions and the relative importance of each dimension (expressed as a weight) may differ for each subject. Ashby and Perrin (1985) showed that this model is a special case of the general Gaussian recognition model (see footnote 6).

the oblique space completely determines the correlation between the variates in the orthogonal space.

Theorem 2: Tanner's (1956) signal detectability model is a special case of the general Gaussian recognition model in which all covariance matrices are equal, with correlation

$$\rho_{xy} = -\cos \theta,$$

where θ is the angle between dimensions in the oblique space.

This result shows that even with oblique dimensions, Tanner's signal detectability model is not as general as the general Gaussian recognition model (which has orthogonal dimensions), because in a two-dimensional space Tanner's model has only one correlation parameter (i.e., θ) whereas the general Gaussian recognition model has one associated with each stimulus. This fact is illustrated in Figure 5, which shows Tanner's model with oblique dimensions (Figure 5a) and the equivalent general Gaussian recognition model (Figure 5b). Unlike the Figure 5b model, however, the general Gaussian recognition model can have a different covariance matrix associated with each perceptual distribution (e.g., look ahead to Figure 6).

Theorem 2 provides a basis for the use of dimensional orthogonality as a test of perceptual independence, although there is one major caveat involved: It requires all covariance matrices in the true psychological space to be equal.

Ashby and Perrin (1985) show that a similar relationship exists between the general Gaussian recognition model and the general Euclidean scaling model. Not only is the general Euclidean scaling model a special case of the general Gaussian recognition model, but, like Tanner's model, when the perceptual space is two-dimensional, it has only one correlation parameter. Thus, the relationship between dimensional orthogonality and perceptual independence does not depend on which model (Tanner's or Tucker's) is used to test orthogonality. The next result summarizes this relationship.

Theorem 3: If the covariance matrices in the general Gaussian recognition model are equal, then dimensional orthogonality holds in Tanner's (1956) signal detectability model and in Tucker's (1972) general Euclidean scaling model if and only if components A and B are perceived independently in all stimulus configurations.

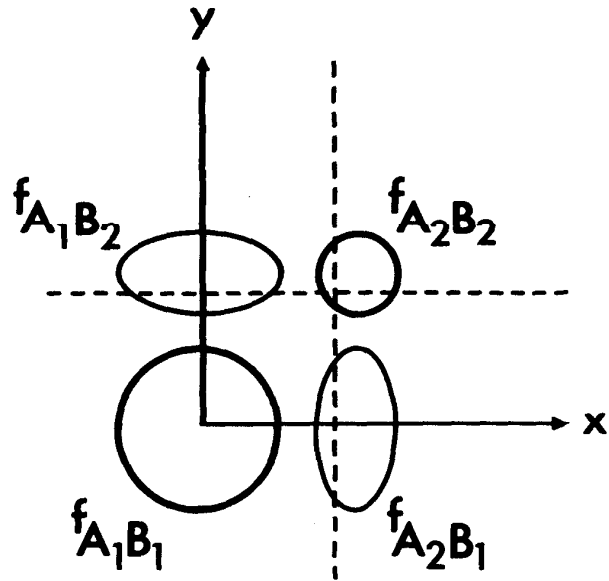


Figure 6. Contours of equal probability and the decision bounds of the general Gaussian recognition model as fitted by Ashby and Perrin (1985) to the data of Townsend, Hu, and Ashby (1980, 1981).

er's (1972) general Euclidean scaling model if and only if components A and B are perceived independently in all stimulus configurations.

To see why this theorem is true, note that in the two-dimensional case, if all covariance matrices are equal they must all have the same correlation ρ . Using the inverse transformation to the one specified in Theorem 2 will lead to a version of Tanner's model in which the angle between the dimensions is related to the correlation ρ via

$$\rho = -\cos \theta.$$

Thus, if dimensional orthogonality holds, θ will be 90° , and so the correlation will be zero, which implies perceptual independence under the conditions of the theorem (i.e., that the general Gaussian recognition model holds). Therefore, a good fit of the general Euclidean scaling model makes dimensional orthogonality an excellent test of perceptual independence.⁹

On the other hand, a good fit is not guaranteed, even if normality is satisfied. A good way to illustrate this fact is to consider a specific example. In particular, consider the previously discussed complete identification experiment of Townsend et al. (1980, 1981) in which all possible combinations of a vertical and horizontal line segment were tachistoscopically presented to subjects.

Ashby and Perrin (1985) fitted the general Gaussian recognition model to these data (under some constraints) and for all subjects obtained a solution similar to the one illustrated in Figure 6. Because the four contours of equal probability are different, the covariance matrices associated with the four stimuli all differ.

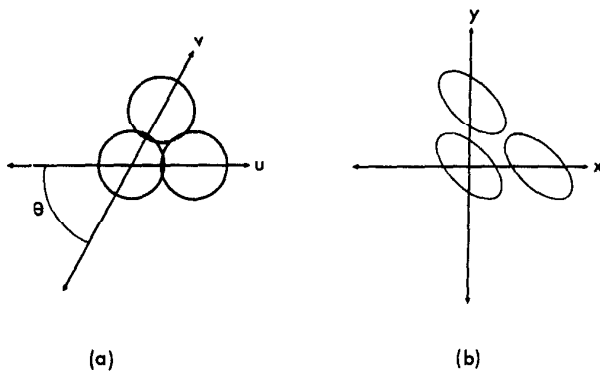


Figure 5. Contours of equal probability from an experiment with three stimuli. (Figure 5a is the representation as conceived by Tanner, 1956, or by the general Euclidean scaling model; Figure 5b shows the representation of the equivalent general Gaussian recognition model.)

⁹ The validity of Tanner's model cannot be tested with his experimental design.

The conditions of Theorem 3 are therefore not met, even though our definition of perceptual independence is satisfied in this experiment within each stimulus configuration.

The problem is that, because of the different covariance matrices, the general Euclidean scaling model will not be able to provide an adequate account of similarity data generated from the Figure 6 model. This conclusion is supported by the fact that Ashby and Perrin (1985) found strong evidence that each of the distance axioms was violated in these data, indicating that no distance-based similarity model is appropriate. Thus, in this case, the orthogonality criterion can not validly be used to make inferences about perceptual independence.

An even stronger test of perceptual independence is obtained if dimensional orthogonality is tested along with sampling independence. Because the two tests are not equivalent, using them together will provide more information than can be obtained from either alone. For example, if both tests are passed, perceptual independence is strongly supported, whereas the failure of both is strong evidence against perceptual independence.

Separability Versus Integrality

Another issue that has come to be closely associated with perceptual independence is that of perceptual *separability* versus *integrality*. According to Garner and Morton (1969), stimulus components are perceptually separable if they "act separately in the organism and thus can go on independently of each other" (p. 236), whereas integral stimulus components are those which "join one another such that it is extremely difficult for S (the subject) to take note of one . . . without at the same time taking note of the other" (Lockhead, 1966, p. 97).

Separable stimulus components can be separately attended to whereas the components of integral stimuli cannot. However, it may be possible to selectively attend to individual stimulus components, but only imperfectly or with some difficulty. Such stimulus components are neither truly separable nor integral. Separability and integrality may thus form two ends of a continuum (Garner, 1974). As we will see, testing between these two notions is complicated by the fact that it is logically possible for a perceiving system to be capable of activating either an integral or a separable strategy.

Operationally, the concepts of separability and integrality have been defined in several ways. The two we consider differ according to the nature of the experimental task to which they are applied. The first is based on what Posner (1964) calls a filtering task (e.g., Garner, 1974; Garner & Felfoldy, 1970).

Operational Definition A: With separable stimulus components, performance on a task that demands a response based on a single component is unaffected by the level of other irrelevant components. With integral components, varying the level of irrelevant components degrades performance.

This is called a filtering task because a plausible strategy is for the subject to filter out all information about the irrelevant components. With separable components this should be easy, and so no interference should be produced by variation along the irrelevant dimension. With integral stimuli, efficient filtering should not be possible and interference should therefore result.

In the special case when their only levels are presence and absence (i.e., features in Garner's, 1974, terminology), both components will be presented on A_2B_2 trials, but on A_1B_2 and A_2B_1 trials only a single component will be presented. In this case, Operational Definition A is closely related to the notion of unlimited capacity (e.g., Townsend & Ashby, 1983), which, in the reaction time literature, is the property stipulating that the performance (e.g., processing time) on an individual stimulus component does not depend on the total number of components to be processed. One difference is that with unlimited capacity, performance must be the same on all components, whereas with separability only one component need be involved. If all three of the stimulus components A , B , and C pass the separability test, then we can also say that the system is unlimited capacity. However, it is logically possible that component A could pass this separability test but not components B and C . Thus, component A can be separated from B and C , but B cannot be separated from C , and thus B and C are integral. In this case, the system is not unlimited capacity.

The second operational definition of the separability-integrality bifurcation is somewhat more problematic than Definition A (Garner, 1974; Garner & Felfoldy, 1970; Lockhead, 1972; Monahan & Lockhead, 1977).

Operational Definition B: Consider a task with stimuli A_1B_1 , A_2B_2 , and either A_1B_2 or A_2B_1 in which the subject is to respond yes whenever a component is at Level 2. If components A and B are separable, then performance on A_2B_2 trials will be at the same level as the performance on A_1B_2 or A_2B_1 trials. If A and B are integral, performance will be better on A_2B_2 trials than on A_1B_2 or A_2B_1 trials.

The idea here is that with integral components, the stimulus is perceived as a Gestalt and so the subject simultaneously processes all dimensions. On trials when both components convey information (A_2B_2 trials), performance should therefore be improved. When the components are separable, they can be separately attended to. If the subject is to respond yes when B is at level 2, all available attention will be focused on component B and redundant information in component A will never be processed. The argument with integral components seems reasonable, but the requirement that performance not be improved for separable components is unduly severe.

The problem is that an ideal observer would allocate some attention to both components because doing so improves performance. For example, consider an auditory signal detection task in which the two stimuli are a noise burst (N) and a pure tone embedded in noise (SN). Suppose that on some trials only one of these stimuli is presented but that on the others the two stimuli are either both N or both SN . It is well known that under these conditions an ideal observer will perform better on simultaneous trials. In fact, if performance is measured by the signal detectability parameter d' , then (Green & Swets, 1966)

$$d'_{\text{simultaneous}} = \sqrt{2}d'_{\text{separate}}$$

In contradiction to Operational Definition B, we can clearly construct such a device that separately attends to the two stimuli on simultaneous trials. Indeed, this would be the natural way to

proceed. Later in this article, we will see several more examples of a redundancy gain with separable components.

In fact, the only way that redundant information will not improve performance is either if no attention is allocated to the redundant information, that is, if it is ignored, or if some sort of mutual inhibition occurs, which we refer to as *negative integrality*. At the other end of the continuum, neither should there be any mutual excitation between separable components, and so redundant information should not make performance too much better.

As presently stated, this definition of separability is quite inexact and, hence, difficult to apply empirically. Exactly how much performance must be improved with separable components to make it better but not too much better will depend on one's choice of performance measure. In a later section we will examine these limits in the case when reaction time is used to measure performance.

To relate separability to the other concepts so far discussed, we need to first recast it into the language of the general recognition theory. Let us begin with Operational Definition A. Although instructions to subjects will differ, note that this separability criterion utilizes the same stimulus ensemble as the complete identification experiment discussed earlier (with stimuli A_1B_1 , A_1B_2 , A_2B_1 , and A_2B_2). Instead of asking for an identification response, subjects are now asked to respond yes whenever component A is at level 2. According to Definition A, the components A and B are separable if the subject's performance on component A is unaffected by the level of B .

