Perceptual interactions of facial dimensions in speeded classification and identification

ROBIN D. THOMAS

Miami University, Oxford, Ohio

The representation underlying the identification and classification of semirealistic line drawings taken from a computer model of the face was investigated by using a speeded classification task and an identification task. These data were analyzed by using a multidimensional extension of signal detection theory, within which varieties of perceptual interactions between dimensions within and across stimuli can be characterized. The dimensions of interest here were eye separation, nose length, and mouth width. The response time and accuracy data from the speeded classification task suggest that processing of a given feature did depend on whether other features were present or absent, but given that other features were present, the results strongly support separability (a macrolevel, across-stimulus form of invariance) for all pairs of facial dimensions used. This separability was confirmed by the subsequent identification task. Owing to its greater resolution, the identification task can reveal interactions that might exist at more microlevels of processing. In fact, the identification data did indicate the presence of perceptual dependence between facial dimensions *within a stimulus* when the dimensions that were varied were close in spatial proximity (i.e., eye separation and nose length). Within the theoretical framework, perceptual dependence can be interpreted as correlated noise between otherwise separate channels (and hence, is logically distinct from separability). This dependence was greatly reduced for dimensions that were more distant (eyes and mouth). The relation between these results and the configural effects that have been observed with faces as stimuli in other studies is discussed.

One question in the study of the perception of faces concerns whether the face is processed as a configuration (holistically) or as a collection of independent features. This question has been addressed from a number of different viewpoints (see Bruce, 1988, for a review). As was noted by Tanaka and Farah (e.g., 1993), just what is meant by "holistic" and "featural" processing has been difficult to tie down. A full analysis of any cognitive-processing system requires several, logically orthogonal characteristics to be specified (Townsend, 1974; Townsend & Ashby, 1983).

In the face recognition literature, "configural," "interactive," or "holistic" processing has been interpreted from many perspectives. In terms of architecture, one view has centered on the equating of parallel processing of facial features with interactive processing and serial with independent processing (Bradshaw & Wallace, 1971; Mathews, 1978). As Tanaka and Farah (1993) astutely observed, it is entirely possible that faces are represented by independent features that can be processed in parallel (see also Townsend & Ashby, 1983). In a similar vein, Sergent (1984) argued that if features are independently processed,

"different" latencies for faces differing by two features should not be faster than "different" latencies for faces differing by the most salient feature. With the feature of chin contour determined to be the most salient, she found that "different" latencies for face pairs differing on both chin contour and another feature were faster than "different" latencies for face pairs differing on the single dimension of chin contour—a result she interpreted as indicating interactive, configural processing. However, an independent parallel processing of the features would predict exactly such statistical race gains (Colonius & Vorberg, 1994; Townsend & Nozawa, 1995); that is, interactivity is not a necessary requirement for a redundancy effect. Sergent also performed multidimensional scaling (MDS) analyses of similarity ratings and response times (RTs) for pairs of faces that were used to derive a perceptual representation of the stimulus configuration. The resulting configuration was examined for evidence of interactive processing. The perceptual representation of individual feature values that emerged (i.e., their locations in the MDS) depended on the characteristics of the other features in the face.

A different approach to the holistic issue includes attempts to identify the basic unit of representation of a face as being the whole face, rather than the individual parts (Farah, Tanaka, & Drain, 1995; Tanaka & Farah, 1991, 1993). These series of studies have employed variants of the part-whole recognition technique devised by Palmer (1977) to identify object parts. In a series of studies, these researchers found that parts of faces were dis-

I thank Helena Kadlec and Todd Maddox for their valuable suggestions on an earlier version of this paper. Very special thanks go to Lester Krueger for his attention and comments, which helped to greatly improve many aspects of this work. This work was supported by NIMH Grant R03 MH58812-02. Correspondence concerning this article should be addressed to R. D. Thomas, Department of Psychology, Miami University, Oxford, OH 45050 (e-mail: thomasrd@miaovx1.acs.muohio.edu).

proportionately more recognizable in whole faces than as isolated parts—disproportionate, that is, to the recognizability of parts and wholes of other, nonface objects or inverted faces. Tanaka and Farah (1993) interpreted this as evidence of holistic processing of the face. However, the data from these experiments cannot be used to rule out a featural representation of upright faces in which the features exhibit interactivity (e.g., a stochastic dependency), as in Sergent's (1984) MDS configurations. Nonetheless, this evidence suggests that, in some tasks, facial features interact.

Two primary goals of the present work were to build on this empirical work by clarifying the nature and degree of interaction in the processing of a face when it is involved in basic cognitive tasks by using unambiguous definitions provided by a powerful theory of perception that is ideally suited for this purpose. The next section describes this theoretical framework, within which configural or emergent properties can be modeled as perceptual interactions of a given kind (Kadlec, 1999a). Following that, two experiments, a speeded classification task and an identification task, serve to illustrate the application of this theory to the perceptual processing of semirealistic drawings of a face.

Constructs of the General Recognition Theory

The approach taken follows a long tradition that has been concerned with how pairs of stimulus attributes are processed by an organism in simple perceptual tasks. An early distinction made between pairs of dimensions classifies them as either *separable* or *integral* (Attneave, 1950; Garner, 1974; Shepard, 1964). A stimulus was said to be composed of separable dimensions if it could be easily analyzed into those dimensions to which selective attention is possible. On the other hand, selective attention to the individual dimensions is extremely difficult, if not impossible, for stimuli composed of integral dimensions. Interactivity, then, in this tradition is characterized as integrality of stimulus dimensions. Typical separable dimensions are shape and color or circle size and radial line orientation. Hue and saturation of a color are among the most commonly studied integral dimensions. The difficulty with this simple distinction is that many empirical findings could not be cleanly fit into either category—a problem that led to a proliferation of dimension types and circularities of definitions (Garner, 1976; Maddox, 1992).

A theoretical framework within which these concepts and additional notions of independence are characterized independently of the tasks used to reveal them is that of multidimensional signal detection theory (SDT; e.g., Olzak, 1986; Wickens, 1992), also termed the *general recognition theory* (GRT) by Ashby and Townsend (1986; see also Kadlec, 1995; Kadlec & Townsend, 1992a, 1992b; Thomas, 1995a, 1999). Like SDT, the GRT separates the decisional process from the perceptual process and, hence, in the multivariate case, affords a number of potential sources of dimensional interactions.

Perceptual processes. Just as in signal detection, the perception of a stimulus is assumed to be noisy, so that a

stimulus is represented by a distribution of perceptual effects accrued across the course of the experiment. So that central concepts can be defined rigorously, a bit of notation will be necessary. For exposition, assume the stimuli are constructed from two physical components, A and B, each sampled at one of two possible levels to obtain a *feature-complete factorial set*. For example, the facial dimensions under investigation in the present work include the separation of the eyes, nose length, and mouth width of semirealistic line-drawing faces taken from a three-dimensional rendering of the face (Parke, 1982). An example of a stimulus is shown in Figure 1.

If one assumes that the two physical dimensions (A = eye-separation in pixels, B = nose length in pixels) map onto two perceptual dimensions, X (perceived eye distance) and Y (perceived nose length), we can denote the density of perceptual effects for an individual face A_iB_j by $f_{A_iB_j}(x, y)$. Figure 2 illustrates equal-likelihood contours for four stimuli that were obtained by cross-sectioning hypothetical perceptual densities.

From each joint distribution, two marginal distributions, one for each dimension, $g_{A_iB_j}(x)$ for component A and $g_{A_iB_j}(y)$ for component B, can be obtained by integrating over (i.e., collapsing across) the opposing dimension. These single-component representations, together with the joint representation, are used in the definitions of perceptual interactions. Component A (B) is *perceptually separable* from B (A) if, and only if, the (marginal) perception of one component does not depend of the physical level of the other component—that is,

$$g_{A_iB_1}(x) = g_{A_iB_2}(x), \text{ for all } x \text{ and for } i = 1, 2$$

$$g_{A_1B_j}(y) = g_{A_2B_j}(y), \text{ for all } y \text{ and for } j = 1, 2.$$

In Figure 2, component A is perceptually separable from B, but B is not perceptually separable from A.

In addition to invariance in performance for the relevant component while the irrelevant component varies, absolute separability should also include the case in which it does not matter whether the opposing dimension exists

Figure 1. An example of the stimuli used in the experiments. On the left is a full face, and on the right is a face with a single component (in this case, eyes). The latter was used in Experiment 1 to assess existence separability.

Figure 2. Contours of equal likelihood for a possible perceptual representation of objects consisting of two dimensions sampled at two levels each, together with their marginal densities. Perceptual and decisional separability hold for component A, but not for B. Also, each stimulus exhibits positive correlation between the dimensions. The decision bounds are what might be used in identification.

or not (Garner, 1974; Townsend & Thomas, 1993), at least in such cases where this is possible. Ashby and Townsend (1986) claim that their definition of perceptual separability (PS) applies equally to the case in which features are either present or absent to the case in which the feature is always present but varies along a continuous dimension. When both possibilities exist within the same experiment, a distinction needs to be made, depending on the kind of dimensions under consideration. For prosthetic dimensions, such as size or brightness, the zero level corresponds to the absence of the dimension; hence, PS as presently defined is adequate. However, for metathetic attributes that vary continuously (such as orientation of a line), absence (e.g., of the line) is not the zero value on the dimension; it is something quite different. To capture this idea, a property termed *existence separability* is introduced. Existence separability of component A from B holds if and only if the marginal distribution for component A is invariant across levels of B and this distribution is the same whether B is present in the stimulus or not—that is, for all x , $g_{A,B_1}(y) = g_{A,B_2}(y) = g_A(y)$, for $i = 1, 2$. Note that existence separability implies PS, but not vice versa. Many experiments, for

example, have shown that circle size and line orientation enjoy separability when both the line and the circle are present (e.g., Garner & Felfoldy, 1970; Shepard, 1964; but see Kadlec & Hicks, 1998). However, size discrimination when the line is present (regardless of angle) is markedly different both in RTs and in accuracy from when the line is omitted and only the circle is present (Thomas, 1995b). As it turns out, in the present study, we will also need to differentiate between PS (when all components are present and varying) and existence separability (components present vs. absent), since the former will hold, but not the latter.

A final concept associated with perceptual interactions is that of *perceptual independence* (PI). Within a given stimulus A_iB_j , PI of components A and B occurs if and only if $f_{A,B_i}(x,y) = g_{A,B_i}(x)g_{A,B_i}(y)$; that is, the perceptual effects of the components are statistically independent. Unlike separability, PI is a symmetric, within-stimulus relation. When the distribution is multivariate normal, PI is equivalent to zero correlation between the dimensions. The resulting equal likelihood contours will be either circles or ellipses with their major and minor axes parallel to the coordinate axes in the perceptual space. When a de-

pendence exists, the ellipses will be tilted, as in Figure 2, in which all the stimuli exhibit a positive correlation.

Decision processes. The form of the decision rules depends on the nature of the task demands. In general, each response demanded by the task is attached to a region in perceptual space, so that if, on a trial, a percept (sampled from the presented stimulus distribution) falls in a given region, the observer emits the associated response. In an identification task, each stimulus is to be given a unique response, so the number of regions equals the number of stimuli. In classification, the number of regions is smaller than the number of stimuli. In other designs, such as the same–different judgment task, more complicated decision models need to be specified (Thomas, 1996). Response regions arise from a partitioning of the space by criteria called *decision bounds*. In the two-dimensional case, the decision bounds can be any one-dimensional (locally connected) subset of points.

One can assume that *decisional separability* (DS) simplifies the geometry of these bounds if the observer's decision about the level of one component does not depend on the perceived value of the other component. Thus, if component A is decisionally separable from B, there exists a decision criterion for component A, x_c , so that if, on a given trial, the perceptual effect for A, x , is to the left of x_c , the observer responds that component A is at Level 1; otherwise, he or she responds that A is at Level 2. DS always leads to decision criteria that are parallel to the perceptual axes in the case of classification and identification.¹ In some paradigms, such as the same–different task, a stronger notion is needed in which DS requires that separate judgments be made on each component and that those judgments are invariant with the perceived value of other components (e.g., Thomas, 1996; see also Olzak, 1986). In Figure 2, component A is decisionally separable from B, but not vice versa; this can be seen from the fact that, to decide if B is at Level 2, the observer would need a larger percept value on the y dimension if A were at Level 2 than if A were at Level 1. Logically, PS and DS are unrelated (Ashby & Townsend, 1986), but they may be tied together in real physical situations (Ashby & Lee, 1991).