In the general recognition theory, performance is a function of both perceptual and decisional processes. It is possible that one of these processes could be separately attending to the individual components, but not the other. Thus, separability could be defined at either the perceptual or the decisional level.

At the perceptual level, Definition A suggests that if the two components are separable, then the perceptual representation of component A should not depend on the level of B . For example, if size and hue are perceptually separable, the perceptual effect of a particular wavelength should not depend on the size of the test stimulus. Similarly, perceived stimulus size should not be affected by wavelength. In the general recognition theory, details of the perceptual representation of components A and B are contained in the joint probability density function $f(x, y)$. The perceptual representation of component A can be obtained by summing (i.e., integrating) this function over all possible perceptual values of component B . Thus, in the general recognition theory, the perceptual representation of component A equals the marginal distribution of component A ,

$$g(x) = \int_{-\infty}^{\infty} f(x, y) dy.$$

Following Operational Definition A, the components A and B are perceptually separable if this marginal distribution of A does not depend on the level of B , that is, if

$$g_{A_1B_1}(x) = g_{A_1B_2}(x),$$

for $i = 1, 2$.

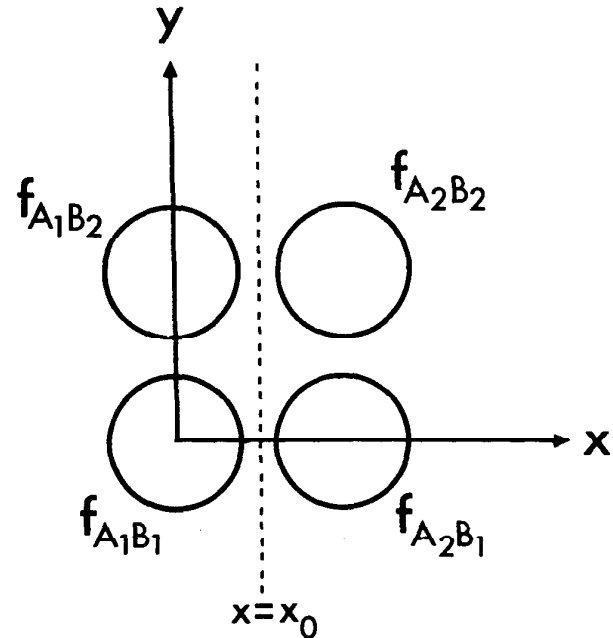


Figure 7. Contours of equal probability and the decision bound for a subject decisionally separating stimulus components A and B in the task of Operational Definition A.

At the decisional level, we interpret Definition A as implying that, under separability, the decision rule the subject selects for component A does not depend on the level of B . Another way of saying this is that the subject's criterion for reporting A (i.e., for saying yes) does not depend on the level of B . In the general recognition theory, presentation of stimulus A_iB_j causes a perceptual value (x, y) to be registered somewhere in the perceptual space. Because perceptual dimension x corresponds to component A , a large value of x is strong evidence that A was at level 2 in the presented stimulus, and a small value is evidence that it was at level 1. Thus, in this task the subject might set a criterion x_0 on x such that the perceptual value (x, y) is recognized as containing level 2 of component A if and only if $x > x_0$. Following Operational Definition A, the components A and B are decisionally separable if this criterion does not depend on the value of y , that is, on the magnitude of the perception associated with B .

An example illustrating this type of decision rule is shown in Figure 7. Note that the decision bound is parallel to the y axis. The subject responds yes to any stimulus that generates a perceptual value falling in the region to the right of the boundary and otherwise responds no. Under this rule of classification, the decision as to which level of component A is contained in the stimulus depends only on the magnitude of the perception along the dimension associated with component A .

One possible criticism of these two definitions is that subjects are never asked to respond to component B . Before concluding that components A and B are separable, we may therefore want to run another condition in which subjects are told to ignore A and respond yes only if the stimulus contains component B at

level 2. In this case perceptual separability occurs if the marginal distribution of B does not depend on the level of A and decisional separability occurs if the decision bound is parallel to the x axis.

Observe now that the conjoining of these two conditions produces a task in which subjects are told to simultaneously report whether component A is at level 2 and whether the level of component B is 2. The responses available to the subject are therefore (a) both A and B are at level 2 (A_2B_2), (b) A is at level 2 but B is not (A_2B_1), (c) A is not at level 2 but B is (A_1B_2), and (d) neither A nor B is at level 2 (A_1B_1). These are precisely the responses available in a complete identification experiment. This reasoning, together with our perceptual and decisional interpretations of Operational Definition A, leads to the following definitions.

Definition 3: Consider the complete identification experiment with stimuli A_1B_1 , A_1B_2 , A_2B_1 , and A_2B_2 . The components A and B are perceptually separable if the perceptual effect of one component does not depend on the level of the other, that is, if

$$g_{A_1B_1}(x) = g_{A_1B_2}(x),$$

for $i = 1, 2$, and

$$g_{A_1B_1}(y) = g_{A_2B_1}(y),$$

for $j = 1, 2$.

Definition 4: Consider the complete identification experiment with stimuli A_1B_1 , A_1B_2 , A_2B_1 , and A_2B_2 . The components A and B are decisionally separable if the decision about one component does not depend on the level of the other, that is, if the decision bounds in the general recognition theory are parallel to the coordinate axes.

Note that this definition of decisional separability was exactly the condition in Theorem 1 that was necessary to relate perceptual independence and sampling independence. Perceptual separability was never mentioned. Thus, an alternative way of stating the first two parts of Theorem 1 is that if decisional separability holds, then perceptual independence is equivalent to sampling independence (when the latter holds for differing decision criteria, cf. Theorem 1). A more powerful test of perceptual independence therefore results if sampling independence is used in conjunction with a test of decisional separability. Because of Theorem 1, perceptual independence is logically implied if both of these conditions are satisfied (for differing decision criteria). Data in which decisional separability holds but sampling independence does not, implies a perceptual dependence. On the other hand, if decisional separability is not found, then Theorem 1 indicates that sampling independence is logically unrelated to perceptual independence.

In addition to their Theorem 1 relationship through sampling independence, it turns out that separability and perceptual independence can be related in other ways. In fact, as the next theorem shows, under certain conditions perceptual and decisional separability imply perceptual independence.

Theorem 4. Consider the complete identification experiment with stimuli A_1B_1 , A_1B_2 , A_2B_1 , and A_2B_2 . If the following three conditions hold, then perceptual separability and decisional separability together imply the perceptual independence of components A and B within each stimulus configuration. Let μ_{ij} be the perceptual mean vector associated with stimulus A_iB_j .

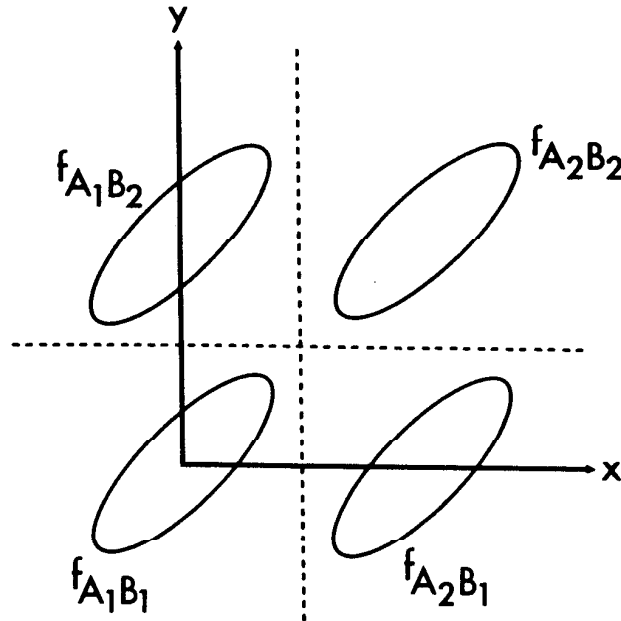


Figure 8. Contours of equal probability from a complete identification experiment in which perceptual and decisional separability hold but perceptual independence fails.

1. All perceptual representations are normally distributed (i.e., the general Gaussian recognition model holds).
2. No two μ_{ij} are equal.
3. The subject is responding optimally (i.e., maximizing probability correct).

To prove this important result, it is necessary to work backward from the decision rule a subject employs (i.e., decisional separability) to his or her perceptual representation of each stimulus. The key to this process is the optimality assumption. For a given set of perceptual distributions, there will, in general, be only one set of decision bounds that maximize response accuracy.

Theorem 4 provides a rationale for using separability to test for perceptual independence. If all of the conditions of the theorem are true, then perceptual and decisional separability logically imply perceptual independence. However, if one or more of Conditions 1-3 fail, then perceptual and decisional separability might hold in the absence of perceptual independence. This fact is illustrated in Figure 8, which shows contours of equal probability from a complete identification experiment in which both perceptual and decisional separability hold but perceptual independence fails. Decisional separability holds because the decision bounds are parallel to the coordinate axes. To see that perceptual separability holds, note that collapsing the perceptual distributions onto, say, the horizontal axis—that is, integrating out y —makes the A_2B_1 and the A_2B_2 marginal distributions identical—that is, $g_{A_2B_1}(x) = g_{A_2B_2}(x)$ —and it also makes the A_1B_1 and the A_1B_2 marginal distributions identical—that is, $g_{A_1B_1}(x) = g_{A_1B_2}(x)$. Because the same thing happens when we collapse onto the vertical dimension, perceptual separability holds.

Conditions 1 and 2 of Theorem 4 are also satisfied in Figure

Table 3

Results of Applying the Theorem 5 Separability Test to the Data of Townsend, Hu, and Ashby (1980, 1981)

Estimated conditional response probabilities	Gap condition subjects				Connected condition subjects			
	1	2	3	4	1	2	3	4
$\hat{P}(a_1b_2 A_2B_2) + \hat{P}(a_2b_2 A_2B_2)$	0.36	0.74	0.53	0.36	0.34	0.65	0.51	0.39
$\hat{P}(a_1b_2 A_1B_2) + \hat{P}(a_2b_2 A_1B_2)$	0.53	0.65	0.61	0.50	0.33	0.61	0.53	0.48
Z	-1.18	0.73	-0.58	-0.97	0.07	0.32	-0.14	-0.62
$\hat{P}(a_2b_1 A_2B_2) + \hat{P}(a_2b_2 A_2B_2)$	0.51	0.74	0.52	0.54	0.58	0.86	0.60	0.63
$\hat{P}(a_2b_1 A_2B_1) + \hat{P}(a_2b_2 A_2B_1)$	0.53	0.72	0.62	0.59	0.54	0.82	0.61	0.55
Z	-0.14	0.15	-0.73	-0.36	0.29	0.32	-0.07	0.58

Note. All Z tests were nonsignificant at the $\alpha = .25$ (two-tailed) level.

8, and so the only condition that prevents the theorem from applying is Condition 3. The subject described by Figure 8 is not responding optimally. With the perceptual distributions of Figure 8, decisional separability adversely affects accuracy. Theorem 1 tells us that we may drop the Gaussian and optimality assumptions of Theorem 4 if sampling independence is in force along with separability, because the latter imply perceptual independence under any distributional conditions.

Because of Theorems 1 and 4, it becomes desirable to have some simple way of testing separability, preferably on the same data set that would be used in a test of sampling independence. One possibility is to extend Operational Definition A to the complete identification experiment. This strategy suggests comparing the probability of correctly recognizing, say, component A across different levels of component B. If these marginal response probabilities are invariant, Operational Definition A suggests that separability is supported. The next result shows that neither perceptual nor decisional separability alone is sufficient to imply this marginal response invariance but that both kinds of separability taken together do guarantee that Operational Definition A is satisfied.

Theorem 5: Consider the complete identification experiment with stimuli $A_1B_1, A_1B_2, A_2B_1,$ and A_2B_2 . If perceptual and decisional separability both hold for components A and B, then marginal response invariance holds, or, in other words, the probability of correctly recognizing one component does not depend on the level of the other. More specifically, for $i = 1, 2$,

$$P(a_1b_1|A_1B_1) + P(a_1b_2|A_1B_1) \stackrel{\text{marginal } a_1|b_1}{=} P(a_1b_1|A_1B_2) + P(a_1b_2|A_1B_2) \quad (4)$$

and for $j = 1, 2$,

$$P(a_1b_j|A_1B_j) + P(a_2b_j|A_1B_j) \stackrel{\text{marginal } a_1|b_j}{=} P(a_1b_j|A_2B_j) + P(a_2b_j|A_2B_j). \quad (5)$$

Neither kind of separability alone, however, implies marginal response invariance.

Marginal response invariance (i.e., Equations 4 and 5) can be used to test separability on the same confusion data needed to test sampling independence. Of course, to use Theorem 1, only decisional separability needs to be verified. Perceptual separability is irrelevant. Toward this end, note that the theorem does not provide a perfect test of decisional separability, because it also assumes perceptual separability. We believe, however, that because observable responses are a function of both perceptual and decisional processes, many observable separability tests will evince this same joint dependence on perceptual and decisional separability. As a consequence, we could define an overall, global kind of separability by conjoining Definitions 4 and 5. One argument against this approach, however, is that the two concepts are functionally autonomous.

The proof of Theorem 5 gives an example in which perceptual but not decisional separability holds and one in which decisional separability holds but not perceptual. It therefore demonstrates that these two kinds of separability are logically unrelated. Although we might expect human subjects to often simultaneously exhibit both kinds of separability, we feel that carefully distinguishing between the two will aid our understanding of this complex literature.

It should also be noted that marginal response invariance does not depend on perceptual independence. In fact, it can hold even if the stimulus components are perceived dependently.

Finally, it is important to note that the converse of Theorem 5 is not true. If one starts with the response probabilities and tries to make inferences back about separability, then care must be taken. If the probability of correctly recognizing one component does not depend on the level of the other, then separability is not logically implied. Of course, if this invariance was found to hold across a variety of experimental conditions and over a number of subjects, then separability would be strongly supported. This same comment can be made about Theorem 6 be-

low, in which another test of the two kinds of separability is suggested. It too does not guarantee perceptual and decisional separability in the sense that if the test is successful the two separabilities are not logically implied. However, the two tests, successfully applied to the same data set, would provide strong converging evidence in favor of perceptual and decisional separability. On the other hand, if the probability of correctly recognizing one component does depend on the level of the other, then it is impossible that both types of separability are true. This means that the test is strong in the sense of falsifiability, even for a single experiment. The same is certainly true for Theorem 6.

A very natural way to illustrate the use of Theorem 5 is to apply it to the data of Townsend et al. (1980, 1981). Recall that in this experiment four subjects were tachistoscopically shown all possible combinations of a horizontal (component A) and a vertical (component B) line segment. There were two conditions. On trials when the segments were both presented, they were either physically connected (connected condition) or separated by a gap (gap condition). Townsend et al. (1980, 1981) found that sampling independence held in this data, and so, because of Theorem 1, a test of separability is especially important.

Table 3 contains the estimated probabilities needed to test marginal response invariance and the corresponding Z statistic on their difference. Because Townsend et al. (1980, 1981) only tested sampling independence on A_2B_2 trials, we fixed $i = 2$ in Equation 4 and $j = 2$ in Equation 5. As can be seen, none of these differences are significant; hence, the null hypothesis of perceptual and decisional separability cannot be rejected for any of the subjects in either condition.

The notion that subjects are separating the stimulus components in this task is also supported by the general Gaussian recognition model fits to this data reported by Ashby and Perrin (1985). In an earlier section, we saw that very good fits were obtained with the model illustrated in Figure 6. Note that both perceptual and decisional separability hold in this case even though all covariance matrices are different.

These results greatly strengthen the argument of Townsend et al. (1980, 1981) that subjects independently perceived the vertical and horizontal line segments in this experiment.¹⁰ Together with the conclusions we drew when discussing whether the orthogonality criterion might be applied to these data, we have tentatively learned the following about the Townsend et al. experiment: (a) Subjects perceived the horizontal and vertical line segments independently, (b) they perceptually and decisionally separated the components, and (c) the covariance matrices of the perceptual distributions were not all equal. The more separate tests that are applied to the same data, the stronger the conclusions that can be reached.

In the remainder of this section, we examine the possibility of testing for perceptual and decisional separability in a task other than the complete identification experiment. Consider first the situation in Figure 9. Here we see the hypothetical recognition space from an experiment presenting only stimuli A_1B_2 and A_2B_1 . We see immediately that the decision rule adopted by the subject simultaneously uses information from both dimensions. In fact, the subject is simply comparing x with y , responding a_2b_1 if x is greater than y and otherwise responding a_1b_2 .

Even though the stimuli in this experiment and their perceptual

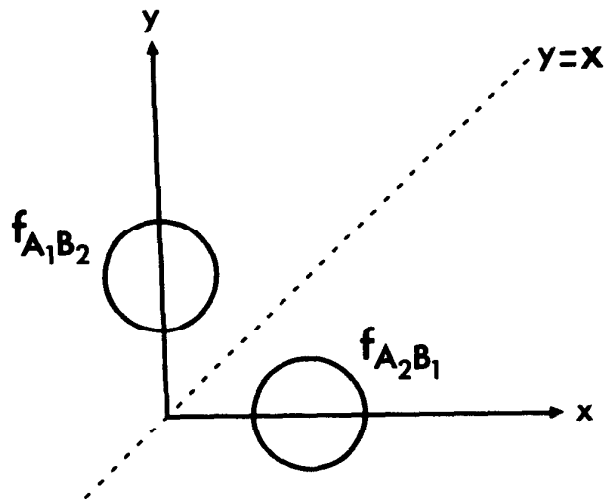


Figure 9. Contours of equal probability and the optimal decision bound from an experiment with stimuli A_1B_2 and A_2B_1 .

representations are the same as in Figure 3, notice that our decisional separability criterion (decision bounds that are parallel to the coordinate axes) is violated.¹¹ By deleting two stimuli (A_1B_1 and A_2B_2), our conclusions about decisional separability have been fundamentally altered. The subject may be separating the stimulus configurations but not into the experimenter-defined components. Notice that the stimulus components have not changed, only the task we are asking of our subjects. It makes sense that if subjects are able to construct decision bounds that separate the components A and B in an experiment with stimuli A_1B_1 , A_1B_2 , A_2B_1 , and A_2B_2 , they should be able to in an experiment with stimuli A_1B_2 and A_2B_1 . The problem is that if they were to do so in the latter experiment, their performance would be adversely affected. By using only stimuli A_1B_2 and A_2B_1 and by asking subjects to try their best to identify presented stimuli, we are asking them not to use a decision rule that separates the components. The fact that they do not separate components under these conditions is not evidence that they cannot.

To test separability we need to arrange the task so that subjects are not penalized for decisionally separating the stimulus components. Unlike the complete identification task, the experiment with stimuli A_1B_2 and A_2B_1 does not satisfy this criterion. In fact, neither does the task of Operational Definition B.

¹⁰ With more complex letterlike stimuli composed of equal-length vertical and horizontal line segments, Townsend and Ashby (1982) found strong evidence that sampling independence was violated. Thus, these conclusions apparently do not generalize across a wide variety of experimental conditions. We feel, however, that this is all the more reason to develop rigorous tests of concepts such as separability and perceptual independence.

¹¹ Perceptual separability cannot be tested. For example, to see if the perceptual effects of component A depend on the level of B , a given level of A (e.g., A_1) must be paired with different levels of B (e.g., both B_1 and B_2). This criterion is not met in an experiment with only stimuli A_1B_2 and A_2B_1 .

The Definition B task differs from the task of Definition A only in the stimulus ensemble it uses. In both tasks, subjects are to respond yes whenever a component is presented at level 2. The Definition A task uses all the stimuli of the complete identification experiment, whereas the B task uses A_1B_1 and A_2B_2 but only one of A_1B_2 and A_2B_1 .

As a concrete example, suppose subjects are told to respond yes whenever component A is at level 2 and that the stimulus ensemble contains A_1B_1 , A_2B_1 , and A_2B_2 . Subjects who decisionally separate components A and B in this task will construct a decision bound parallel to the y axis (as in Figure 7) and respond yes to any perceptual sample falling to the right of this bound and no to any sample falling to the left. In this case, no redundancy gain will occur (accuracy on A_2B_2 trials will equal accuracy on A_2B_1 trials), and so Operational Definition B is satisfied. Unfortunately, however, the decision bound of Figure 7 is not optimal under the response instructions of Operational Definition B. The optimal rule (i.e., maximizes response accuracy), which leads to a redundancy gain, is shown in Figure 10.

A subject, perfectly fulfilling the experimenter's instructions, will therefore not decisionally separate components A and B . This situation is similar to the one of Figure 9, and the conclusion we reached there is still valid. Decisional separability can be rejected only if the subject fails to exhibit it under conditions in which it is advantageous. In the task associated with Operational Definition B, decisional separability is not advantageous; hence, if the subject fails to exhibit it, no conclusions can be drawn.

In summary, we have examined two different operational definitions of separability and integrality commonly used in the literature. Of these, Operational Definition A appears to be on more solid theoretical grounds. However, translating Definition A into the language of the general recognition theory required distinguishing between perceptual and decisional separability. For example, it is conceivable that a subject perceptually separates a pair of stimulus components but then integrates the perceptual information somewhere in the decision-making process. In fact, this is the situation occurring in Figure 10. Using Definitions 3 and 4, it is possible, in principle, to distinguish this case from one in which the information integration occurs in both the perceptual and decisional processes. Theorems 1 and 4 indicate that perceptual and decisional separability are closely related to the other major concepts discussed in this article and that a powerful test of perceptual independence results if separability is tested in conjunction with sampling independence (e.g., via Theorem 5).

Performance Parity

In our discussions of both sampling independence and perceptual and decisional separability, we saw that the most powerful experimental paradigm is the complete identification experiment with stimuli A_1B_1 , A_1B_2 , A_2B_1 , and A_2B_2 . At the same time, however, we have encountered several paradigms that omit one or more of these stimuli. For example, Tanner (1956) failed to include stimulus A_2B_2 , and Operational Definition B excludes stimulus A_1B_2 (or A_2B_1).