In the GRT, the probability of responding R_k on trials given stimulus $A_i B_j$ is equal to the probability that a random sample from the perceptual distribution associated with stimulus $A_i B_j$ falls in the response region associated with the response R_k . In more formal terms, this probability is

$$P(R_k | A_i B_j) = \int \int_{\mathcal{R}_k} f_{A_i B_j}(x, y) dx dy, \quad (1)$$

where $f_{A_i B_j}(x, y)$ is the perceptual distribution (density) function for stimulus $A_i B_j$ and $\mathcal{R}_k$ is the region in the x – y plane attached to response R_k to be used in either classification or identification.

Motivation for the Experiments

To the problem of identifying the nature of the perceptual and decisional processes underlying object recogni-

tion, different approaches can be taken. One approach is to systematically develop parametric models that differ in their assumptions and fit them (i.e., estimate parameters) to obtained data to determine which GRT-based model yields the best overall performance relative to the number of freely varying parameters (Ashby & Lee, 1991). From this best-fitting and most parsimonious model, inferences regarding the perceptual process are immediate. A different methodological approach is to derive various theorems from basic axioms that relate properties holding among observable quantities, such as response probabilities or RTs to unobservable perceptual and/or decisional process assumptions (Ashby & Maddox, 1994; Ashby & Townsend, 1986; Kadlec & Townsend, 1992a; Thomas, 1999). Then, empirical data can be queried for the satisfaction of those properties. This latter approach affords a possible nonparametric analysis of the data and is usually implementable for simpler experimental designs than would be required for model fitting. Both types of analyses will be performed.

In the following two experiments, the technology of the GRT is applied in order to assess the underlying perceptual representations for schematic faces (Figure 1) that observers use when performing a speeded classification task (Experiment 1) and an identification task (Experiment 2) utilizing the same observers. These two tasks were chosen because, together, they can provide the most information regarding the perceptual representation of an object, which is the focus of this study. The speeded classification task was run before the identification task to encourage a DS strategy on the latter task, which is a useful assumption in modeling identification data. Because perception, and not decision, is the important issue here, methodological devices to ensure the simplest decision strategy are desirable (Olzak, 1986). In addition, RTs from classification can corroborate accuracy findings related to separability. Finally, it is much easier to evaluate existence separability (see above) within this task. The identification task, on the other hand, will yield additional information regarding any within-stimulus perceptual correlation and can provide convergent support for any of the separability conclusions obtained from speeded classification. Although it may be interesting for other reasons to model the joint performance on these tasks, which would necessarily rest on decisional processes and how they change from one task to another (Nosofsky, 1986), the present work's focus is on the perceptual representation of the face, irrespective of variations in optional decision processes.

The facial features of interest are the separation of the eyes, the nose length, and the width of the mouth. These features were chosen because, with the exception of hair, they are collectively the most important in recognition of a face (Bruce, 1988) and the role of spatial proximity of features in mediating any interaction can be assessed. Given the above review of the face recognition studies, it is expected that these dimensions will give rise to an interaction of some kind—that is, either integrality and/or dependence between facial attributes. The virtue of ap-

plying the technology of the GRT lies in identify the specific type of interaction.

EXPERIMENT 1

Speeded Classification

In several sessions, observers had to classify a presented face according to the value of one component. Three attributes of a schematic representation of the face (see Figure 1) were selected as those to be varied across two levels: the separation between the eyes, the length of the nose, and the width (horizontal length) of the mouth. Additional pseudofaces were included in the stimulus set, in which only one feature was drawn (with the outline). These single-feature stimuli were included to assess existence separability. For an individual observer, only two of the three attributes varied: eyes and nose or eyes and mouth. These attribute pairs were selected because they were either close in spatial proximity or farther apart in spatial proximity. The third, unvarying, feature was always present and set at the Level 1 value, except in the single-feature pseudofaces. It is hypothesized that facial features near in spatial proximity will be more likely to induce perceptual interactions (i.e., perceptual integrality) than those farther apart. The third possible attribute pair, nose and mouth, was not selected, since these are the least attended to parts of a face that is shown frontally (Fraser & Parker, 1986; Haig, 1984). The data collection and analyses were conducted on the individual observer level, since this has been shown to be the preferred method with respect to accurately characterizing performance (Ashby, Maddox, & Lee, 1994).

Method

Subjects. Two males and 2 females (all 20–25 years of age) were paid for their service. All had normal or corrected-to-normal vision.

Stimuli and Apparatus. Two of the three attributes (eye separation and nose length for Observers 1 and 4, eye separation and mouth for Observers 2 and 3) could vary across two levels each (the third attribute was held constant). Actual dimension values were chosen so that discrimination accuracy along one dimension was around 70%–80% in pilot work (Thomas, 2001). In addition, two faces for each relevant dimension condition were constructed by omitting all facial attributes other than the relevant dimension sampled at each of the two levels (see the right panel of Figure 1 for one such face example). Thus, for each type of classification, there were six possible stimuli for an individual observer: four full faces and two single-feature faces. The experiment was conducted on a PC with 14-in. SVGA color monitors (640 pixels wide $\times$ 400 pixels high $\times$ 256 colors). The stimuli were drawn in gray color ($R = 224, G = 224, B = 224$) on a white background and measured 6 cm top to bottom. The observers sat approximately 90 cm from the screen so that the stimulus subtended about 3.8° of visual angle.

Procedure. After one session of practice, each observer performed five sessions of classification (576 trials each), in which the observer was instructed to attend to one dimension and ignore any variation on the irrelevant one. For the first four sessions, the dimension that was deemed relevant changed for each session (order was counterbalanced across observers). Each of these first four sessions was divided into six blocks of 96 trials each so that the observer could rest between blocks. The last session was split into two halves, one for each dimensional classification, with three blocks of 96 tri-

als in each half. Within each block, each stimulus occurred with equal frequency. On each trial, a fixation cross appeared for an exponentially distributed (minimum = 300 msec, mean = 750 msec, cut off at 2,000 msec) amount of time, followed by the stimulus drawn at the center of the screen. The stimulus remained on the screen for no more than 105 msec, with the display terminating in the event of a response. If the observer did not make a response within 5 sec, the trial ended with no response recorded. The observer indicated his or her response by pressing a key on a four-key response box in which one of the middle two keys was marked "Level 1" and another of the middle two keys was marked "Level 2." The observer used the index finger of each hand for each response (mapping of the response to key was counterbalanced across observers). RT in milliseconds was measured from stimulus onset.

Theoretical Analyses and Results

Qualitative tests: Marginal RT invariance and marginal response invariance. Mean RTs (in milliseconds), standard deviations (*SDs*), and response frequencies for each observer are summarized in Table 1. Rows correspond to stimuli, and column groups (of three columns each) correspond to observer responses. Within each column group, three statistics are offered: mean RT (*MRT*), the *SD* for RT, and the frequency with which each response was given for that stimulus. Occasionally, the observer did not respond in the 5-sec trial window; hence, some row frequencies do not add to the total number of stimulus presentations (see the Method section). Because the statistical analyses of RTs will involve the cumulative distribution functions, no outliers were trimmed.

Ashby and Maddox (1994; Maddox & Ashby, 1996) extended the GRT to the speeded classification task by enlisting a processing time assumption termed the *RT-distance hypothesis* (Ashby, Boynton, & Lee, 1994; Ashby & Maddox, 1991, 1994; see also Murdock, 1985). Simply stated, identification time on a given trial is inversely related to the (Euclidean) distance from the percept to the (nearest) decision bound. Together with the RT-distance hypothesis, the PS and DS of stimulus components imply that the entire RT distribution for classifying a stimulus according to the value of a relevant dimension is invariant across physical levels of the irrelevant component. Formally, if R_{ak} denotes the event of reporting that component A is at Level k , assuming both PS and DS, for component A, they showed, for $i, k = 1, 2$ and for all $t > 0$,

$$P_{FA}(RT_{ak} \leq t | A_i B_1) = P_{FA}(RT_{ak} \leq t | A_i B_2), \quad (2A)$$

where the subscript FA denotes a filtering task in which the relevant component is A. Ashby and Maddox refer to Equation 2A as *marginal RT invariance*.²

Response probability can be used to assess separabilities in this paradigm as well. *Marginal response invariance* (MRI; Ashby & Townsend, 1986) holds if the probability of classifying component A at a given level does not depend on the physical level of component B; MRI holds when, for $i, k = 1, 2$,

$$P_{FA}(R_{ak} | A_i B_1) = P_{FA}(R_{ak} | A_i B_2). \quad (2B)$$

If one includes trials in which only component A is present (or B), existence separability can be assessed (see

Table 1
Response Times (MRTs, in Milliseconds), Standard Deviations (SDs), and Response Frequencies (Freq)
for Speeded Classification Judgments

		Observer 1						Observer 4					
Condition	Components	MRT	SD	Freq	MRT	SD	Freq	MRT	SD	Freq	MRT	SD	Freq
Eye Separation, Mouth Width		Close			Far			Close			Far		
Eyes relevant	A ₁ B ₁	987.3	356.6	160	1,136.0	474.5	80	625.4	162.0	209	805.3	186.0	31
	A ₁ B ₂	1,030.6	370.3	142	1,116.0	385.1	98	643.6	166.6	211	810.9	186.3	31
	A ₂ B ₁	1,069.8	388.8	119	1,018.7	395.0	121	820.7	207.2	33	658.6	141.7	206
	A ₂ B ₂	1,071.9	378.2	117	1,059.7	356.4	122	764.8	183.9	27	673.1	171.2	213
	A ₁	1,034.1	355.9	134	1,172.0	435.7	106	666.0	163.3	218	753.0	183.2	22
	A ₂	1,146.3	375.3	101	944.7	348.3	139	811.7	186.6	51	647.4	155.7	188
		Thin			Wide			Thin			Wide		
Mouth relevant	A ₁ B ₁	674.0	215.6	164	809.3	287.3	75	688.1	195.5	180	731.3	174.4	60
	A ₁ B ₂	759.7	327.5	58	728.8	269.6	181	799.0	238.0	62	713.5	211.7	177
	A ₂ B ₁	692.1	205.9	163	775.5	286.1	77	582.6	187.3	181	774.5	230.6	59
	A ₂ B ₂	729.2	202.3	70	715.4	220.2	170	752.2	224.2	73	685.6	175.3	165
	B ₁	641.0	216.5	171	850.5	356.3	69	698.9	213.6	195	789.1	190.5	45
	B ₂	715.8	242.1	83	753.5	274.8	155	792.6	241.3	115	757.5	211.6	125
		Observer 2						Observer 3					
		MRT	SD	Freq	MRT	SD	Freq	MRT	SD	Freq	MRT	SD	Freq
Eye Separation, Nose Length		Close			Far			Close			Far		
Eyes relevant	A ₁ B ₁	545.7	96.1	193	508.6	146.8	47	540.8	95.5	216	574.9	89.1	24
	A ₁ B ₂	534.9	93.4	181	529.4	136.1	59	535.6	81.8	198	582.0	77.3	42
	A ₂ B ₁	559.2	108.6	92	516.4	102.1	148	564.0	98.0	94	598.3	92.7	145
	A ₂ B ₂	597.4	227.9	84	549.8	140.5	156	555.9	101.9	80	585.5	89.7	160
	A ₁	632.2	147.5	75	493.1	95.2	164	578.4	106.2	132	576.3	102.8	108
	A ₂	600.0	183.6	8	483.9	113.8	232	505.1	83.2	12	554.7	84.0	228
		Short			Long			Short			Long		
Nose relevant	A ₁ B ₁	574.1	113.4	133	547.1	114.3	107	594.5	89.6	156	572.3	99.0	84
	A ₁ B ₂	603.7	151.9	75	525.4	87.8	165	612.9	106.1	85	574.0	89.3	155
	A ₂ B ₁	580.0	131.5	90	556.4	150.2	150	579.2	94.0	165	568.4	116.2	75
	A ₂ B ₂	577.9	133.6	60	537.9	107.5	180	584.3	120.3	63	568.2	83.1	177
	B ₁	528.0	111.0	168	595.8	127.5	72	562.3	84.8	168	581.1	95.5	72
	B ₂	539.5	88.8	124	575.0	135.0	116	568.5	101.8	119	599.2	100.4	121

Note—Rows correspond to stimuli, column groups (three columns each) correspond to observer responses.

above). If the observer's task is to classify the stimulus according to the level of A given the stimulus set A₁, A₂, A₁B₁, A₁B₂, A₂B₁, and A₂B₂, if existence separability holds for component A, DS holds on trials in which the stimulus contains both components, and the decision criterion on the *x* dimension is the same for the single-component trials as for the two-components trials, then, for *i, k* = 1, 2,

$$P_{FA}(R_{ak} | A_i) = P_{FA}(R_{ak} | A_i, B_1) \\ = P_{FA}(R_{ak} | A_i, B_2), \quad (3A)$$

and if the RT-distance hypothesis is correct for all *t* > 0,³

$$P_{FA}(RT_{ak} \leq t | A_i) = P_{FA}(RT_{ak} \leq t | A_i, B_1) \\ = P_{FA}(RT_{ak} \leq t | A_i, B_2). \quad (3B)$$

Equations 2 and 3, together with the signal detection analyses, will be used to assess all three kinds of separability—perceptual, decisional, and existence. Pairs of RT distributions for correct responses⁴ used to assess PS and

DS are presented in Figure 3 (A–D) for each observer and response type.

If, in addition, existence separability holds, the RT distribution for trials without the irrelevant component should equal the RT distributions with the component present (which should also equal each other; Equation 3B).⁵ The RT distribution to be compared with the single-feature case, the *combined* distribution, was found by combining the RTs for the given response type across both levels of the other feature and estimating a cumulative distribution from these RTs.⁶ The distributions for the single-feature and the combined-feature trials are graphed according to response type in Figure 4 (A–D). Tables 2A and 2B summarize marginal RT invariance tests.

The response frequencies of Table 1 were subjected to the MRI tests of Equations 2B and 3A, which are summarized in Table 3.⁷ They largely corroborate the results from the RT analyses.

Parametric analyses: Signal detection parameters. If one makes the additional assumption that the perceptual representations are Gaussian, (parametric) signal

Observer 1 Filtering RT Distributions
(Eye Separation/Mouth Width)

Figure 3A: Response time distributions for speeded classification judgments to test for perceptual separability for Observer 1.

Observer 2 Filtering RT Distributions
(Eye Separation/Nose Length)

Figure 3B: Response time distributions for speeded classification judgments to test for perceptual separability for Observer 2.

Observer 3 Filtering RT Distributions
(Eye Separation/Nose Length)

Figure 3C: Response time distributions for speeded classification judgments to test for perceptual separability for Observer 3.

Observer 4 Filtering RT Distributions
(Eye Separation/Mouth Width)

Figure 3D: Response time distributions for speeded classification judgments to test for perceptual separability for Observer 4.

Observer 2 Single Feature versus Combined
(Eye Separation/Nose Length)

Figure 4B: Response time distributions (single feature vs. combined) for speeded classifications to test for existence separability for Observer 2.

Observer 1 Single Feature versus Combined
(Eye Separation/Mouth Width)

Figure 4A: Response time distributions (single feature vs. combined) for speeded classifications to test for existence separability for Observer 1.

Figure 4D: Response time distributions (single feature vs. combined) for speeded classifications to test for existence separability for Observer 4.

Figure 4C: Response time distributions (single feature vs. combined) for speeded classifications to test for existence separability for Observer 3.

Table 2A
***p* Values for Difference in Response Time Distributions (Kolmogorov–Smirnov Tests):**
Perceptual Separability If (H_0 : Perceptual Separability)

	Observer 1 (A = Eyes, B = Mouth)	Observer 2 (A = Eyes, B = Mouth)	Observer 3 (A = Eyes, B = Mouth)	Observer 4 (A = Eyes, B = Mouth)
$P(RT_{A1} < t A_1B_1) = P(RT_{A1} < t A_1B_2)$	>.80	>.80	>.80	.281
$P(RT_{A2} < t A_2B_1) = P(RT_{A2} < t A_2B_2)$	.24	.42	>.80	>.80
$P(RT_{B1} < t A_1B_1) = P(RT_{B1} < t A_2B_1)$	.069	>.80	.19	>.80
$P(RT_{B2} < t A_1B_2) = P(RT_{B2} < t A_2B_2)$	>.80	>.80	>.80	>.80

Table 2B
***p* Values for Difference in Response Time Distributions (Kolmogorov–Smirnov Tests):**
Existence Separability If (H_0 : Existence Separability)

	Observer 1 (A = Eyes, B = Mouth)	Observer 2 (A = Eyes, B = Mouth)	Observer 3 (A = Eyes, B = Mouth)	Observer 4 (A = Eyes, B = Mouth)
$P(RT_{A1} < t A_1) = P(RT_{A1} < t A_1B_1 \text{ or } A_1B_2)$	>.80	<.0005	<.0005	.004
$P(RT_{A2} < t A_2) = P(RT_{A2} < t A_2B_1 \text{ or } A_2B_2)$	.009	<.0005	<.0005	.004
$P(RT_{B1} < t B_1) = P(RT_{B1} < t A_1B_1 \text{ or } A_2B_1)$	.005	<.0005	<.0005	>.80
$P(RT_{B2} < t B_2) = P(RT_{B2} < t A_1B_2 \text{ or } A_2B_2)$	.166	.153	.0198	.007

detection analyses of the speeded classification accuracy results can also be used to assess both types of separability much in the same way that identification results are studied using the macroanalyses of Kadlec and Townsend (1992a, 1992b; Kadlec & Hicks, 1998; Macmillan & Ornstein, 1998). Parametric analyses are often of higher statistical power, so they are useful when the null hypothesis has theoretical meaning (in this case, separability is the null hypothesis). The perceptual distribution for a given (bivariate) stimulus, assuming normality, has five parameters: location parameters (mean on each dimension—i.e., μ_X and μ_Y), spread parameters (variance on each dimension, σ_X^2 and σ_Y^2), and an association parameter (correlation between dimensions, ρ). For trials in which the observer is to classify on the basis of one component (designated as component A), where the other component (component B) is held constant at Level 1 (short), denote the sensitivity index by d'_{AB_1} , which is the distance in *SD* units (assuming equal variance on the x perceptual dimension) between $\mu_{x_{A_1B_1}}$ and $\mu_{x_{A_2B_1}}$. For trials in which nose length is constant at Level 2 (long), the corresponding sensitivity index is d'_{AB_1} , which is the distance between $\mu_{x_{A_1B_2}}$ and $\mu_{x_{A_2B_2}}$, the perceptual mean on the x dimension for stimulus A_1B_2 and A_2B_2 , respectively, in common *SD* units. Under the joint hypotheses of DS and PS and equal variance on the x dimension—that is, $\sigma_{x_{A_1B_1}}^2 = \sigma_{x_{A_2B_1}}^2$, it is easy to show that $d'_{AB_1} = d'_{AB_2}$ (Kadlec & Townsend, 1992a). If one defines

$$d'_{AB_j} = \frac{\mu_{x_{A_1B_j}} - \mu_{x_{A_2B_j}}}{\sqrt{\frac{1}{2}(\sigma_{x_{A_1B_j}}^2 + \sigma_{x_{A_2B_j}}^2)}},$$

for $j = 1, 2$, one can drop the assumption that $\sigma_{x_{A_1B_j}}^2 = \sigma_{x_{A_2B_j}}^2$; however, this new d' is not estimable from data with only one degree of freedom.

Corroborating evidence for the two separabilities can also be obtained by estimation of the location of the decision criterion for each component. In signal detection analyses, this is usually expressed in terms of a bias measure β_{AB_j} , the ratio of the two marginal densities for component A,

$$\frac{g_{A_2B_j}(x)}{g_{A_1B_j}(x)},$$

at $x = x_c$, the decision criterion for component A. A bias to say that A is at Level 1 would be evidenced by

$$\beta_{AB_j} = \frac{g_{A_2B_j}(x_c)}{g_{A_1B_j}(x_c)} > 1;$$

that is, x_c is to the right of the point where the marginal densities cross. However, a better measure of bias, denoted by c_{AB_j} , is the signed distance in common standard deviation units between x_c and the x value where the two marginal densities cross, say x_0 . This bias measure is functionally related to β_{AB_j} by $\log(\beta_{AB_j}) = c_{AB_j}d'_{AB_j}$. Macmillan and Creelman (1990) have provided a number of reasons why this signal detection measure of bias is to be preferred over other bias measures based on signal detection theory (SDT), including the likelihood ratio, β_{AB_j} , not the least of which is that a statistical test exists for c_{AB_j} , but not for β_{AB_j} (although for some recent developments on this issue, see Kadlec, 1999b).⁸ Both measures essentially provide the same kind of information. For example, if no bias exists, β_{AB_j} and c_{AB_j} will be at the appropriate values (equal to 1 and 0, respectively). Figure 5 shows a pair of marginal densities for component A, with B at Level 1 with the various decision criteria x_c , c_{AB_1} , and β_{AB_1} illustrated.

Kadlec and Townsend (1992a) show that under the *joint* hypothesis of PS and DS, $\beta_{AB_1} = \beta_{AB_2}$. Because we already

Table 3
Proportion of Level 1 Responses and
Marginal Response Invariance Test Results for Filtering Task

Condition	Feature Pair			
	Eye Separation, Mouth Width (A = Eyes, B = Mouth Width)		Eye Separation, Nose Length (A = Eyes, B = Nose Length)	
	Observer 1	Observer 4	Observer 2	Observer 3
Perceptual Separability Tests				
	a_1	a_1	a_1	a_1
Component A relevant				
A_1B_1	0.667	0.871	0.804	0.900
A_1B_2	0.592	0.879	0.754	0.825
z	1.701	-0.276	1.320	2.386
p value	.089	.783	.187	.017
A_2B_1	0.496	0.138	0.383	0.393
A_2B_2	0.490	0.113	0.350	0.333
z	0.138	0.845	0.758	1.365
p value	.890	.398	.449	.172
	b_1	b_1	b_1	b_1
Component B relevant				
A_1B_1	0.686	0.750	0.554	0.650
A_2B_1	0.679	0.754	0.375	0.688
z	0.165	-0.106	3.935	-0.873
p value	.869	.916	.000	.383
A_1B_2	0.243	0.259	0.313	0.354
A_2B_2	0.292	0.307	0.250	0.263
z	-1.212	-1.147	1.523	2.174
p value	.226	0.251	.128	.030
Existence Separability Tests				
	a_1	a_1	a_1	a_1
Component A relevant				
A_1B_*	0.629	0.875	0.779	0.863
A_1	0.558	0.908	0.314	0.550
z	1.824	-1.357	11.808	8.678
p value	.068	.175	.000	.000
A_2B_*	0.493	0.125	0.367	0.363
A_2	0.421	0.213	0.033	0.050
z	1.824	-2.967	10.541	9.783
p value	.068	.003	.000	.000
	b_1	b_1	b_1	b_1
Component B relevant				
A_*B_1	0.683	0.752	0.465	0.669
B_1	0.713	0.813	0.700	0.700
z	-0.821	-1.852	-6.038	-0.851
p value	.412	.064	.000	.395
A_*B_2	0.267	0.283	0.281	0.308
B_2	0.349	0.479	0.517	0.496
z	-2.226	-5.103	-6.081	-4.837
p value	.026	.000	.000	.000

Note— p values < .05 are in boldface.

know that these separabilities imply $d'_{AB_1} = d'_{AB_2}$, we can then conclude that $c_{AB_1} = c_{AB_2}$.

Existence separability, under the assumption of Gaussian distributions, can also be assessed via SDT analyses. If existence separability for component A holds, then $d'_A = d'_{AB_1} = d'_{AB_2}$.⁹ If, in addition, DS for component A holds in the two-component case and the decision bound for the single-component trials is the same as the decision bound for A in the two-component case (i.e., x_c), then $\beta_A = \beta_{AB_1} = \beta_{AB_2}$, and thus, because of the previous proposition, $c_A = c_{AB_1} = c_{AB_2}$.¹⁰

Parameters of signal detection and their variances were estimated in the usual way (Macmillan & Creelman, 1990).