Frequently, when a paradigm includes an impoverished stimulus ensemble, perceptual independence is tested indirectly by

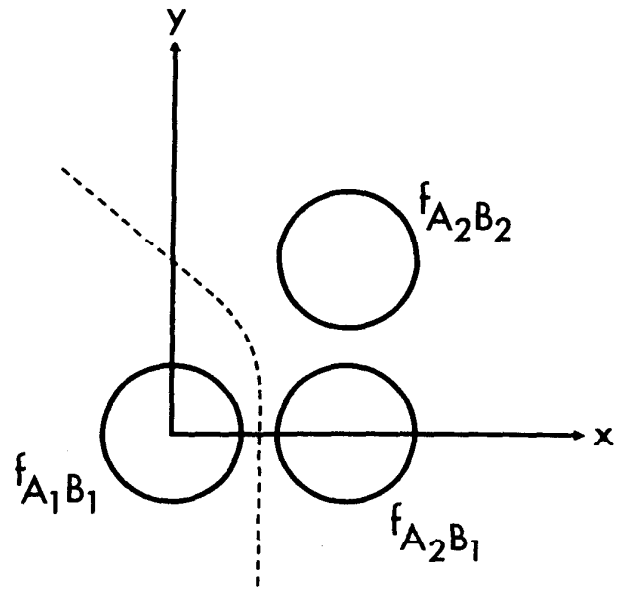


Figure 10. Contours of equal probability and the optimal decision bound for the task of Operational Definition B.

what Garner and Morton (1969) call the *performance parity criterion*. According to them, performance parity results if "performance on two perceptual tasks carried out simultaneously is the same as the sum of the performances on each task carried out separately" (p. 235). More generally, instead of an additive relation between the performances, a multiplicative or some other relationship may exist. In this sense performance parity has been a rather vague concept in general, although specific realizations have sometimes been well defined. Even worse, as we will demonstrate, it is occasionally unrelated or only weakly related to perceptual independence, which it is often used to test (Garner & Morton, 1969, also make this point). For instance, a typical procedure would be to compare performance in the A_1B_2 and the A_2B_1 conditions with performance in the A_2B_2 or simultaneous condition. The idea is that under "independence" there should be no inhibition or facilitation between the two stimuli in the simultaneous condition, and so performance in the A_2B_2 condition should be some simple function of the performances in the separate conditions. We will see below, however, that performance parity is an across-stimulus averaged type of result that is usually at most indirectly related to perceptual independence.

The form that the principle of parity will assume depends on the specific performance measure adopted. For example, with some measures the function is additive (Garner & Morton's, 1969, definition), whereas with others it is multiplicative. In our first and most general definition of parity we therefore leave the exact form of the function unspecified.

Definition 5: Performance parity results if performance in the A_2B_2 condition is some function of the performances in the A_1B_2 and the A_2B_1 conditions, that is, if

$$R_1(A_2B_2) = Q[R_2(A_2B_1), R_3(A_1B_2)],$$

where R_i is an observable overall performance measure and Q is a function into the real numbers.

Additive Parity

This definition is completely general. It places no restrictions on the function Q relating performance in the A_2B_2 condition to performance in the A_1B_2 and A_2B_1 conditions, and it even allows a different overall performance measure to be used in the different conditions. In its most common form, the same performance measure is used in each condition and the function Q is linear. This yields an additive notion of parity and the form considered by Garner and Morton (1969).

Definition 6: Additive parity results if performance in the A_2B_2 condition equals the sum of the performances in the A_1B_2 and the A_2B_1 conditions, that is, if

$$R(A_2B_2) = R(A_1B_2) + R(A_2B_1).$$

We already encountered one form of additive parity when we discussed Tanner's (1956) use of the orthogonality criterion as a test of perceptual independence. He chose $(d')^2$, from signal detection theory, as his performance measure and argued that perceptual independence must hold if

$$(d'_{AB})^2 = (d'_A)^2 + (d'_B)^2.$$

The notion of additive parity as a possible criterion of perceptual independence seems to come from two different sources. First is the fact that the variance of the sum of two random variables equals the sum of the individual variances if and only if the random variables are uncorrelated (Garner & Morton, 1969). This fact may be of limited value in testing perceptual independence, however, for at least two reasons. The first, mentioned before, is that uncorrelated random variables are not necessarily independent unless they are normally distributed. The second reason has to do with the nature of the experimental paradigms most amenable to tests of perceptual independence. The complete identification experiment with stimuli A_1B_1 , A_1B_2 , A_2B_1 , and A_2B_2 is a good example. Under perceptual independence we expect stimulus A_2B_2 to be perceived as a random ordered pair, (x, y) , whose components are independent random variables, that is, we expect stimulus components A and B to cause separate and independent perceptual effects. We do not ordinarily expect stimulus A_2B_2 to be perceived as a sum, $x + y$.

On the other hand, depending on task instructions, the sum may play an important role in the decision process. If the subject is asked to respond in a certain manner whenever the perceptual magnitude is large, then a natural decision strategy is to compute $x + y$. However, in a complete identification experiment where stimuli are created by factorially combining components, a perceptual space something like the one in Figure 3 should result. In this case the sum $x + y$ is never computed and so no use can be made of the fact that if two variables are uncorrelated, the variance of their sum equals the sum of their variances.

A second motivation for additive parity as a test of perceptual independence comes from information theory (Garner & Morton, 1969). The joint uncertainty $H(W, Z)$, or information, of two random variables W and Z equals the sum of the uncertainties of each variable if and only if the random variables are statistically independent (Garner, 1962; McGill, 1954; Shannon & Weaver, 1949). That is, under statistical independence,

$$H(W, Z) = H(W) + H(Z). \quad (6)$$

The amount of dependence between the two random variables can be determined by computing the contingent uncertainty (often called the transmitted information):

$$T(W, Z) = H(W) + H(Z) - H(W, Z). \quad (7)$$

Clearly, $T(W, Z)$ is zero if and only if the two random variables W and Z are statistically independent. (As a test of statistical independence, $T(W, Z)$ is therefore more useful than the widely known Pearson correlation, because it does not require the assumption of normal distributions.)

Early applications of information theory in psychology (e.g., Attneave, 1959; Garner, 1962) focused on observable stimulus and response variables. Garner and Morton (1969; see their Equation 3) presented the most sophisticated application of this approach to the problem of testing for perceptual independence. Basically, they suggested that to test for perceptual independence in a complete identification experiment with stimulus components A and B , one should use information theory to measure the contingent uncertainty between stimulus component A and response component b , $T_B(A, b)$, with stimulus component B partialled out. Because this term describes a sort of crossing over from one processing channel to another, they argued that it should be zero if perceptual independence occurs. In an analogous fashion, the same argument can be made for $T_A(B, a)$. However, the next result shows that this test is more logically related to separability than to perceptual independence (as defined in Definition 1).

Theorem 6: Consider the complete identification experiment with stimuli A_1B_1 , A_1B_2 , A_2B_1 , and A_2B_2 . If perceptual and decisional separability hold for components A and B , then the partial contingent uncertainties satisfy

$$T_B(A, b) = T_A(B, a) = 0. \quad (8)$$

Thus, the partial contingent uncertainties $T_B(A, b)$ and $T_A(B, a)$ can equal zero even if a perceptual dependence exists. Conversely, both of these terms can be greater than zero even if perceptual independence occurs. This could happen, for instance, if the subject sometimes confuses the response codes and responds, say, a_2b_1 when a_1b_2 was intended.

The test that

$$T_B(A, b) = T_A(B, a) = 0$$

is therefore logically unrelated to the perceptual independence of components A and B . On the other hand, because it can be used to test for perceptual and decisional separability, it is indirectly related to perceptual independence through Theorem 4, in exactly the same way as marginal response invariance (i.e.,

Theorem 5). In fact, the tests of Theorems 5 and 6 are closely related. If the two separabilities hold in a set of data, marginal response invariance and Equation 8 should both be found within the usual statistical error. If either marginal response invariance of Equation 8 is false, at least one of the two types of separability must be violated.

Recall that marginal response invariance was verified in the Townsend et al. (1980, 1981) data. If it held because the separability assumptions are valid, we expect Equation 8 to also hold in these data. In fact, this is exactly what happens. Table 4 contains estimates of the two partial contingent uncertainties, $T_B(A, b)$ and $T_A(B, a)$, for each of the four subjects in both the gap and the connected conditions of the Townsend et al. experiment. In spite of the fact that the maximum possible value for either uncertainty is only one bit, the obtained estimates still appear strikingly close to zero. Without attempting a statistical analysis, it seems reasonably safe to conclude that, as expected, Equation 8 is supported in the Townsend et al. experiment. Although the conjunction of marginal response invariance and Equation 8 does not logically imply the two forms of separability, their confirmation strongly supports this interpretation.

Perhaps it should not be too surprising that Equation 8 does not provide a direct test of perceptual independence. As a rule of thumb, the degree of dependency between stimulus and response will usually not tell much about the degree of perceptual dependency between a pair of stimulus components. By focusing on the perceptual effects, however, instead of on stimuli and responses, information theory does have the potential to provide a perfect test of perceptual independence. To see this, note that the uncertainty or information in the perceptual distribution of the stimulus pair $A_i B_j$ is

$$H_{A_i B_j}(x, y) = - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{A_i B_j}(x, y) \log f_{A_i B_j}(x, y) dx dy.$$

Similarly, the uncertainty in the marginal perceptual distribution associated with component A is

$$H_{A_i B_j}(x) = - \int_{-\infty}^{\infty} g_{A_i B_j}(x) \log g_{A_i B_j}(x) dx.$$

Following Equation 7 we can define the contingent uncertainty between components A and B in the perceptual distribution associated with the stimulus $A_i B_j$ as

$$T_{A_i B_j}(x, y) = H_{A_i B_j}(x, y) + H_{A_i B_j}(y) - H_{A_i B_j}(x). \quad (9)$$

From Equation 6 it is clear that this value equals zero if and only if perceptual independence holds.

Unfortunately, of course, in the experimental paradigms considered here, the Equation 9 uncertainties are unobservable, and at present, no way to estimate them is known. Thus, in a sense, Equation 9 is no more useful as an empirical test of perceptual independence than is Definition 1. On the other hand, if some estimates of the perceptual distributions or their uncertainties can be found, then Equation 9 will provide a simpler test because it only involves three statistics and one equality. In contrast, a test based on Definition 1 requires an infinite number of checks because there the equality must hold for all values of x and y .

Table 4

Results of Applying the Theorem 6 Separability Test to the Data of Townsend, Hu, and Ashby (1980, 1981)

Condition/subject	$\hat{T}_B(A, b)$	$\hat{T}_A(B, a)$
Gap		
1	.014	.006
2	.003	.002
3	.012	.008
4	.009	.002
Connected		
1	.000	.000
2	.007	.005
3	.004	.000
4	.006	.001

Performance Parity and Reaction Time

Another area where a performance parity criterion has been used to test perceptual independence is reaction time (RT). There is a widespread belief that the response time to a stimulus component should be a function of the perceptual value of that component. For example, an extreme perceptual value may make identification or discrimination, or whatever the task is, fairly easy, with the result that RT is short. This being the case, perceptual independence of a pair of stimulus components should cause the pair of associated RT components to be statistically independent, a condition known as stochastic independence.

This prediction is often difficult to test because the individual RT components are not usually observable on $A_2 B_2$ trials. To test for stochastic independence, some sort of processing model is needed. To simplify the discussion in this section, we will assume that the two levels of components A and B are presence and absence. This is the usual interpretation of levels in the reaction time literature. In this case, because both stimulus components must be fully processed on $A_2 B_2$ trials in a complete identification paradigm (or at least processed enough for them to be recognized), a plausible model might assume that processing of the stimulus components is parallel and exhaustive (see, e.g., Townsend & Ashby, 1983). In this case, the RT on $A_2 B_2$ trials is determined by the last component to be recognized, and under stochastic independence the RT distribution function can be written as

$$P[RT(A_2 B_2) \leq t] = P(T_A \leq t)P(T_B \leq t), \quad (10)$$

where T_i is the time to recognize stimulus component i . Given a way to estimate the component distribution functions $P(T_A \leq t)$ and $P(T_B \leq t)$, this equation permits a test of stochastic independence, and, assuming the unique relationship between perceptual values and processing time, it also permits a test of perceptual independence.