Appendix A provides details of the computation and construction of the relevant test statistics. Table 4 presents the parameter estimates of sensitivity (d') and decision criteria (β and c) for the component classifications for each observer. Listed separately are the results relevant to assessing PS in the context of the whole face and to assessing existence separability where classification of single-feature stimuli is compared with the combined classification of whole faces.

Discussion

Two conclusions can be drawn. First, the RT analyses, the MRI tests, and the signal detection analyses provided

Figure 5. The different decision criteria for unidimensional signal detection theory. See the text for their definitions.

evidence consistent with PS and DS when all the features of the face were present in the stimulus. In 11 out of the 16 RT comparisons (Table 2A), the p value for accepting the null hypothesis was greater than .80, and all but one were greater than .19. Because statistical power is always greater than the obtained p value (Lehman, 1983), confidence in accepting the null seems warranted.¹¹ Only one of the MRI tests of PS yielded significance (Observer 2, nose length relevant and short). Only one of the sensitivity parameter comparisons (nose length classification for Observer 3; Table 4) yielded a significant difference, and only two decision criteria comparisons differed (Observer 3, eye separation, and Observer 2, nose length; Table 4).

Second, existence separability can be rejected for both pairs of features. In only two of the RT tests was the p value greater than .2 (i.e., favoring the null), and most (12 of 16) were less than .025. Similarly, more than half of the MRI tests between the single-feature and the combined-feature cases failed. Inspection of the signal detection results, which have higher power than the nonparametric MRI tests (Kadlec, 1993), suggests that most of these differences may be due to shifts in criteria when comparing features in the context of a whole face versus in isolation.

All decision criteria comparisons of whole face versus single feature differed significantly (all $ps < .01$). Some of the sensitivity measures differed as well (Table 4).

EXPERIMENT 2 Identification

The speeded classification test can yield information about separability but does not speak to PI, the within-stimulus source of interaction. Equations 2 and 3, as well as the marginal detection equalities of the signal detection macroanalysis, hold regardless of any underlying correlation between the dimensions. For this information, one needs to turn to the identification task, in which a larger number of degrees of freedom in the data are available. This task can provide convergent evidence regarding the result of separability as well.

In this experiment, the same stimulus sets were used without the single-feature faces, with the same observers now making identifications for each of the four faces in their respective sets. The same individuals were used in order to keep the perceptual and decisional representations assessed in Experiment 1 constant for Experiment 2. Also,

Table 4
Signal Detection Estimates for Filtering Task and Significance Levels for Equality of Parameters

Feature Pair	Condition	d'	β	c	d'	β	c	
Perceptual Separability Tests								
Eye separation, mouth width (A = eyes, B = mouth width)	Eye separation relevant	Observer 1			Observer 4			
		AB ₁	0.443	1.097	0.209	2.219	1.051	0.020
	AB ₂	0.264	1.026	0.101	2.384	0.950	-0.021	
	p for difference in d'	.326			.221			
	p for difference in c	.250			.534			
	Mouth width relevant	A ₁ B	1.188	0.883	-0.105	1.320	1.018	0.014
		A ₂ B	1.013	0.962	-0.039	1.193	1.112	0.091
	p for difference in d'	.332			.454			
	p for difference in c	.448			.366			
	Eye separation, nose length (A = eyes, B = nose length)	Eye separation relevant	Observer 2			Observer 3		
AB ₁			1.157	1.380	0.278	1.550	2.187	0.505
AB ₂		1.077	1.174	0.148	1.365	1.412	0.253	
p for difference in d'		.646			.251			
p for difference in c		.135			.002*			
Nose length relevant		A ₁ B	0.624	0.894	-0.177	0.759	1.005	0.006
		A ₂ B	0.355	0.842	-0.495	1.125	0.924	-0.070
p for difference in d'		.103			.035*			
p for difference in c		.001*			.369			
Existence Separability Test								
Eye separation, mouth width (A = eyes, B = mouth width)	Eye separation relevant	Observer 1			Observer 4			
		A	0.347	0.992	-0.026	2.299	1.758	0.267
	AB (combined)	0.348	1.056	0.156	2.125	1.001	0.001	
	p for difference in d'	.993			.143			
	p for difference in c	.022*			<.001*			
	Mouth width relevant	B	0.950	1.082	0.082	0.939	1.482	0.419
		AB (combined)	1.096	0.923	0.073	1.255	1.070	0.054
	p for difference in d'	.332			.040			
	p for difference in c	.036*			<.001*			
	Eye separation, nose length (A = eyes, B = nose length)	Eye separation relevant	Observer 2			Observer 3		
A			1.348	0.290	-1.160	1.780	0.258	-0.760
AB (combined)		1.110	1.269	0.214	1.441	1.707	0.371	
p for difference in d'		.049			.01*			
p for difference in c		.001*			<.001*			
Nose length relevant		B	0.477	1.145	0.282	0.533	1.147	0.254
		AB (combined)	0.490	0.849	-0.334	0.937	0.970	-0.032
p for difference in d'		.925			.010			
p for difference in c		.001*			<.001*			

* $p < .05$.

because DS will be assumed for the modeling analyses, Experiment 1 preceded Experiment 2 in the hope that the prior speeded classification judgments will encourage DS in the identification task (Goldstone, 1995). Some indirect tests using signal detection analyses analogous to those described in the results of Experiment 1 can provide support for this assumption (Kadlec & Hicks, 1998).

Method

Subjects. The same 4 observers as those from Experiment 1 participated in Experiment 2.

Stimuli and Apparatus. The four full-featured faces used for a given observer in Experiment 1 were used here. For example, the stimulus set for Observer 1 included faces with close eyes and thin mouth, close eyes and wide mouth, far eyes and thin mouth, and far eyes and

wide mouth. The nose was always present but unchanging in all the stimuli for Observers 1 and 4. The mouth was always present but unchanging for Observers 2 and 3. The same computers, displays, and response boxes as those used in Experiment 1 were used here.

Procedure. Each observer served in one practice session and two experimental sessions. Each session consisted of six blocks of 92 trials each (total, 552 trials/session). Each stimulus occurred with equal frequency. On each trial, a variable foreperiod (exponential, with a minimum of 300 msec, a mean of 750 msec, up to 1,700 msec) containing a fixation cross drawn at the center of the screen was followed by the stimulus. The stimulus was on for 105 msec or less if a response occurred. If the observer did not make a response within 5 sec, the trial ended with no response recorded. Each face was associated with one of four buttons on a response box, which the observer had to press to indicate the identity of the face. The two leftmost buttons were to be pressed by the middle and index fingers of the left hand (going from left to right), whereas the two rightmost

Table 5
Observed and Predicted Identification Confusions

Stimulus	Eye Separation, Mouth Width							
	Observer 1				Observer 4			
	1	2	3	4	1	2	3	4
1(1,1)*	78	66	73	55	159	74	34	9
	80	59	76	56	162	76	28	9
2(1,2)	59	65	45	105	86	155	11	24
	56	71	41	106	82	156	16	21
3(2,1)	69	49	87	66	23	16	165	72
	66	52	83	70	23	14	168	71
4(2,2)	40	66	42	126	9	25	90	152
	43	62	46	122	10	27	87	151

Stimulus	Eye Separation, Nose Length							
	Observer 2				Observer 3			
	1	2	3	4	1	2	3	4
1(1,1)	93	115	55	13	148	74	42	12
	97	111	49	18	138	80	44	12
2(1,2)	42	167	31	36	78	138	28	33
	49	159	33	35	77	142	29	27
3(2,1)	17	29	126	104	30	23	147	76
	14	34	132	95	32	27	151	65
4(2,2)	9	41	82	144	11	55	93	117
	6	42	76	151	15	44	99	125

Note—Rows correspond to stimuli and columns to responses. Top row: observed frequency; bottom row: predicted frequency of most efficient GRT model. *(Level on first feature, level on second feature).

buttons were assigned the index and middle fingers of the right hand. The response and RTs (which are not analyzed¹²) were measured.

Theoretical Analyses and Results

In an identification task, the data to be predicted by a model are response probabilities (or equivalently, response frequencies) given each stimulus (see Equation 1). The observed response probabilities (or frequencies) are summarized in an identification-confusion matrix in which the rows correspond to the presented stimuli and the columns denote the observer responses. Identification-confusion matrices for all 4 observers are shown in Table 5. Again, occasionally, the observer did not respond within the 5-sec trial window, leading to some rows summing to less than 276, the total number of stimulus presentations (see the Method section). Two types of analyses on these data were performed. First, two qualitative tests, one assessing PI and the other assessing PS (both assuming DS), were evaluated nonparametrically. Second, parametric analyses using signal detection estimates and hierarchical model fitting were conducted on each observer's data. The latter yields the most efficient GRT model from which conclusions regarding separability and independence can be drawn. Analysis of the data from Observer 1 should be interpreted with caution, since this observer had difficulty discriminating the eye separation in this task (overall accuracy was 33%, chance is at 25%).

Qualitative tests for independence and separability: Sampling independence and marginal response invariance. Ashby and Townsend (1986) proposed two analyses of identification data that can be used to assess hypothesized perceptual properties. The *sampling independence (SI) test* confirms PI and DS. SI is a property

of the data (rather than a theoretical construct) and can be evaluated directly from the confusion matrix. SI in stimulus A_iB_j holds if and only if

$$P(R_{a_1b_1} | A_iB_j) = [P(R_{a_1b_1} | A_iB_j) + P(R_{a_1b_2} | A_iB_j)] \\ \times [P(R_{a_1b_1} | A_iB_j) + P(R_{a_2b_1} | A_iB_j)], \quad (4)$$

where $R_{a_i b_j}$ denotes that response $a_i b_j$ was given. The conditional probabilities for other $a_i b_j$ are defined analogously. In Theorem 1 of Ashby and Townsend, PI within A_iB_j and DS are shown to imply that SI holds for A_iB_j for all responses. Consequently, if SI is found in the data, support for PI is obtained. Conversely, if SI fails, PI and/or DS fails. Given that DS may be verified from other tests, such a failure might implicate a perceptual dependence.¹³ Table 6 gives the outcomes for the SI test for each stimulus and observer.

PS and DS can be assessed by using the MRI test from above. MRI, in an identification-confusion matrix, holds when the probability of correctly recognizing one component does not depend on the physical level of the other—that is, for $i = 1, 2$,

Table 6
Obtained p Values for Sampling Independence Tests

Stimulus	Eyes/Mouth Width		Eyes/Nose Length	
	Observer 1	Observer 4	Observer 2	Observer 3
A_1B_1	.635	.156	< .0001	.111
A_1B_2	< .0001	.622	< .0001	.172
A_2B_1	.467	.186	.027	.203
A_2B_2	< .0001	.222	.013	< .0001

Note— H_0 : Sampling independence holds. Small p value are evidence in favor of perceptual correlation.

Table 7
Proportion of Level 1 Responses and
Marginal Response Invariance Test Results for Identification Task

Condition	Feature Pair			
	Eye Separation, Mouth Width (A = Eyes, B = Mouth Width)		Eye Separation, Nose Length (A = Eyes, B = Nose Length)	
	Observer 1	Observer 4	Observer 2	Observer 3
Component A	<u>a₁</u>	<u>a₁</u>	<u>a₁</u>	<u>a₁</u>
A ₁ B ₁	0.529	.0844	0.754	0.804
A ₁ B ₂	0.453	0.873	0.757	0.779
z	1.796	−0.978	−0.099	0.734
p value	.072	.328	.921	.463
A ₂ B ₁	0.435	0.141	0.167	0.192
A ₂ B ₂	0.387	0.123	0.181	0.239
z	1.152	0.628	−0.449	−1.346
p value	.249	.530	.653	.178
Component B	<u>b₁</u>	<u>b₁</u>	<u>b₁</u>	<u>b₁</u>
A ₁ B ₁	0.555	0.699	0.536	0.688
A ₂ B ₁	0.576	0.681	0.518	0.641
z	−0.482	0.460	0.426	1.172
p value	.630	.645	.670	.241
A ₁ B ₂	0.380	0.351	0.264	0.384
A ₂ B ₂	0.299	0.359	0.330	0.377
z	1.985	−0.178	−1.677	0.175
p value	.047	.859	.094	.861

Note—*a*₁ responses are composed of *a*₁*b*₁ and *a*₁*b*₂ responses; *b*₁ responses are composed of *a*₁*b*₁ and *a*₂*b*₁ responses. Comparisons significant at the .05 level are in boldface.

$$P(R_{a_1b_1} | A_iB_1) + P(R_{a_1b_2} | A_iB_1) \\ = P(R_{a_1b_1} | A_iB_2) + P(R_{a_1b_2} | A_iB_2), \quad (5A)$$

and for $j = 1, 2$,

$$P(R_{a_1b_j} | A_1B_j) + P(R_{a_2b_j} | A_1B_j) \\ = P(R_{a_1b_j} | A_2B_j) + P(R_{a_2b_j} | A_2B_j). \quad (5B)$$

In Theorem 5, Ashby and Townsend (1986) show that PS and DS for both components A and B imply that MRI holds (both Equations 5A and 5B).¹⁴ Table 7 gives the results of the MRI tests.