Unfortunately, there is much to go wrong with this argument. In addition to the extra assumption relating RT and the perceptual values, it requires precise knowledge of the subject's processing strategy. For example, Equation 10 is not correct if the subject serially processes the stimulus components. If this were the case, a test of Equation 10 would fail, but it would be in-

appropriate to conclude that perceptual independence therefore also fails.

Ignoring these problems for the moment, consider how an Equation 10 test of perceptual independence might be implemented. The function on the left is easily estimated from the A_2B_2 data, but the functions on the right are unobservable. One obvious approach is to estimate these functions from the data of the A_2B_1 and the A_1B_2 conditions. Note that such a test is based on a multiplicative performance parity criterion inasmuch as it says that independence holds if performance in the simultaneous (A_2B_2) condition, as measured by the empirical *RT* cumulative distribution function, equals the product of the performances in the separate (A_2B_1 and the A_1B_2) conditions.

The test is valid, of course, only if the time to process, say, component A is the same in the separate A_2B_1 condition as it is in the simultaneous A_2B_2 condition. If the levels of A and B are presence and absence, then, in the sense discussed earlier, the simultaneous A_2B_2 condition will involve a heavier processing load than either separate condition. We saw earlier that in the reaction time literature (e.g., Townsend & Ashby, 1983), the property that the performance (e.g., processing time) on an individual stimulus component does not depend on the total processing load (e.g., number of components to be processed) is called *unlimited capacity*. Thus, such a test of perceptual independence is valid only if capacity is unlimited. Many of the perceptual independence tests that are based on a performance parity criterion implicitly assume unlimited capacity.

Although in certain circumstances this relationship to unlimited capacity may make performance parity an undesirable way to test for perceptual independence, it may actually be an advantage when testing for separability. This is because of the close relationship between unlimited capacity and separability that we uncovered earlier. Recall that if the two levels of each component are presence and absence, then separability, as defined in Operational Definition A, is equivalent to unlimited capacity, at least if it holds for all components.

Using reaction times, Miller (1982) employed a paradigm similar to the task of Operational Definition B to test whether a simple visual signal (an asterisk) and an auditory signal (a bell) produce separate activations. Miller used the four stimuli of the complete identification task (i.e., A_1B_1 , A_1B_2 , A_2B_1 , and A_2B_2) but the response instructions of Operational Definition B (respond yes if either component is at level 2). His idea was that if the auditory and visual signals produced separate activations, the *RT* distribution function on simultaneous A_2B_2 trials should satisfy

$$P[RT(A_2B_2) \leq t] = P[RT(A_2B_1) \leq t] + P[RT(A_1B_2) \leq t] - P(T_A \leq t \text{ and } T_B \leq t | A_2B_2), \quad (11)$$

where T_i is the random finishing time of component i .

A derivation of Equation 11 is given in the proof of Theorem 7 below. It is shown there to depend on the following traditional *RT* definition of unlimited capacity (e.g., Townsend & Ashby, 1983):

$$P(T_A \leq t | A_2B_1) = P(T_A \leq t | A_2B_2) \quad (12)$$

and

$$P(T_B \leq t | A_1B_2) = P(T_B \leq t | A_2B_2). \quad (13)$$

Thus, in effect, Miller (1982) equates separability and unlimited capacity by assuming that if components A and B are separable, these two equalities will hold. Although, as we have mentioned, there are certain dangers involved in the use of *RT* to make inferences about perceptual events, this seems a reasonable approach. The components A and B are assumed separable if, for example, performance on component A is identical in the separate A_2B_1 condition to what it is in the simultaneous A_2B_2 condition.

Note that this is an Operational Definition A of separability. Even though Miller is using a paradigm similar to the one of Definition B, his test is based on a notion of separability derived from Definition A.

In the language of a popular theory of *RT*, the completion of this task requires four processing stages: namely, stimulus encoding, comparison, response selection, and execution (e.g., Smith, 1968; Sternberg, 1969). According to the general recognition theory, stimulus encoding is the stage during which the perceptual representation is generated, and comparison is the process of deciding, say, to which response quadrant the representation belongs. It is traditional in *RT* theory to assume that changing the number of components in this task affects the duration of only the comparison process (e.g., Sternberg, 1969). Thus, although other interpretations are possible, by assuming Equations 12 and 13, Miller appears to be making an assumption about the decision process, in particular, that the time it takes to make a decision about one component does not depend on whether the other component is present. In our terminology, this is a form of decisional separability.

Next note that the last term on the right in Equation 11 should equal the probability that $RT \leq t$ in a simultaneous A_2B_2 condition in which a correct response requires that both components be completed. With the paradigm that Miller (1982) used, however, this probability is not estimable, because the components T_A and T_B are not observable. However, because probabilities can never be negative, the following inequality is implied by Equations 12 and 13:

$$P[RT(A_2B_2) \leq t] \leq P[RT(A_2B_1) \leq t] + P[RT(A_1B_2) \leq t].$$

The better the performance, the greater the probability that *RT* will be less than a given time t , and so this condition states that performance in the A_2B_2 condition cannot be too much better than performance in either of the two separate conditions if separate activation occurs. Miller (1982) found this prediction to be consistently violated and on this basis argued that the visual and auditory signals were interacting in a mutually excitatory fashion. Responses were faster in the simultaneous condition than separability predicts.

By using our revised version of Operational Definition B and the Equations 12 and 13 definition of *RT* separability, Miller's test can be extended in the other direction (as well as proven rigorously). In addition, as with other tests discussed in this article, it is not logically restricted to experiments in which the stimulus component levels are presence and absence. Thus, we can generalize Equations 12 and 13 to the following.

Definition 7: Consider an experiment with stimuli A_1B_1 , A_1B_2 , A_2B_1 , and A_2B_2 . The components A and B are temporally sep-

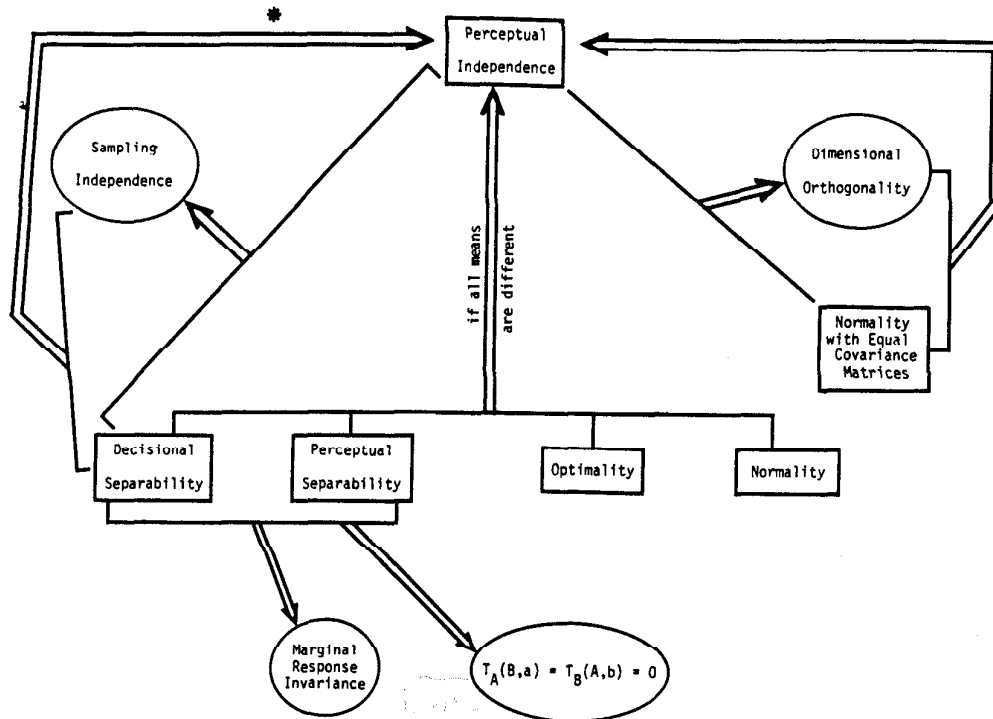


Figure 11. Schematic illustrating the relationships between some major concepts. (Arrows mean that the antecedent conditions taken together logically imply the consequent; the star on the upper left-hand arrow signifies that this implication requires sampling independence to hold for differing decision criteria.)

arable if the processing time of one component does not depend on the level of the other, that is, if

$$P(T_{A_i} \leq t | A_i B_1) = P(T_{A_i} \leq t | A_i B_2),$$

for $i = 1, 2$, and

$$P(T_{B_j} \leq t | A_1 B_j) = P(T_{B_j} \leq t | A_2 B_j),$$

for $j = 1, 2$, where, for example, T_{A_i} is the random time to process component A at level i .

Although temporal separability seems to be a form of decisional separability, it is impossible to specify the exact relationship between the two without an explicit theory relating the perceptual representation to the processing time of a stimulus. For example, one possibility that has been explored in the case of unidimensional stimulus representations is that processing time is inversely related to distance from the decision bound (e.g., Thomas, 1971; Thomas & Meyers, 1972). At any rate, we present Definition 7 because it seems intuitively to be related to separability and because it allows us to clarify and extend Miller's (1982) approach. The next result indicates that, in contradiction to Operational Definition B, temporal separability does predict better performance in the simultaneous condition with redundant stimulus components.

Theorem 7: Consider an experiment with stimuli $A_1 B_1$, $A_1 B_2$, $A_2 B_1$, and $A_2 B_2$ where the subject is instructed to indicate whether

either component is at level 2. If the components A and B are temporally separable, then

$$\begin{aligned} & \max\{P[RT(A_2 B_1) \leq t], P[RT(A_1 B_2) \leq t]\} \\ & \leq P[RT(A_2 B_2) \leq t] \\ & \leq P[RT(A_2 B_1) \leq t] + P[RT(A_1 B_2) \leq t]. \end{aligned}$$

This theorem provides a fairly powerful performance parity test of temporal separability. It also clearly illustrates the weakness of Operational Definition B. Even if components A and B are separable, performance in the simultaneous condition is better than in either separate condition.

Conclusion

The goal of this article has been to develop a theory of perceptual independence and the important notions that are closely associated with it in the perceptual literature. Toward this end a taxonomy of dimensional concepts as they relate to perceptual behavior was developed within the context of a general recognition theory. Simple yet rigorous definitions of perceptual independence as well as sampling independence, dimensional orthogonality, perceptual and decisional separability, and performance parity were offered within the scheme of this general recognition theory.

The theory used to investigate these concepts assumed only that a single presentation of a stimulus induces a perceptual effect in a multidimensional space, which in turn elicits a response. The extreme generality of this theory should make our conclusions quite robust.

Sampling independence, dimensional orthogonality, and perceptual and decisional separability were shown to be sensitive to different structures in a data set. Furthermore, we saw that none of these conditions is equivalent to perceptual independence and that each condition, by itself, provides only a weak test of whether subjects independently perceive a pair of stimulus components.

Theorems relating the major concepts to response probabilities and to each other are summarized in Figure 11. The observable conditions are circled, and all conditions that are typically unobservable are enclosed in rectangles. Note that, by themselves, none of the observable conditions implies (or is implied by) perceptual independence, and thus, if they are to be used to test independence some ancillary assumptions must be made. Fortunately, most of these assumptions are at least partially testable and/or commonly made in the psychological literature. For ex-

ample, as Figure 11 indicates, marginal response invariance and the test involving partial contingent uncertainties each provide a method of checking perceptual and decisional separability, and thus, when these tests are used in conjunction with sampling independence, a powerful method of testing perceptual independence results.