Parametric analyses: Macrosignal detection and hierarchical GRT model fitting of identifications. Two types of parametric analyses assuming normality are employed here: the macroanalyses of Kadlec and Townsend (1992a; Kadlec & Hicks, 1998), which will assess PS and DS together, and hierarchical GRT model fitting, which will provide estimates for the values of the detection parameters, including correlation.¹⁵ The latter is especially of interest, given the finding of a failure of SI. The macroanalyses will provide convergent support for DS that was suggested by the MRI tests, an assumption that will be used in the model-fitting analyses. Because the SDT tests are parametric, they afford higher statistical power than the nonparametric tests (Kadlec, 1993), which is useful given the acceptance of separability.

Macro SDT analysis. This set of analyses of performance in the identification task is exactly analogous to the signal detection analysis described in Experiment 1 for the speeded classification task. The principal differ-

ence is how the response cells of the confusion matrix are combined to arrive at the appropriate hit and false alarm rates in each of the d' and c comparisons. Recall that, under the joint hypotheses of DS and PS for component A, $d'_{AB_1} = d'_{AB_2}$, and $c_{AB_1} = c_{AB_2}$, and similarly for component B. Consider the computation of d'_{AB_2} , the (standardized) difference between the marginal density $g_{A_1B_2}(x)$ and $g_{A_2B_2}(x)$. If a hit is defined as correctly responding that A was at Level 2, identification responses $R_{a_2b_1}$ and $R_{a_2b_2}$ are considered hits on component A. Hence, the hit rate is the observed proportion $\hat{P}(R_{a_2b_1} \text{ or } R_{a_2b_2} | A_2B_2)$, and the false alarm rate is $\hat{P}(R_{a_2b_1} \text{ or } R_{a_2b_2} | A_1B_2)$. The detection parameters computed from the identification responses can be subjected to the same statistical tests as those described Appendix A. Table 8 lists the detection parameter estimates and the results of the statistical tests for equality needed to assess the two separabilities for each component.

Hierarchical modeling. Fitting a specific GRT model to an observed identification-confusion matrix consists of searching for values of the model's parameters so that the predicted probabilities (Equation 1) maximally match the observed response probabilities or frequencies (Wickens, 1982). Appendix B outlines the technical details of this, together with information regarding computing various statistical indices used in making inferences for nested and nonnested models.

The logic of hierarchical model fitting, as applied to the problem of understanding dimensional interactions, is as follows. To make the discussion concrete, consider the present two-stimulus component case, A = eye sep-

Table 8
Marginal Detection Parameter Estimates for Identification Task and
Significance Levels for Equality of Parameters

Feature Pair	Condition	d'	β	c	d'	β	c
Eye separation, mouth width (A = eyes, B = mouth width)	Eye separation relevant	Observer 1			Observer 4		
	AB ₁	0.236	0.989	−0.044	2.086	0.937	−0.313
	AB ₂	0.168	0.966	−0.203	2.301	0.980	−0.009
	p for difference in d'	.683			.100		
	p for difference in c	.057			.729		
	Mouth width relevant						
	A ₁ B	0.445	0.963	−0.084	0.904	1.060	0.070
	A ₂ B	0.717	0.887	−0.168	0.833	0.050	0.054
	p for difference in d'	.100			.669		
	p for difference in c	.309			.847		
Eye separation, nose length (A = eyes, B = nose length)	Eye separation relevant	Observer 2			Observer 3		
	AB ₁	1.650	0.792	−0.141	1.730	0.989	−0.007
	AB ₂	1.610	0.8424	−0.107	1.480	1.040	0.030
	p for difference in d'	.892			.093		
	p for difference in c	.837			.0624		
	Nose length relevant						
	A ₁ B	0.721	0.824	−0.269	0.786	1.080	0.098
	A ₂ B	0.490	0.908	−0.198	0.676	1.020	0.024
	p for difference in d'	.141			.511		
	p for difference in c	.368			.376		

aration (near or far apart) and B = nose length (short or long). The most general GRT model (assuming DS) allows all perceptual parameters and both decision criteria to vary freely. For four stimuli, this model has 22 free parameters (5 for each stimulus plus 2 decision param-

eters). However, the identification-confusion matrix has only 12 freely varying data cells¹⁶ so this model would perfectly fit the data and, hence, be untestable. Consequently, a good place to start is at the most restrictive (reasonable) GRT model: one that assumes both PI and

Figure 6. The hierarchy of possible general recognition theory models for a 4×4 confusion matrix. The most (interesting) restricted model is at the top; models are nested within other models that are connected by a downward arrow. The most general models possible are at the bottom of the graph. The number of parameters for each model is in parentheses.

PS on both dimensions. From this point, different GRT models can be developed in which one or more of the restrictive assumptions are relaxed. Figure 6 depicts a hierarchy of possible GRT models that can be fit to a 16-cell confusion matrix of response frequencies.

At the top of the figure is the GRT model that assumes PS and PI. From that, if the assumption of PS on one of the dimensions—say, nose length (B)—is dropped, the more general model PS_B, PI , would result. Alternatively, the independence assumption can be relaxed by allowing a correlation between the perceptual dimensions that is the same value in all of the stimulus representations (PS, one correlation, ρ). The models in the upper part of Figure 6 are nested within models in the lower part if they can be connected by traveling down a path of arrows. That is, an upper model that is connected to a model in the lower part of the figure is a special case of the latter, in which some of the parameters have been fixed. The models at the bottom of the figure are the most general possible, given the number of degrees of freedom in the data. The two models on the lower rightmost branch are included for completeness but are difficult to interpret psychologically. One assumes an equal difference in means for one dimension across levels of the other (for both dimensions) and one common variance. This model allows the correlation to be different across stimuli. The other model simply assumes that the means form a rectangle in the perceptual space. For all models, without loss of generality, the mean of stimulus A_1B_1 is assumed to be (0, 0), and the variances are set equal to 1.¹⁷

This nesting of models allows us to determine whether the assumptions of PS and PI are warranted. This is possible because one can test whether relaxing the assumptions (by allowing the relevant parameters to vary) significantly improves the fit of the model to the data (see Appendix B). Hierarchical modeling analyses were performed on the identification data of all 4 observers. Table 9 gives the fit statistics, including percentage of variance accounted for given each model in the hierarchy in Figure 6. For models that are nested within another, results from the chi-square test described in Appendix B are listed in Table 10.

To choose the best model for an individual observer, we start with the model that has the smallest Akaike's (1974) information criterion (AIC; See Appendix B). Following this choice, the χ^2 tests were consulted to determine whether this model significantly improved fit over models nested within it ($p < .1$). When the likelihood ratio tests were combined with AIC, the best model for Observers 2, 3, and 4 assumed PS on both dimensions and a correlation between the features that was constant across faces. This model was especially successful for the Observers 2 and 3, whose relevant features were eye separation and nose length. The improvement was only marginally significant for Observer 4, whose features were farther apart on the face (eye separation and mouth width). For Observer 1, the model that assumed an equal difference in means for one dimension across lev-

els of the other and one common variance best described the data. This model allowed the correlation to be different across stimuli. However, because accuracy was low for this observer, caution should be given when interpreting the model fits. Contours of equal likelihood for the best model are graphed in Figures 7A–7D for each observer. From these figures and associated parameter estimates in Table 11, the nature of the perceptual representation of these faces, as experienced in this task, can be seen.

Discussion

SI fails for the features of eye separation and nose length (Observers 2 and 3) and is marginally rejected for eye separation and mouth width. This implies that either PI or DS does not hold. MRI invariance is strongly supported for both pairs of components for most observers (except Observer 1, whose accuracy was very low). This finding corroborates the conclusions of PS and DS found in Experiment 1. Separability is also strongly supported by the failure of any of the relevant detection parameter comparisons to yield significance. Given these latter two observations, the failure of SI is likely due to the presence of a perceptual dependence, especially for eye separation and nose length, rather than a failure of DS.¹⁸ The model fits strongly supported PS, in agreement with the other analyses. Table 10 illustrates that the central source of the improvement in fit for some of the more general models is the relaxation of the PI assumption, which, as is shown in Table 11, was due to a negative correlation between the separation of the eyes and the other manipulated feature, nose length or mouth width. Intuitively, this means that the farther apart the eyes appear for an individual face (variability is due to perceptual noise), the shorter the nose or the narrower the mouth is sampled to be. The largest correlations were for Observers 2 and 3. A smaller correlation between eye separation and mouth width was found for Observer 4. Thus, a perceptual interaction, in the form of a perceptual dependence, in fact did emerge for the facial features used in the present experiments.

GENERAL DISCUSSION

The speeded classification judgments of Experiment 1 indicated that the PS and DS for pairs of facial features—eye separation and nose length or eye separation and mouth width—held when accuracy and RTs were analyzed. For the latter measure, the entire cumulative distribution function was examined, in which any difference (e.g., variance) between RTs for judgments along one dimension across levels of the other could be detected. Existence separability, which evaluates invariance of perception of one attribute when the presence or absence of the other attribute varies, did not hold for either dimension pair. PS and DS were confirmed by the nonparametric tests, the signal detection analyses, and the model fits performed on the identification data, using the same observers and stimuli as those in Experiment 2. The iden-

Table 9
Summary of Fit Statistics for the General Recognition Theory Models, Experiment 2

Model	No. of Parameters	Observer 1 (A = Eye Separation, B = Mouth Width)			Observer 4 (A = Eye Separation, B = Mouth Width)			Observer 2 (A = Eye Separation, B = Nose Length)			Observer 3 (A = Eye Separation, B = Nose Length)		
		-InL	AIC	%var	-InL	AIC	%var	-InL	AIC	%var	-InL	AIC	%var
PS, PI	4	1,473.7	2,955.4	77.53	1,144.7	2,297.4	99.60	1,283.1	2,574.1	91.95	1,292.5	2,592.9	96.92
PS _A , PI	7	1,471.7	2,957.5	81.22	1,144.6	2,303.1	99.59	1,281.7	2,577.4	93.23	1,292.1	2,598.	96.46
PS _B , PI	7	1,471.4	2,956.8	85.06	1,143.9	2,301.8	99.52	1,283.0	2,580.0	92.10	1,291.8	2,597.6	97.26
PS, one ρ	5	1,470.6	2,951.2	84.27	1,143.1	2,296.1	99.76	1,257.3	2,524.5	98.85	1,283.4	2,576.9	98.31
PS _A , one ρ	8	1,468.6	2,953.2	87.52	1,143.0	2,301.9	99.73	1,257.2	2,530.4	98.75	1,282.6	2,581.2	98.73
PI	10	1,469.3	2,958.6	88.22	1,143.8	2,307.6	99.70	1,281.6	2,583.2	93.33	1,291.1	2,602.2	97.14
PS _B , one ρ	8	1,468.2	2,952.4	89.53	1,142.5	2,301.1	99.82	1,257.2	2,530.3	98.93	1,282.4	2,580.8	98.69
PS	8	1,466.9	2,949.9	91.95	1,141.8	2,299.7	99.86	1,254.6	2,525.3	99.32	1,280.4	2,576.9	99.30
PS _A	11	1,464.4	2,950.9	95.64	1,141.6	2,305.1	99.89	1,252.5	2,527.0	99.97	1,279.4	2,580.9	99.62
One ρ	11	1,466.1	2,954.2	93.42	1,142.2	2,306.5	99.87	1,255.5	2,532.9	99.51	1,281.6	2,585.1	99.17
PS _B	11	1,464.5	2,951.0	96.25	1,141.0	2,303.9	99.95	1,253.9	2,529.7	99.41	1,279.1	2,580.2	99.58
Equal μ , $\sigma = 1$	10	1,463.8	2,947.6	97.06	1,141.6	2,303.3	99.86	1,253.7	2,527.4	99.41	1,279.5	2,579.1	99.52
μ rectangle	10	1,464.5	2,949.1	95.53	1,141.6	2,303.1	99.57	1,252.5	2,524.9	99.96	1,278.8	2,577.7	99.73

Note — -InL, minus log likelihood; AIC, Akaike's information criterion; %var, percent of variance accounted for by model. For the GRT models: PI, perceptual independence, PS, perceptual separability (when followed by a letter, PS on that dimension only). One ρ assumes a common, equal correlation between dimensions across all stimuli, Equal μ , $\sigma = 1$ assumes an equal distance between means on a given dimension across levels of the other dimension with all variances = 1, μ rectangle assumes the means form a rectangle, but the variances are allowed to vary. Decisional separability and equality of variance on a given dimension across levels of that same dimension are assumed throughout. Observers who experienced the same set of dimensions are grouped together (e.g., Observers 1 and 4).