In addition, as we have already indicated, although normality may not be simple to test, it is very commonly assumed in psychological theorizing. The assumption of optimality is also reasonably popular, and there does exist some empirical evidence that with enough motivation and practice, subjects are capable of performing at near optimal levels in the kinds of tasks we have been considering (Ashby & Gott, 1985).

To the extent that the normality and optimality assumptions are reasonable, the existence of perceptual and decisional separability can be used to infer perceptual independence (because of Theorem 4). However, if marginal response invariance and the cross-contingent uncertainty condition are used to test for perceptual and decisional separability, nothing is gained by not also testing sampling independence before deciding on perceptual

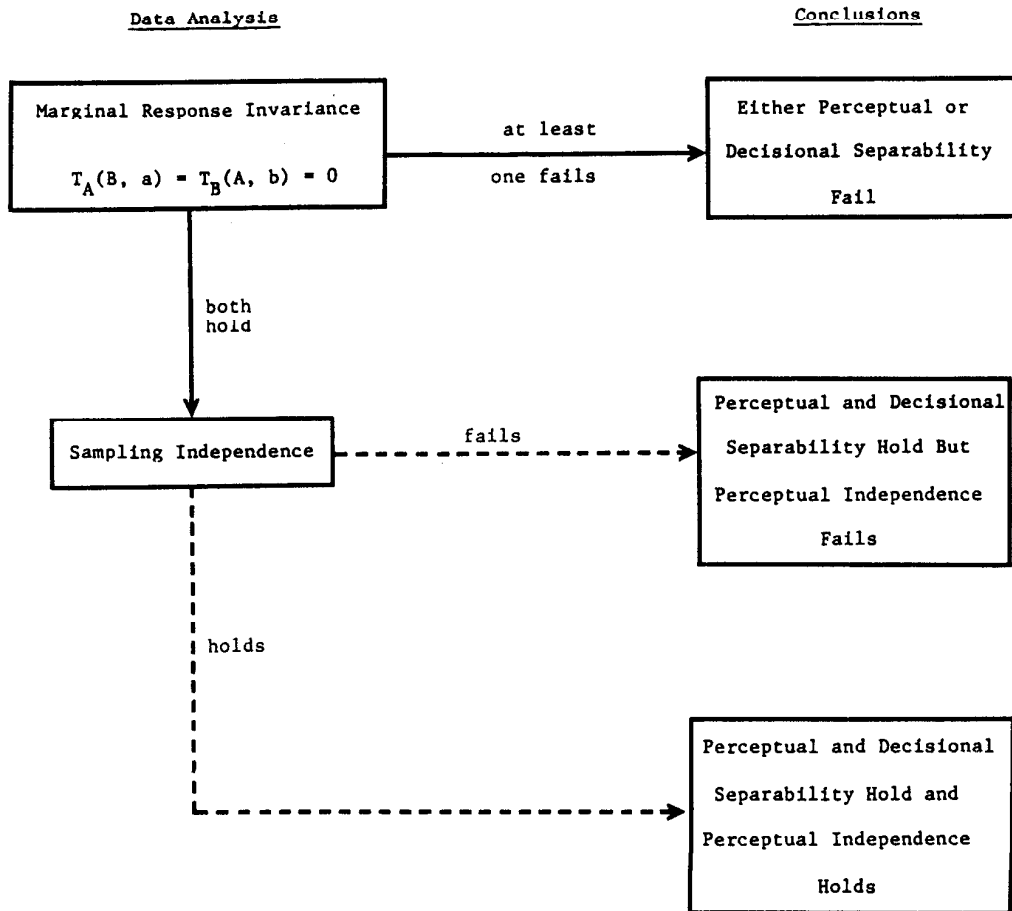


Figure 12. Flow chart illustrating the procedure for testing perceptual and decisional separability and perceptual independence. (Also indicated are the conclusions that can be drawn from every possible outcome of these tests.)

independence. If all the assumptions of Theorem 4 hold (including perceptual and decisional separability), then sampling independence will hold (because perceptual independence will hold). On the other hand, suppose perceptual independence fails but the subject is not responding optimally. If marginal response invariance and the cross-contingent uncertainty condition hold, then perceptual and decisional separability are strongly supported, but of course it would be a mistake to infer perceptual independence on this basis. In fact, under these conditions sampling independence will fail, and so perceptual independence can be ruled out.

The assumption in Figure 11 that we feel is least tenable is the one needed to relate dimensional orthogonality to perceptual independence, namely, that the covariance matrices associated with each perceptual distribution are all equal (see Theorem 3). For example, recall that we found evidence of the violation of this assumption in the Townsend et al. (1980, 1981) data (e.g., see Figure 6). Other damaging evidence comes from Ashby and Perrin (1985), who applied the general recognition theory to the study of empirical dissimilarity judgments. As previously mentioned, they showed that standard (Euclidean) distance-based dissimilarity models are contained within the general Gaussian recognition model as a special case in which all covariance matrices are equal. Thus, in this case, the general Gaussian recognition model, like the distance-based dissimilarity models, predicts that empirical dissimilarity judgments must satisfy the standard distance axioms. However, Ashby and Perrin (1985) showed that with unequal covariance matrices, the general Gaussian recognition model often predicts violations of the distance axioms. Because empirical dissimilarity judgments have frequently been found to violate distance axioms (e.g., Krumhansl, 1978; Tversky, 1977; Tversky & Gati, 1982), it appears that the assumption that all covariance matrices are equal is not generally true. One way to test its validity is to check if empirical dissimilarity judgments satisfy the distance axioms. If they do, then dimensional orthogonality becomes an excellent test of perceptual independence.

How can the experimentalist best make use of the results in this article? Although we have discussed several experimental paradigms, the one most useful for investigating independence-related concepts appears to be the complete identification experiment in which the stimulus ensemble is created by factorially combining several levels of two or more physical components. The nature of the components is relatively unimportant with regard to the tests developed here. As illustration, we applied several of the tests to a data set reported by Townsend et al. (1980, 1981). In this case the components were horizontal and vertical line segments, and the two component levels were presence and absence. It is important to note, however, that none of the tests require such binary valued levels (e.g., presence-absence).

To illustrate how our results might be used, we sometimes referred to a thought experiment that investigated the perceptual processing of the size and the wavelength of light reflected off of geometric figures such as squares. In this experiment two sizes and two wavelengths, each one or two jnds apart, are selected. Four stimuli are now constructed by factorial combination. Thus, for example, stimulus A_1B_2 might be a square of width A_1 cm

cut from paper reflecting a light of B_2 nm. A complete identification experiment with these stimuli is run and a full 4×4 confusion matrix is estimated. The tests developed above can now be applied. Figure 12 is a flow chart of the testing procedure, indicating the conclusions that can be drawn from every possible outcome.

First, the two tests developed in Theorems 5 and 6 are applied. If either is violated, then perceptual and/or decisional separability are false. If both tests are satisfied then because of converging evidence, the two forms of separability, although not logically implied, are strongly supported. In this case, sampling independence can now be used to test for perceptual independence. The dotted arrows signify that these implications are strongly indicated rather than logically implied.

Our primary goal in writing this article was to make sense of what we felt was a somewhat bewildering literature. To accomplish this goal required a general theory of perceptual performance that could interrelate the many different models and approaches this literature contains. We feel the general recognition theory developed above satisfies this requisite. For example, it contains both signal detection theory and the general Euclidean scaling model as special cases. Thus, although we focused here on perceptual independence as an organizing concept, the general recognition theory has the potential to interrelate other important concepts and results from a disparate aggregate of psychological literatures.

References

- Ashby, F. G. (1982, August). *Stimulus similarity in a geometric model of alphabetic confusion*. Paper presented at the 15th Annual Mathematical Psychology Meetings, Princeton University, Princeton, NJ.
- Ashby, F. G., & Gott, R. E. (1985). *Decision rules in the perception and categorization of multidimensional stimuli*. Unpublished manuscript.
- Ashby, F. G., & Perrin, N. A. (1985). *Toward a unified theory of similarity and recognition*. Unpublished manuscript.
- Attneave, F. (1959). *Applications of information theory to psychology*. New York: Holt, Rinehart & Winston.
- Carroll, J. D., & Chang, J. J. (1972, March). *IDOSCAL (Individual Differences in Orientation Scaling): A generalization of INDSCAL allowing IDIOSyncratic reference systems as well as analytic approximation to INDSCAL*. Paper presented at the annual meetings of the Psychometric Society, Princeton, NJ.
- Davison, M. L. (1983). *Multidimensional scaling*. New York: Wiley.
- Fukunaga, K. (1972). *Introduction to statistical pattern recognition*. New York: Academic Press.
- Garner, W. R. (1962). *Uncertainty and structure as psychological concepts*. New York: Wiley.
- Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.
- Garner, W. R. (1978). Aspects of a stimulus: Features, dimensions, and configurations. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization*. Hillsdale, NJ: Erlbaum.
- Garner, W. R., & Felfoldy, F. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, 1, 225-241.
- Garner, W. R., & Morton, J. (1969). Perceptual independence: Definitions, models, and experimental paradigms. *Psychological Bulletin*, 72, 233-259.
- Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts.

- Green, D. M., & Birdsall, T. G. (1978). Detection and recognition. *Psychological Review*, 85, 192-206.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Green, D. M., Weber, D. L., & Duncan, J. E. (1977). Detection and recognition of pure tones in noise. *Journal of the Acoustical Society of America*, 62, 948-954.
- Hefner, R. A. (1958). *Extensions of the law of comparative judgment to discriminable and multidimensional stimuli*. Unpublished doctoral dissertation, University of Michigan.
- Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology*, 148, 574-591.
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160, 106-154.
- Krumhansl, C. L. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review*, 85, 445-463.
- Kruskal, J. B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29, 1-27.
- Kruskal, J. B. (1964b). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 29, 115-129.
- Letttvin, J. Y., Maturana, H. R., McCulloch, W. S., & Pitts, W. H. (1959). What the frog's eye tells the frog's brain. *Proceedings of the Institute of Radio Engineers*, 47, 1940-1951.
- Lockhead, G. R. (1966). Effects of dimensional redundancy on visual discrimination. *Journal of Experimental Psychology*, 72, 95-104.
- Lockhead, G. R. (1972). Processing dimensional stimuli: A note. *Psychological Review*, 79, 410-419.
- 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.
- MacCallum, R. C. (1976). Effects on INDSCAL of nonorthogonal perceptions of object dimensions. *Psychometrika*, 19, 97-116.
- McGill, W. J. (1954). Multivariate information transmission. *Psychometrika*, 19, 97-116.
- Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, 14, 247-279.
- Monahan, J. S., & Lockhead, G. R. (1977). Identification of integral stimuli. *Journal of Experimental Psychology: General*, 106, 94-110.
- Morrison, O. F. (1976). *Multivariate statistical methods*. New York: McGraw-Hill.
- Peterson, W. W., Birdsall, T. G., & Fox, W. C. (1954). The theory of signal detectability. *Transactions of the IRE Professional Group on Information Theory*, 4, 171-212.
- Posner, M. I. (1964). Information reduction in the analysis of sequential tasks. *Psychological Review*, 71, 491-504.
- Regan, D. (1982). Visual information channeling in normal and disordered vision. *Psychological Review*, 89, 407-444.
- Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.
- Shepard, R. N. (1962a). The analysis of proximities: Multidimensional scaling with an unknown distance function I. *Psychometrika*, 27, 125-140.
- Shepard, R. N. (1962b). The analysis of proximities: Multidimensional scaling with an unknown distance function II. *Psychometrika*, 27, 219-246.
- Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, 1, 54-87.
- Smith, E. E. (1968). Choice reaction time: An analysis of the major theoretical positions. *Psychological Bulletin*, 69, 77-110.
- Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), *Attention and performance* (Vol. 2, pp. 276-315). Amsterdam: North Holland Press.
- Tanner, W. P. (1956). Theory of recognition. *Journal of the Acoustical Society of America*, 28, 882-888.
- Tanner, W. P., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, 61, 401-409.
- Taylor, D. A. (1976). Holistic and analytic processes in the comparison of letters. *Perception & Psychophysics*, 20, 187-190.
- Thomas, E. A. C. (1971). Sufficient conditions for monotone hazard rate: An application to latency-probability curves. *Journal of Mathematical Psychology*, 8, 303-332.
- Thomas, E. A. C., & Myers, J. L. (1972). Implications of latency data for threshold and nonthreshold models of signal detection. *Journal of Mathematical Psychology*, 9, 263-285.
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34, 273-286.
- Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.
- Townsend, J. T., & Ashby, F. G. (1982). Experimental test of contemporary mathematical models of visual letter recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 834-864.
- Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. New York: Cambridge University Press.
- Townsend, J. T., Hu, G. G., & Ashby, F. G. (1980). A test of visual feature sampling independence with orthogonal straight lines. *Bulletin of the Psychonomic Society*, 15, 163-166.
- Townsend, J. T., Hu, G. G., & Ashby, F. G. (1981). Perceptual sampling of orthogonal straight line features. *Psychological Research*, 43, 259-275.
- Townsend, J. T., & Landon, D. E. (1983). Mathematical models of recognition and confusion in psychology. *Mathematical Social Sciences*, 4, 25-71.
- Tucker, L. R. (1972). The relations between multidimensional scaling and three-mode factor analysis. *Psychometrika*, 37, 3-27.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327-352.
- Tversky, A., & Gati, I. (1982). Similarity, separability, and the triangle inequality. *Psychological Review*, 89, 123-154.
- Wandell, B. A. (1982). Measurement of small color differences. *Psychological Review*, 89, 281-302.
- Wandmacher, J. (1976). Multicomponent theory of perception. *Psychological Research*, 39, 17-37.
- Young, F. W., & Hammer, R. M. (in press). *Multidimensional scaling: Theory and applications*. Hillsdale, NJ: Erlbaum.
- Zinnes, J. L., & MacKay, D. B. (1983). Probabilistic multidimensional scaling: Complete and incomplete data. *Psychometrika*, 48, 27-48.

(Appendix follows on next page)

Appendix

Proof of Theorem 1

1. Under the conditions of the theorem,

$$\begin{aligned} P(a_2 b_2 | A_i B_j) &= \int_{y_0}^{\infty} \int_{x_0}^{\infty} g_{A_i B_j}(x) g_{A_i B_j}(y) dx dy \\ &= [1 - G_{A_i B_j}(x_0)] [1 - G_{A_i B_j}(y_0)], \end{aligned}$$

where $G_{A_i B_j}(z)$ is the cumulative distribution function associated with $g_{A_i B_j}(z)$. Similarly,

$$\begin{aligned} P(a_2 b_1 | A_i B_j) &= \int_{-\infty}^{y_0} \int_{x_0}^{\infty} g_{A_i B_j}(x) g_{A_i B_j}(y) dx dy \\ &= [1 - G_{A_i B_j}(x_0)] G_{A_i B_j}(y_0), \end{aligned}$$

and

$$\begin{aligned} P(a_1 b_2 | A_i B_j) &= \int_{y_0}^{\infty} \int_{-\infty}^{x_0} g_{A_i B_j}(x) g_{A_i B_j}(y) dx dy \\ &= G_{A_i B_j}(x_0) [1 - G_{A_i B_j}(y_0)]. \end{aligned}$$

Substituting these expressions into both sides of Definition 2 verifies the theorem.

2. With decision bounds parallel to the coordinate axes, the left-hand side of Definition 2 can be rewritten as

$$P(a_2 b_2 | A_i B_j) = \int_{y_0}^{\infty} \int_{x_0}^{\infty} f_{A_i B_j}(x, y) dx dy. \quad (A-1)$$

In a similar fashion, the right-hand side can be rewritten as

$$\begin{aligned} & [P(a_1 b_2 | A_i B_j) + P(a_2 b_2 | A_i B_j)] \\ & \quad \times [P(a_1 b_1 | A_i B_j) + P(a_2 b_1 | A_i B_j)] \\ &= \left[\int_{y_0}^{\infty} \int_{-\infty}^{x_0} f_{A_i B_j}(x, y) dx dy + \int_{y_0}^{\infty} \int_{x_0}^{\infty} f_{A_i B_j}(x, y) dx dy \right] \\ & \quad \times \left[\int_{-\infty}^{y_0} \int_{x_0}^{\infty} f_{A_i B_j}(x, y) dx dy + \int_{y_0}^{\infty} \int_{x_0}^{\infty} f_{A_i B_j}(x, y) dx dy \right] \\ &= \left[\int_{y_0}^{\infty} \int_{-\infty}^{\infty} f_{A_i B_j}(x, y) dx dy \right] \left[\int_{-\infty}^{\infty} \int_{x_0}^{\infty} f_{A_i B_j}(x, y) dx dy \right] \\ &= \left[\int_{y_0}^{\infty} g_{A_i B_j}(y) dy \right] \left[\int_{x_0}^{\infty} g_{A_i B_j}(x) dx \right] \\ &= \int_{y_0}^{\infty} \int_{x_0}^{\infty} g_{A_i B_j}(x) g_{A_i B_j}(y) dx dy. \quad (A-2) \end{aligned}$$

Differentiation of Equations A-1 and A-2 yields

$$f_{A_i B_j}(x_0, y_0) = g_{A_i B_j}(x_0) g_{A_i B_j}(y_0). \quad (A-3)$$

By assumption, this relation holds for all x_0 and y_0 , so components A and B are perceived independently.

3. We first show that sampling independence does not imply perceptual independence. It suffices to construct an example in which sampling independence holds but perceptual independence does not and in which

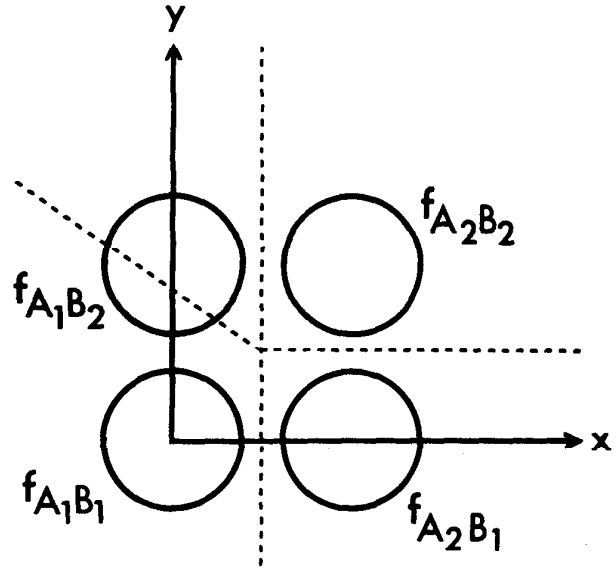


Figure A-1. Contours of equal probability and the decision bounds for a subject perceiving components A and B independently but violating sampling independence.

the decision bounds are not parallel to the coordinate axes. To do this, consider the model illustrated in Figure 3 but assume that all contours of equal probability are identical ellipses in which the major axis is parallel to the abscissa. By the above proof, this model predicts sampling independence. Now rotate the figure (but leave the coordinate axes unchanged) about the point (x_0, y_0) a total of θ° , where θ is not a multiple of 90 . This rotates the densities and the decision axes but not the original coordinates. Clearly, neither perceptual independence nor the parallelity condition is satisfied in this rotated model. However, because rotation preserves distances, the volumes corresponding to the conditional response probabilities of Definition 2 will be invariant for any rotation. Thus, sampling independence will hold even though neither condition of the theorem is satisfied.

To show that perceptual independence, by itself, does not imply sampling independence, consider the model illustrated in Figure A-1. This model is identical to the one of Figure 3 except for the decision bound separating the $A_1 B_2$ and the $A_1 B_1$ distributions. Thus, perceptual independence of components A and B holds in each stimulus configuration. However, note that of all the terms in the sampling independence definition (Definition 2), only the term $P(a_1 b_2 | A_i B_j)$ is affected by this change. Because this term only appears in the right-hand side of the defining equation, sampling independence can not hold in both the Figure 3 and the Figure A-1 models. However, we know already that it holds in the Figure 3 model, and so the model of Figure A-1 is an example in which perceptual independence holds but sampling independence does not.

To find an example in which neither perceptual independence nor sampling independence holds, consider again Figure A-1. Suppose, however, that instead of circles, the contours are all identical ellipses whose major and minor axes agree with the coordinate axes. As before, this model exhibits perceptual but not sampling independence. Now rotate the figure (but not the coordinate axes) about the point (x_0, y_0) (i.e., the point where the decision bounds intersect) a total of θ° , where $0 < \theta < 90$. Perceptual independence now fails and because such a rotation does

not affect the conditional response probabilities, sampling independence will also fail.

Finally, Figure 3 illustrates a case in which both perceptual independence and sampling independence hold.

Proof of Theorem 2

It suffices to prove the result in the two-dimensional case. Call the dimensions in the oblique space u and v and the dimensions in the orthogonal space x and y . Assume for simplicity, and without loss of generality, that dimensions u and x are coincident. Let θ be the angle between u and v .

The circular contours of equal probability in the oblique space have coordinates

$$\begin{bmatrix} u \\ \sqrt{1-u^2} \end{bmatrix},$$

and the transformation to orthogonal dimensions is accomplished by the matrix

$$\begin{bmatrix} 1 & -\cos \theta / \sin \theta \\ 0 & 1 / \sin \theta \end{bmatrix}.$$

Thus, the contours of equal probability in the orthogonal space have coordinates

$$\begin{bmatrix} 1 & -\cos \theta / \sin \theta \\ 0 & 1 / \sin \theta \end{bmatrix} \begin{bmatrix} u \\ \sqrt{1-u^2} \end{bmatrix} = \begin{bmatrix} u - \sqrt{1-u^2} \cos \theta / \sin \theta \\ \sqrt{1-u^2} / \sin \theta \end{bmatrix},$$

which implies

$$x = u - y \cos \theta$$

and

$$y = \sqrt{1-u^2} / \sin \theta.$$

Eliminating u we obtain

$$x^2 + y^2 + (2 \cos \theta)xy = 1.$$

This is the equation of the ellipse that is the equal probability contour of a bivariate normal distribution with correlation

$$\rho_{xy} = -\cos \theta.$$

Ashby and Perrin (1985) present a more detailed description of the parameter mappings between the two models.

Proof of Theorem 3

The proof is sketched in the paragraph immediately following the theorem.