Table 10
 χ^2 Tests for Nested General Recognition Theory

Restricted	General	Observer 1 (Eye Separation, Mouth Width)			Observer 4 (Eye Separation, Mouth Width)			Observer 2 (Eye Separation, Nose Length)			Observer 3 (Eye Separation, Nose Length)		
		χ^2	df	p value	χ^2	df	p value	χ^2	df	p value	χ^2	df	p value
PS, PI	PS _A , PI	3.93	3	.269	0.24	3	.972	2.79	3	.425	0.74	3	.863
PS, PI	PS _B , PI	4.58	3	.205	1.55	3	.671	0.17	3	.982	1.30	3	.730
PS, PI	PS, one ρ	6.15	1	.013*	3.23	1	.072†	51.62	1	.000***	18.02	1	.000***
PS _A , PI	PS _A , one ρ	6.27	1	.012*	3.21	1	.073†	48.93	1	.000***	18.95	1	.000***
PS _A , PI	PI	4.81	3	.186	1.55	3	.671	0.17	3	.982	1.98	3	.577
PS _B , PI	PI	4.16	3	.245	0.24	3	.972	2.79	3	.425	1.43	3	.700
PS _B , PI	PS _B , one ρ	6.39	1	.011*	2.76	1	.097†	51.63	1	.000***	18.84	1	.000***
PS, one ρ	PS _A , one ρ	4.05	3	.256	0.22	3	.974	0.10	3	.992	1.67	3	.643
PS, one ρ	PS _B , one ρ	4.83	3	.185	1.08	3	.782	0.19	3	.979	2.12	3	.548
PS, one ρ	PS	7.36	3	.061†	2.45	3	.484	5.25	3	.155	6.02	3	.111
PS _A , one ρ	PS _A	8.30	3	.040*	2.81	3	.422	9.46	3	.024*	6.37	3	.095†
PS _A , one ρ	One ρ	5.01	3	.171	1.44	3	.696	3.49	3	.322	2.09	3	.554
PI	One ρ	6.47	1	.011*	3.10	1	.078†	52.25	1	.000***	19.06	1	.000***
PS _B , one ρ	One ρ	4.24	3	.237	0.58	3	.900	3.40	3	.334	1.64	3	.650
PS _B , one ρ	PS _B	7.39	3	.060†	3.13	3	.372	6.63	3	.085†	6.53	3	.088†
PS	Equal μ , $\sigma = 1$	6.25	2	.044*	0.40	2	.818	1.89	2	.389	1.79	2	.409
PS	μ rectangle	4.82	2	.090†	0.56	2	.754	4.35	2	.114	3.18	2	.204
PS	PS _A	4.99	3	.173	0.57	3	.902	4.32	3	.229	2.02	3	.568
PS	PS _B	4.85	3	.183	1.76	3	.625	1.57	3	.665	2.63	3	.452

Note — H_0 : Restricted model is adequate. †Indicates $p < .01$, * $p < .05$, *** $p < .0001$.

Figures 7. Panels A–B: Equal likelihood contours of the best-fitting general recognition theory model for Observers 1 (A) and 2 (B). Panels C–D: Equal likelihood contours of the best-fitting general recognition theory model for Observers 3 (C) and 4 (D).

tification data, however, revealed the presence of a within-stimulus perceptual dependence. From an examination of the best-fitting model, this dependence was a negative correlation, which was especially large for the features of eye separation and nose length, the pair of features closest in spatial proximity.

The failure of existence separability bears much resemblance to the phenomenon in face recognition re-

ferred to as the *face superiority effect* (Homa, Haver, & Schwartz, 1976; Tanaka & Farah, 1993). This is the finding that observers are better able to perceive attributes or recognize individual features of previously seen features when the features are in the context of a whole face rather than in isolation. If the RT distributions for the whole-face trials (i.e., the combined RTs) in Experiment 1 had always been greater (i.e., faster) than those for the single-

Table 11
Parameter Estimates for the Best-Fitting Model
for Each Observer

Stimulus (Parameter)	Observer 1	Observer 2	Observer 3	Observer 4
A_1B_1				
ρ	0	-.392	-.250	-.125
A_1B_2				
μ_X	0.15	0	0	0
μ_Y	0.571	0.605	0.711	0.872
ρ	.275	-.392	-.250	-.125
A_2B_1				
μ_X	0.205	1.63	1.61	2.20
μ_Y	0.073	0	0	0
ρ	.023	-.392	-.250	-.125
A_2B_2				
μ_X	0.355	1.63	1.61	2.20
μ_Y	0.645	0.605	0.711	0.872
ρ	.222	-.392	-.250	-.125
X criterion	.05	.70	.84	1.08
Y criterion	.20	.08	.42	.50

feature case, this effect would have been essentially replicated. However, an examination of Figures 5A–5D shows that there is little systematicity in the distribution orderings even within an observer. In addition, the criterion shifts found in the detection analyses were not consistent across features within or across observers either. There is, at present, no explanation for this lack of systematicity.

Much of the literature on face perception argues that faces are encoded as configurations in which the relations between features is the primary currency used in the perceptual representation. This view was often inferred from the disproportionately deleterious effects of inversion on recognition memory for faces, relative to other stimuli (Farah et al., 1995; Valentine, 1988; Yin, 1969) and the relative recognizability of parts of a face in whole faces (Tanaka & Farah, 1993). These kinds of effects have been shown to hold for other stimulus types when expertise of recognition for them has been developed (Diamond & Carey, 1986). Consequently, that PS and DS was obtained here appears to be somewhat surprising. However, the task demands of the speeded classification task, which include the instruction to selectively attend to one component, are, in a sense, pleas for DS (see also Farah et al., 1995, for a similar result).

The finding of a within-stimulus perceptual dependence in the present study is quite consistent with the notion of configural or relational encoding of an individual face.¹⁹ Ashby (1989) and, recently, Kadlec (1999a) have suggested that, when a correlation exists, the titled axes of the contour can be interpreted as an “emergent” or “configural” feature, as in principal component analyses. In the case of eye separation and nose length, this emergent feature could be the area of the triangle determined by the feature placements and the conservation of this area across perceptual samples of an individual face. The fact that features that are nearer to each other generated a larger correlation seems to suggest that spatial proximity mediates the configural encoding. That is, the face may not

be encoded as a pure configuration or unitary whole that might predict relationships across facial features to be of the same strength regardless of location. Rather, the encoding is of the separate features (or areas) of the face, with spatial information registered.

The level of interaction, in terms of specificity, is limited to the processing of an individual face, rather than the categorization across faces. That is, if the results found here are correct, one’s ability to evaluate, say, the size of human noses across faces is not affected by the physical characteristics of the other features. However, the perception of an *individual* nose is contingent upon the perception of the displacement of the eyes (similarly for the mouth, but to a lesser degree). A process model that is consistent with this interpretation would assume that separate channels for facial features (or areas as in receptive fields) exist but that there is significant crosstalk between them (Ashby, 1989, 2000). The degree of crosstalk between the channels is a function of the spatial proximity of the features in the stimulus. The validity of a model could be assessed by methods of identifying architecture (Schweickert & Townsend, 1989; Townsend & Schweickert, 1989; Townsend & Thomas, 1994) with some recent results regarding the efficiency of processing in different architectures (Townsend & Nozawa, 1995). The results reported in this experiment place constraints on how the information in these architectures would have to be represented.

REFERENCES

- AKAIKE, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723.
- ASHBY, F. G. (1989). Stochastic general recognition theory. In D. Vickers & P. L. Smith (Eds.), *Human information processing: Measures, mechanisms, and models* (pp. 435–457). Amsterdam: Elsevier.
- ASHBY, F. G. (1992). Multivariate probability distributions. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 1–34). Hillsdale, NJ: Erlbaum.
- ASHBY, F. G. (2000). A stochastic version of general recognition theory. *Journal of Mathematical Psychology*, **44**, 310–329.
- ASHBY, F. G., BOYNTON, G., & LEE, W. W. (1994). Categorization response time with multidimensional stimuli. *Perception & Psychophysics*, **55**, 11–27.
- ASHBY, F. G., & GOTT, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 33–53.
- ASHBY, F. G., & LEE, W. W. (1991). Predicting similarity and categorization from identification. *Journal of Experimental Psychology: General*, **120**, 150–172.
- ASHBY, F. G., & MADDOX, W. T. (1991). A response time theory of perceptual independence. In J. P. Doignon & J. C. Falmagne (Eds.), *Mathematical psychology: Current developments*. New York: Springer-Verlag.
- ASHBY, F. G., & MADDOX, W. T. (1994). A response time theory of separability and integrality in speeded classification. *Journal of Mathematical Psychology*, **38**, 423–466.
- ASHBY, F. G., & MADDOX, W. T., & LEE, W. W. (1994). On the dangers of averaging across subjects when using multidimensional scaling and the similarity choice model. *Psychological Science*, **5**, 144–151.
- ASHBY, F. G., & PERRIN, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review*, **95**, 124–150.
- ASHBY, F. G., & TOWNSEND, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, **93**, 154–179.
- ATTNEAVE, F. (1950). Dimensions of similarity. *American Journal of Psychology*, **63**, 516–556.

- BRADSHAW, J. L., & WALLACE, G. (1971). Models for the processing and identification of faces. *Perception & Psychophysics*, **9**, 443-448.
- BRUCE, V. (1988). *Recognizing faces*. Hillsdale, NJ: Erlbaum.
- COLONIUS, H., & VÖRBERG, D. (1994). Distribution inequalities for parallel models with unlimited capacity. *Journal of Mathematical Psychology*, **38**, 35-58.
- DIAMOND, R., & CAREY, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, **115**, 107-117.
- FARAH, M. J., TANAKA, J. W., & DRAIN, H. M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 628-634.
- FRASER, I., & PARKER, D. (1986). Reaction time measures of feature saliency in a perceptual integration task. In H. D. Ellis, M. A. Jeeves, F. Newcombe, & A. Young (Eds.), *Aspects of face processing* (pp. 112-128). Dordrecht: Martinus Nijhoff.
- GARNER, W. R. (1974). *The processing of information and structure*. New York: Wiley.
- GARNER, W. R. (1976). Interaction of stimulus dimensions in concept and choice processes. *Cognitive Psychology*, **8**, 98-123.
- GARNER, W. R., & FELDOLDY, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, **1**, 225-241.
- GOLDSTONE, R. L. (1995). Effects of categorization on color perception. *Psychological Science*, **6**, 298-304.
- GOUREVITCH, V., & GALANTER, E. (1967). A significance test for one-parameter isosensitivity curves. *Psychometrika*, **32**, 25-33.
- HAIG, N. D. (1984). The effect of feature displacement on face recognition. *Perception*, **13**, 104-109.
- HOMA, D., HAVER, B., & SCHWARTZ, T. (1976). Perceptibility of schematic face stimuli: Evidence for a perceptual Gestalt. *Memory & Cognition*, **4**, 176-185.
- KADLEC, H. (1993, August). *Type I error and power of a hierarchical general recognition model: Testing and signal detection analyses of dimensional interactions*. Paper presented at the 26th Annual Mathematical Psychology Meeting, Norman, OK.
- KADLEC, H. (1995). Multidimensional signal detection analyses (MSDA) for testing separability and independence: A Pascal program. *Behavior Research Methods, Instruments, & Computers*, **27**, 442-458.
- KADLEC, H. (1999a, August). *Emergent properties as perceptual interactions in the general recognition theory*. Paper presented at the 32nd Annual Mathematical Psychology Meeting, Santa Cruz, CA.
- KADLEC, H. (1999b). Statistical properties of d' and β estimates of signal detection theory. *Psychological Methods*, **4**, 22-43.
- KADLEC, H., & HICKS, C. (1998). Invariance of perceptual spaces and perceptual separability of stimulus dimensions. *Journal of Experimental Psychology*, **24**, 80-104.
- KADLEC, H., & TOWNSEND, J. T. (1992a). Implications of marginal and conditional detection parameters for the separabilities and independence of perceptual dimensions. *Journal of Mathematical Psychology*, **36**, 325-374.
- KADLEC, H., & TOWNSEND, J. T. (1992b). Signal detection analyses of dimensional interactions. In F. G. Ashby (Ed.), *Multidimensional probabilistic models of perception and cognition* (pp. 181-227). Hillsdale, NJ: Erlbaum.
- LEHMAN, E. L. (1983). *Testing statistical hypotheses*. New York: Wiley.
- MACMILLAN, N. A., & CREELMAN, C. D. (1990). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.
- MACMILLAN, N. A., & ORNSTEIN, A. S. (1998). The mean-integral representation of rectangles. *Perception & Psychophysics*, **60**, 250-262.
- MADDOX, W. T. (1992). Perceptual and decisional separability. In F. G. Ashby (Ed.), *Multidimensional probabilistic models of perception and cognition* (pp. 147-180). Hillsdale, NJ: Erlbaum.
- MADDOX, W. T., & ASHBY, F. G. (1996). Perceptual separability, decisional separability, and the identification-speeded classification relationship. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 795-817.
- MATHEWS, M. L. (1978). Discrimination of Identikit constructions of faces: Evidence for a dual processing strategy. *Perception & Psychophysics*, **23**, 153-161.
- MILLER, J. (1996). The sampling distribution of d' . *Perception & Psychophysics*, **58**, 65-72.
- MURDOCK, B. B., JR. (1985). An analysis of the strength-latency relationship. *Memory & Cognition*, **13**, 511-521.
- NOSOFSKY, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, **115**, 39-57.
- NOSOFSKY, R. M., & SMITH, J. E. K. (1992). Similarity, identification, and categorization: Comment on Ashby and Lee (1991). *Journal of Experimental Psychology: General*, **121**, 237-245.
- OLZAK, L. A. (1986). Widely separated spatial frequencies: Mechanism interactions. *Vision Research*, **26**, 1143-1153.
- OTT, R. L. (1993). *An introduction to statistical methods and data analysis*. Belmont, CA: Duxbury.
- PALMER, S. E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology*, **9**, 441-474.
- PARKE, F. I. (1982). Parameterized models for facial animation. *IEEE: Computer Graphics & Applications*, **2**, 76-82.
- SCHWEICKERT, R., & TOWNSEND, J. T. (1989). A trichotomy: Interactions of factor prolonging sequential and concurrent mental processes in stochastic discrete mental (PERT) networks. *Journal of Mathematical Psychology*, **33**, 328-347.
- SEIGEL, S., & CASTELLAN, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.
- SERGEANT, J. (1984). An investigation into component and configural processes underlying face recognition. *British Journal of Psychology*, **75**, 221-242.
- SHEPARD, R. N. (1964). Attention and the metric structure of stimulus space. *Journal of Mathematical Psychology*, **1**, 54-87.
- TAKANE, Y., & SHIBAYAMA T. (1992). Structures in stimulus identification data. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 335-362). Hillsdale, NJ: Erlbaum.
- TANAKA, J. W., & FARAH, M. J. (1991). Second-order relational properties and the inversion effect: Testing a theory of face perception. *Perception & Psychophysics*, **50**, 367-372.
- TANAKA, J. W., & FARAH, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology*, **46A**, 225-245.
- THOMAS, R. D. (1995a). Gaussian general recognition theory and perceptual independence. *Psychological Review*, **102**, 192-200.
- THOMAS, R. D. (1995b). *A mathematical model of perceptual interactions in a same-different task*. Unpublished doctoral dissertation, Indiana University.
- THOMAS, R. D. (1996). Separability and independence of dimensions within the same-different judgment task. *Journal of Mathematical Psychology*, **40**, 318-341.
- THOMAS, R. D. (1999). Assessing sensitivity in multidimensional space: Some problems and a definition of a general d' . *Psychonomic Bulletin & Review*, **6**, 224-238.
- THOMAS, R. D. (2001). Characterizing perceptual interactions in face identification using multidimensional signal detection theory. In M. J. Wenger & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition* (pp. 193-227). Hillsdale, NJ: Erlbaum.
- TOWNSEND, J. T. (1974). Issues and models concerning the processing of a finite number of inputs. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 133-186). New York: Wiley.
- TOWNSEND, J. T., & ASHBY, F. G. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge: Cambridge University Press.
- TOWNSEND, J. T., & NOZAWA, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology*, **25**, 321-359.
- TOWNSEND, J. T., & SCHWEICKERT, R. (1989). Toward the trichotomy method of reaction times: Laying the foundation of stochastic mental networks. *Journal of Mathematical Psychology*, **33**, 309-327.
- TOWNSEND, J. T., & THOMAS, R. D. (1993). On the need for a general theory of similarity. In S. C. Masin (Ed.), *Foundations of perceptual theory* (pp. 297-368). Amsterdam: North-Holland.

- TOWNSEND, J. T., & THOMAS, R. D. (1994). Stochastic dependencies in parallel and serial models: Effects on systems factorial interactions. *Journal of Mathematical Psychology*, **38**, 1-34.
- VALENTINE, T. (1988). Upside-down faces: A review of the effects of inversion upon face recognition. *British Journal of Psychology*, **79**, 471-491.
- WICKENS, T. D. (1982). *Models for behavior: Stochastic processes in psychology*. San Francisco, CA: Freeman.
- WICKENS, T. D. (1992). Maximum-likelihood estimation of a multivariate Gaussian rating model with excluded data. *Journal of Mathematical Psychology*, **36**, 213-234.
- YIN, R. K. (1969). Looking at upside down faces. *Journal of Experimental Psychology*, **81**, 141-145.

NOTES

1. Decisionally separable decision criteria are not all that is possible. One class of decision criteria that is important is the class of optimal criteria, often defined as those that maximize accuracy in a typical experiment (i.e., in the absence of payoffs and with equal prior probabilities). In an identification task, given a perceptual effect (x, y) , an observer responding optimally would select the stimulus label associated with the largest probability density value. The optimal decision strategy often does not satisfy DS (Ashby & Gott, 1988).

2. Other conditions besides the one described here are usually included in a typical Garner classification task (Garner, 1974). However, in order to be meaningfully applied to tests of separability, these conditions require auxiliary assumptions, such as unlimited capacity and no context effects (Ashby & Maddox, 1994), and, hence, are not discussed here.

3. This is essentially an extension of Theorem 5 of Ashby and Townsend (1986) and Theorem 2 of Ashby and Maddox (1994). The proof of Equation 3a is straightforward, given the definition of existence separability and noting that

$$P_{FA}(R_{a_2} | A_i B_j) = \int_{-\infty}^{\infty} \int_{x_c}^{\infty} f_{A_i B_j}(x, y) dx dy = \int_{x_c}^{\infty} g_{A_i B_j}(x) dx \\ = \int_{x_c}^{\infty} g_{A_i}(x) dx = P_{FA}(R_{a_2} | A_i).$$

For Equation 3b, the proof is almost exactly the same as in Ashby and Maddox's Theorem 2, with the additional integral

$$\int_{x_c + h^{-1}(t)}^{\infty} g_{A_i}(x) dx = P_{FA}(T_2 \leq t | A_i),$$

where T_2 is the time to respond A was at Level 2, inserted in the next to last line (Ashby & Maddox, 1994, bottom of p. 457), x_c is the common decision bound, and $h^{-1}(t)$ entails the RT-distance hypothesis.

4. Even though the theory accounts for errors, it is certain that guessing has also occurred. Assuming that guessing occurs on trials for which the observer has no stimulus information, then for any stimulus, there is equal probability of emitting each response, and the RTs are simple RTs. Consequently, guessing trials would yield results consistent with separability. Because fewer errors overall occurred than correct responses, a disproportionate amount of the former would be due to guessing. Hence, the RT data from errors may yield an unacceptably high probability of accepting the null hypothesis of separability.

5. These predictions were assessed by a Kolmogorov-Smirnov (K-S) test for the equality of distributions (Seigel & Castellan, 1988). One determines the time, t , for which the two cumulative RT distributions are maximally different. The difference in cumulative proportion, D , is converted to a test statistic,

$$X^2 = 4D^2 \frac{n_1 n_2}{n_1 + n_2},$$

where n_1 and n_2 are the number of observations in each distribution, respectively. This X^2 follows a chi-square distribution with 2 degrees of freedom for large sample sizes.

6. This combining procedure was done because there does not appear to be available a generalized K-S test for more than two groups.

7. In tests of equality of (binomial) proportions, the test statistic is defined as

$$z = \frac{(p_1 - p_2)}{\sqrt{p(1-p) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}},$$

where $p = (r_1 + r_2)/(n_1 + n_2)$, n_i is the number of trials used to estimate p_i , and r_i is the number of "successes" within those trials used to compute p_i . Under the null hypothesis, this z follows a standard normal distribution (Ott, 1993). Table 3 summarizes the results of the MRI tests. Note that since the probability of an a_1 response given a stimulus is $1 -$ the probability of an a_2 response, it is sufficient to test Equations 2B and 3A on only one classification response. Again, for the assessment of existence separability, responses were combined in the two-feature case to obtain one proportion.

8. Although a number of "nonparametric" bias indices supposedly exist, deeper study into some reveals that either they really are parametric or they are not good indices of bias by other considerations of what a bias measure should be (Macmillan & Creelman, 1990).

9. The right equality is simply a restatement of Proposition 1a of Kadlec and Townsend (1992a), noting that existence separability implies PS. The left-hand side follows, since existence separability is defined as $g_{A_i B_j}(x) = g_{A_i B_2}(x)$ for all x . In the Gaussian case this is equivalent to $\mu_{x_{A_i B_j}} = \mu_{x_{A_i B_1}} = \mu_{x_{A_i B_2}}$ and $\sigma_{x_{A_i B_j}} = \sigma_{x_{A_i B_1}} = \sigma_{x_{A_i B_2}}$, which immediately implies the result (regardless of whether one assumes a common variance across levels of A).

10. The proof of this proposition follows the same line of argument as in the proof of Proposition 2a in Kadlec and Townsend (1992a).

11. Marginal p values are difficult to interpret, although the high number of observations suggest that power, even in these cases, is likely to be high as well.

12. Pilot data from a simple RT task indicate that the first fingers of each hand have similar motor time distributions but that significant and unpredictable differences exist for other fingers. Because the identification task requires four unique responses and, hence, fingers, it was believed an analysis of RTs would be too contaminated by these motor differences.

13. One uses a standard two-by-two (two components and two levels, each yielding four responses for each stimulus) chi-square contingency table test to evaluate SI. To illustrate this, the sum of the probabilities in the first bracket on the right side of Equation 4 is simply the aggregate proportion of a_1 responses for stimulus $A_i B_j$, and the second sum is simply the proportion of b_1 responses. For Observer 4, given stimulus $A_1 B_1$ (close eyes, thin mouth), the observed proportion of *close eyes* responses is the sum of the first and second columns in the first row of this observer's confusion matrix (159 + 74) divided by the total number of stimulus $A_1 B_1$ presentations (276), which is equal to .844. The proportion of *thin mouth* responses is the sum of the first and third columns divided by the row total, .699. SI holds if the product of these two proportions (.590) is (statistically) the same as the proportion of $a_1 b_1$ responses, .576 as determined by a chi-square test with 1 degree of freedom. For the 2×2 case, verification of Equation 4 implies that all other response combinations for that stimulus will be verified as well. This is because the row and column proportions sum to one.