Proof of Theorem 4

First note that Condition 2 together with perceptual separability implies a rectangular configuration for the μ_{ij} in the x, y plane. To prove the theorem, it suffices to successively consider pairs of stimuli each with a component in common. For instance, consider the pair A_1B_2 and A_2B_2 . If the subject is responding optimally and the perceptual distributions associated with these stimuli are bivariate normal, then the optimal decision bound satisfies the equation (e.g., Ashby & Gott, 1985; Fukunaga, 1972)

$$\begin{aligned} & (\sigma_{x_{12}}^2 - \sigma_{x_{22}}^2)x^2 + (\sigma_{y_{12}}^2 - \sigma_{y_{22}}^2)y^2 \\ & + (2\text{COV}_{12} - 2\text{COV}_{22})xy + C_1x + 2(\mu_{y_{22}}\sigma_{y_{22}}^2 - \mu_{y_{12}}\sigma_{y_{12}}^2 \\ & + \mu_{x_{22}}\text{COV}_{22} - \mu_{x_{12}}\text{COV}_{12})y - C_2 = 0, \end{aligned}$$

where C_1 and C_2 are constants depending on the parameters of the two distributions and, for example, $\mu_{y_{12}}$ is the mean on the y dimension of the perceptual distribution of stimulus A_1B_2 . Analogous definitions hold for the other terms. Because of perceptual separability,

$$\sigma_{y_{12}}^2 = \sigma_{y_{22}}^2 \quad \text{and} \quad \mu_{y_{12}} = \mu_{y_{22}},$$

and so this equation reduces to

$$\begin{aligned} & (\sigma_{x_{12}}^2 - \sigma_{x_{22}}^2)x^2 + (2\text{COV}_{12} - 2\text{COV}_{22})xy + C_1x \\ & + 2(\mu_{x_{22}}\text{COV}_{22} - \mu_{x_{12}}\text{COV}_{12})y - C_2 = 0. \quad (\text{A-4}) \end{aligned}$$

Because of decisional separability, however, the relevant decision bound satisfies

$$x - x_0 = 0.$$

By the assumptions of the theorem, Equation A-4 must reduce to this form. This can occur only if

$$\sigma_{x_{12}}^2 = \sigma_{x_{22}}^2,$$

$$\text{COV}_{12} = \text{COV}_{22}, \quad \text{and}$$

$$\mu_{x_{22}}\text{COV}_{22} - \mu_{x_{12}}\text{COV}_{12} = 0.$$

Call the mutual covariance cov . Then the last restriction reduces to

$$(\mu_{x_{22}} - \mu_{x_{12}})\text{cov} = 0,$$

which requires either that $\mu_{x_{22}} = \mu_{x_{12}}$ or $\text{cov} = 0$. If the former is true, then because $\mu_{y_{12}} = \mu_{y_{22}}$ it follows that $\mu_{12} = \mu_{22}$, which violates an assumption of the theorem. Therefore $\mu_{x_{12}} \neq \mu_{x_{22}}$, and so it must be true that $\text{cov} = \text{COV}_{12} = \text{COV}_{22} = 0$. Because we are assuming normality, perceptual independence follows (see Lemma 1). The proof for other pairs follows in an analogous fashion.

Proof of Theorem 5

The probability of correctly recognizing component A_1 when the stimulus is A_1B_j is

$$P(\text{reporting } a_1 | A_1B_j) = P(a_1b_1 | A_1B_j) + P(a_1b_2 | A_1B_j).$$

If decisional separability holds, then

$$\begin{aligned} & P(a_1b_1 | A_1B_j) + P(a_1b_2 | A_1B_j) \\ & = \int_{-\infty}^{y_0} \int_{-\infty}^{x_0} f_{A_1B_j}(x, y) dx dy + \int_{y_0}^{\infty} \int_{-\infty}^{x_0} f_{A_1B_j}(x, y) dx dy \\ & = \int_{-\infty}^{x_0} g_{A_1B_j}(x) dx. \end{aligned}$$

If perceptual separability holds, this expression reduces to

$$P(a_1b_1 | A_1B_j) + P(a_1b_2 | A_1B_j) = \int_{-\infty}^{x_0} g_{A_1}(x) dx,$$

which does not depend on the level of component B . A similar derivation produces

$$P(a_2b_1 | A_2B_j) + P(a_2b_2 | A_2B_j) = \int_{x_0}^{\infty} g_{A_2}(x) dx,$$

which also does not depend on the level of component B , verifying the theorem for component A . An analogous argument can be made for component B .

To show that neither kind of separability alone implies a performance constancy, consider the following two counterexamples. First, consider the model of Figure 3 but assume that the decision bounds are rotated counterclockwise θ° where $0 < \theta < 90$. In this case perceptual separability holds but decisional separability does not and

$$P(a_2b_2|A_2B_2) + P(a_2b_1|A_2B_2) > P(a_2b_2|A_2B_1) + P(a_2b_1|A_2B_1), \quad (\text{A-5})$$

and so constancy fails.

Second, again consider the Figure 3 model but now assume that the A_2B_2 distribution is shifted to the right by some amount. In this case perceptual separability fails but decisional separability holds. The inequality of Equation A-5 is still valid, and so constancy again fails.

Proof of Theorem 6

We prove the result for $T_B(A, b)$. The proof for $T_A(B, a)$ follows in an analogous fashion. First note that (see, e.g., Garner, 1962)

$$T_B(A, b) = \sum_m P(B_m) T_{B_m}(A, b),$$

where $P(B_m)$ is the a priori probability of presenting component B at level m . For simplicity we assume $P(B_m) = 1/2$ for $m = 1, 2$ and $P(A_k) = 1/2$ for $k = 1, 2$. Although the theorem does not depend on this assumption, the derivations are simpler if it is true. Now

$$T_{B_m}(A, b) = H_{B_m}(A) + H_{B_m}(b) - H_{B_m}(A, b).$$

We examine each term on the right in turn. First, note that because

$$\begin{aligned} P(A_1) &= P(A_2) = 1/2, \\ H_{B_m}(A) &= 1, \end{aligned} \quad (\text{A-6})$$

for $m = 1, 2$. Second, note that

$$\begin{aligned} H_{B_m}(b) &= \sum_j P_{B_m}(b_j) \log P_{B_m}(b_j) \\ &= -\frac{1}{2} \sum_j \left[\sum_k \sum_i P(a_i b_j | A_k B_m) \right] \\ &\quad \times \log \frac{1}{2} \sum_k \sum_i P(a_i b_j | A_k B_m). \end{aligned}$$

Because of decisional separability

$$\begin{aligned} \sum_i P(a_i b_1 | A_k B_m) &= \int_{-\infty}^{y_0} \int_{-\infty}^{\infty} f_{A_k B_m}(x, y) dx dy \\ &= \int_{-\infty}^{y_0} g_{A_k B_m}(y) dy \\ &= G_{A_k B_m}(y_0). \end{aligned}$$

Similarly,

$$\sum_i P(a_i b_2 | A_k B_m) = 1 - G_{A_k B_m}(y_0) = \bar{G}_{A_k B_m}(y_0).$$

Therefore,

$$\begin{aligned} H_{B_m}(b) &= -\frac{1}{2} \left[\sum_k G_{A_k B_m}(y_0) \log \frac{1}{2} \sum_k G_{A_k B_m}(y_0) \right. \\ &\quad \left. + \sum_k \bar{G}_{A_k B_m}(y_0) \log \frac{1}{2} \sum_k \bar{G}_{A_k B_m}(y_0) \right]. \end{aligned}$$

Because of perceptual separability,

$$G_{A_1 B_m}(y_0) = G_{A_2 B_m}(y_0) = G_{B_m}(y_0).$$

And so,

$$\begin{aligned} H_{B_m}(b) &= -\frac{1}{2} [2G_{B_m}(y_0) \log G_{B_m}(y_0) \\ &\quad + 2\bar{G}_{B_m}(y_0) \log \bar{G}_{B_m}(y_0)] \\ &= -[G_{B_m}(y_0) \log G_{B_m}(y_0) \\ &\quad + \bar{G}_{B_m}(y_0) \log \bar{G}_{B_m}(y_0)]. \end{aligned} \quad (\text{A-7})$$

Similarly,

$$\begin{aligned} H_{B_m}(A, b) &= -\sum_k \sum_j P_{B_m}(A_k, b_j) \log P_{B_m}(A_k, b_j) \\ &= -\frac{1}{2} \sum_k \sum_j \left[\sum_i P(a_i b_j | A_k B_m) \right] \log \frac{1}{2} \sum_i P(a_i b_j | A_k B_m) \\ &= -\frac{1}{2} \sum_k [G_{A_k B_m}(y_0) \log \frac{1}{2} G_{A_k B_m}(y_0) \\ &\quad + \bar{G}_{A_k B_m}(y_0) \log \frac{1}{2} \bar{G}_{A_k B_m}(y_0)] \\ &= -\frac{1}{2} [2G_{B_m}(y_0) \log \frac{1}{2} G_{B_m}(y_0) \\ &\quad + 2\bar{G}_{B_m}(y_0) \log \frac{1}{2} \bar{G}_{B_m}(y_0)] \\ &= 1 - [G_{B_m}(y_0) \log G_{B_m}(y_0) \\ &\quad + \bar{G}_{B_m}(y_0) \log \bar{G}_{B_m}(y_0)]. \end{aligned} \quad (\text{A-8})$$

Adding Equations A-6 and A-7 and subtracting A-8 yields

$$T_{B_m}(A, b) = 0,$$

for $m = 1, 2$, and thus,

$$\begin{aligned} T_B(A, b) &= \sum_m P(B_m) T_{B_m}(A, b) \\ &= 0. \end{aligned}$$

Proof of Theorem 7

To establish the lower bound, note that

$$\begin{aligned} P[RT(A_2 B_2) \leq t] &= P[\min(T_A, T_B) \leq t | A_2 B_2] \\ &= 1 - P(T_A > t \text{ and } T_B > t | A_2 B_2) \\ &= 1 - \int_t^\infty \int_t^\infty f_{A_2 B_2}(t_a, t_b) dt_a dt_b, \end{aligned}$$

where $f_{A_2B_1}(t_a, t_b)$ is the joint density function on the completion times of components A_1 and B_1 . Similarly,

$$\begin{aligned} P[RT(A_2B_1) \leq t] &= P(T_A \leq t|A_2B_1) \\ &= 1 - P(T_A > t|A_2B_1) \\ &= 1 - \int_t^\infty f_{A_2}(t_a) dt_a. \end{aligned}$$

Using the definition of temporal separability,

$$\int_t^\infty f_{A_2}(t_a) dt_a = \int_0^\infty \int_t^\infty f_{A_2B_2}(t_a, t_b) dt_a dt_b,$$

and so

$$\begin{aligned} P[RT(A_2B_1) \leq t] &= 1 - \int_0^\infty \int_t^\infty f_{A_2B_2}(t_a, t_b) dt_a dt_b \\ &\leq 1 - \int_t^\infty \int_t^\infty f_{A_2B_2}(t_a, t_b) dt_a dt_b \\ &= P[RT(A_2B_2) \leq t]. \end{aligned}$$

In an analogous fashion it can be shown that

$$P[RT(A_1B_2) \leq t] \leq P[RT(A_2B_2) \leq t],$$

thus verifying the lower bound.

To establish the upper bound, note that

$$\begin{aligned} P[RT(A_2B_2) \leq t] &= P[\min(T_A, T_B) \leq t|A_2B_2] \\ &= P(T_A \leq t \text{ and } T_B \leq t|A_2B_2) + P(T_A \leq t \text{ and } \\ &\quad T_B > t|A_2B_2) + P(T_A > t \text{ and } T_B \leq t|A_2B_2) \\ &= [P(T_A \leq t \text{ and } T_B \leq t|A_2B_2) \\ &\quad + P(T_A \leq t \text{ and } T_B > t|A_2B_2)] \\ &\quad + [P(T_A > t \text{ and } T_B \leq t|A_2B_2) \\ &\quad - P(T_A \leq t \text{ and } T_B \leq t|A_2B_2)]. \end{aligned}$$

Using the definition of temporal separability, the terms in the first set of square brackets equal $P[RT(A_2B_1) \leq t]$, and the terms in the second set of square brackets equal $P[RT(A_1B_2) \leq t]$, thus establishing the upper bound.

Received June 8, 1984

Revision received August 6, 1985 ■