14. The sum of the probabilities on the left-hand side of Equation 5A assessing MRI is the proportion of correct a_i responses when component B is at Level 1, and the sum on the right is the proportion of correct a_i responses when component B is at Level 2. For Observer 4, focusing on separability for the A component for stimuli with A at Level 1 and B at Level 1, the first sum is found from columns 1 and 2 of the first row in the confusion matrix relative to the row total (233/276 = .844). The second sum (i.e., B is at Level 2) is found from the first two columns of the second row of the confusion matrix (241/276 = .873). These two proportions are compared, using a z test for proportions, with the null hypothesis of equality supporting separability of the A component, as described in note 10.

15. Kadlec and Townsend (1992a) discuss assessing PI by using a microsignal detection analysis applied to identification data in which various conditional detection parameters can be estimated. This would be especially useful if independence was supported by the nonparametric SI test. Given that SI failed and that the MRI tests, together with the macrosignal detection analyses reported in this section, indicated that

DS holds, we can conclude that perceptual dependence exists. Hierarchical GRT modeling is better at diagnosing the sign and magnitude of this dependence (Kadlec, 1993; Kadlec & Townsend, 1992a, p. 355) than are the microsignal detection analyses. Hence, it was felt that these tests would be redundant and unnecessary.

16. The number of degrees of freedom in an $r \times r$ confusion matrix is $r \times (r - 1)$ because the row proportions sum to 1. Here, $r = 4$ stimuli, so the 4×4 matrix yields 12 degrees of freedom. A model with as many or more parameters than degrees of freedom in the data will fit the data perfectly.

17. This restriction on the variance assumes that the (marginal) variance on a given dimension is the same across levels of that dimension [i.e., $\sigma_X^2(A_1B_j) = \sigma_X^2(A_2B_j)$]. If only two levels of a dimension are sampled (as in the present case), discriminability differences owing to unequal variances are not identifiable from those caused by changes in the means.

18. Despite all the evidence marshaled in support of DS, one reviewer still questioned its use in the model-fitting analyses. It is true that the joint occurrence of a specific form of perceptual integrality (mean-shift; Macmillan & Ornstein, 1998) and nonorthogonal decision bounds,

whose angle matches exactly the angle formed by the parallelogram of means, could produce the entire constellation of results. The unlikelihood of obtaining exactly this configuration seems too great. Another difficulty with allowing for nondecisionally separable bounds in modeling of the identification data has to do with choosing what form of nonseparability to investigate. The only a priori justifiable choice is that of optimal responding. Consequently, all the models of Figure 6 were refit, assuming an optimal decision strategy (i.e., one that maximizes accuracy). Because, in all cases, the best likelihoods obtained were so inferior to those obtained assuming DS, even accounting for the differing number of parameters, this decision strategy could be discounted.

19. It should be noted that the across-stimulus integrality observed in the MDS configuration obtained by Sergent (1984) may be related to the within-stimulus dependencies revealed here. Ashby and Perrin (1988) established the mathematical relationship that exists between Euclidean MDS and the GRT. They found that MDS is equivalent to a GRT model that assumes within-stimulus independence. The only way for MDS to accommodate any failures of PI is to displace the means and/or variances within a dimension across stimuli—that is, to exhibit a perceptual integrality.

APPENDIX A

Consider the data from Observer 2 as an example. Let R_{a_i} denote the event that the observer has classified component A as an a_i . Assuming PS and $\sigma_{X_{A_1B_1}}^2 = \sigma_{X_{A_2B_1}}^2$ to compute the sensitivity for component A when B is at Level 1, d'_{AB_1} , a *hit* is defined as responding R_{a_2} when component A is at Level 2, and a *false alarm* occurs when R_{a_2} is given although component A is at Level 1. Now,

$$\begin{aligned} d'_{AB_1} &= \Phi^{-1}(\text{hit rate}) - \Phi^{-1}(\text{false alarm rate}) \\ &= \Phi^{-1}[\hat{P}(R_{a_2}|A_2B_1)] - \Phi^{-1}[\hat{P}(R_{a_2}|A_1B_1)], \end{aligned}$$

where $\Phi(z) = P(Z \leq z)$ is the distribution function of the standard normal variate. From the data for Observer 2, then,

$$\hat{P}(R_{a_2}|A_2B_1) = \frac{148}{240} = 0.61667,$$

and

$$\hat{P}(R_{a_2}|A_1B_1) = \frac{47}{240} = 0.1958.$$

Thus, $d'_{AB_1} = 1.157$. The decision criterion c_{AB_1} is computed using

$$\begin{aligned} c_{AB_1} &= -0.5[\Phi^{-1}(\text{hit rate}) + \Phi^{-1}(\text{false alarm rate})] \\ &= -0.5\{\Phi^{-1}[\hat{P}(R_{a_2}|A_2B_1)] + \Phi^{-1}[\hat{P}(R_{a_2}|A_1B_1)]\} \\ &= .278. \end{aligned}$$

From Gourevitch and Galanter (1967), the variance of d' is approximated by

$$\begin{aligned} \text{Var}(d'_{AB_1}) &= \frac{(1 - \text{false alarm rate})(\text{false alarm rate})}{N_{A_1B_1}[\phi(1 - \text{false alarm rate})]^2} + \frac{(1 - \text{hit rate})(\text{hit rate})}{N_{A_2B_1}[\phi(1 - \text{hit rate})]^2} \\ &= \frac{[1 - \hat{P}(R_{a_2}|A_1B_1)][\hat{P}(R_{a_2}|A_1B_1)]}{N_{A_1B_1}[\phi(1 - \hat{P}(R_{a_2}|A_1B_1))]^2} + \frac{[1 - \hat{P}(R_{a_2}|A_2B_1)][\hat{P}(R_{a_2}|A_2B_1)]}{N_{A_2B_1}[\phi(1 - \hat{P}(R_{a_2}|A_2B_1))]^2}, \end{aligned}$$

where $N_{A_iB_1}$ is the number of trials in which A_iB_1 was presented and

$$\phi(z) = \frac{d}{dz} \Phi(z),$$

the value of the density function of the standard normal variate at z . Miller (1996) provided an exact formula for the variance that requires significant computational efforts to compute. He found, however, that for $N >$

APPENDIX A (Continued)

100, the Gourevitch and Galanter approximation was quite accurate. Because all N s are nearly 240, the computationally simpler formation is used here. For the data obtained here,

$$\text{Var}(d'_{AB_1}) = \frac{(0.8042)(0.1958)}{(240)(.28873)^2} + \frac{(0.3833)(.6167)}{(240)(.37068)^2} = .01504.$$

By a similar computation, $d'_{AB_2} = 1.077$, with a variance .0153. A test statistic can be formed to test the equality of d'_{AB_1} and d'_{AB_2} ,

$$G = \frac{|d'_{AB_1} - d'_{AB_2}|}{\sqrt{\text{Var}(d'_{AB_1}) + \text{Var}(d'_{AB_2})}},$$

which has a standard normal distribution. Here, $G = .4613$. Hence, the p value for a two-tailed test is given by $p = 2 - 2\Phi(.4613) = .646$. To test for equality of the decision criteria, the variance of c'_{AB_j} is related to the variance of d'_{AB_j} by $\text{Var}(c'_{AB_j}) = .25 \text{Var}(d'_{AB_j})$ (Macmillan & Creelman, 1990), and this can be used as the variances in the formula for G above.

To assess existence separability, trials from two whole-face stimulus presentations in which the relevant dimension is held constant are combined to compute a component d' to compare with the single-featured d' . Consider Observer 2's responses to component A (eye separation). Denote by d'_{A*} the standardized distance between $\mu_{x_{A_1*}}$ and $\mu_{x_{A_2*}}$, where $*$ indicates that we have collapsed across levels of B (size). In this case, d'_{A*} can be estimated by

$$\begin{aligned} d'_{A*} &= \Phi^{-1}(1 - \text{false alarm rate}) - \Phi^{-1}(1 - \text{hit rate}) \\ &= \Phi^{-1}[\hat{P}(R_{a_2} | \{A_2B_1 \text{ or } A_2B_2\})] - \Phi^{-1}[\hat{P}(R_{a_2} | \{A_1B_1 \text{ or } A_1B_2\})], \end{aligned}$$

as before, so that $d'_{A*} = 1.11007$.

APPENDIX B

One measure of a model's fit to identification data is the likelihood function. For an r -row $\times$ r -column confusion matrix, the likelihood, L , of the data, given the model, is

$$L = \prod_{i=1}^r \left[\frac{n_i!}{\prod_{j=1}^r f_{ij}!} \prod_{j=1}^r P(R_j | S_i)^{f_{ij}} \right], \quad (\text{B1})$$

where n_i is the total number of trials stimulus i was presented, f_{ij} is the observed frequency with which Response R_j was given on trials when Stimulus S_i was presented (the entry for cell i, j in the obtained confusion matrix), and $P(R_j | S_i)$ is the model's predicted response probability for Response j and Stimulus i . Often, for computational convenience, the logarithm of Equation B1 can be used to find the best fitting parameter values,

$$\ln L = \sum_{i=1}^r \ln n_i! + \sum_{i=1}^r \sum_{j=1}^r f_{ij} \ln P(R_j | S_i) - \sum_{i=1}^r \sum_{j=1}^r \ln f_{ij}!. \quad (\text{B2})$$

To fit a specific model to identification data, one uses an iterative routine that searches through the possible parameters values so that Equation B2 is maximized (Wickens, 1982). So long as the number of parameters is less than the degrees of freedom in the confusion matrix, any model can be fit to identification confusions in this manner. One important aspect of Equation B2 needs to be mentioned. If n_i or f_{ij} is very large (e.g., greater than 170) for any i or j , the factorials cannot be computed on many computers. An alternative fit function that can be used that yields equivalent parameter estimates is the middle term of Equation B2,

$$\sum_{i=1}^r \sum_{j=1}^r f_{ij} \ln P(R_j | S_i).$$

This is the only part of the log-likelihood function that depends on the parameters of the model (Takane & Shibayama, 1992). When this term is substituted into the chi-square and AIC formulas (see below), there is no effect on the outcome of the model selection process, because the other terms in the equation are constant across all of the models. For details on the numerical aspects of how this fit statistic was maximized, see Ashby (1992) and Ashby and Lee (1991, Appendix). An alternative algorithm can be found in Wickens (1992).

Another measure of fit is based on the sum of squared deviations (SSE) between the observed frequencies and the predicted frequencies summed over all the cells in the confusion matrix,

$$SSE = \sum_{i=1}^r \sum_{j=1}^r [f_{ij} - \hat{f}_{ij}]^2,$$

APPENDIX B (Continued)

where f_{ij} is the observed frequency with which Response R_j was given on trials when Stimulus S_i was presented (the entry for cell i, j in the obtained confusion matrix) and f_{ij} is the model's predicted frequency for that cell. The test statistic for nested models based on SSE is an F_{obs} ,

$$F_{\text{obs}} = \frac{(SSE_{M_1} - SSE_{M_2}) / (df_{M_1} - df_{M_2})}{SSE_{M_2} / df_{M_2}},$$

where df_{M_i} is the degrees of freedom associated with model M_i . Under the null hypothesis that the restricted model, M_1 is correct, this statistic follows an F with numerator degrees of freedom $(df_{M_1} - df_{M_2})$ and denominator degrees of freedom df_{M_2} .

One problem with SSE is that a deviation between predicted and observed frequencies of a certain size contributes equally to the measure of fit, regardless of the absolute frequency in the cell. For instance, a deviation of 10 observations has the same effect on SSE whether the cell has 5 or 500 observations. However, a discrepancy between 5 and 15 seems more important than a discrepancy between 500 and 510. Likelihood does not suffer from this problem and is, by some, preferred (Nosofsky & Smith, 1992).

More general models will always fit better in the absolute sense, but the maximum-likelihood procedure allows a statistical hypothesis test, a *likelihood ratio test*, to determine whether the improvement from the extra complexity of the more general model is capturing real aspects in the data or whether the simpler, more parsimonious model provides an adequate description. Suppose two models, M_1 and M_2 are nested, in that M_1 is a special case of M_2 . Let L_i be the likelihood associated with M_i (Equation B2). The null hypothesis for this test is that the special case model is correct. The test statistic X^2 is defined from the likelihood ratio L_1 / L_2 to be

$$\begin{aligned} X^2 &= -2 \ln \left(\frac{L_1}{L_2} \right) \\ &= 2[\ln L_2 - \ln L_1]. \end{aligned}$$

Under the null hypothesis, this statistic asymptotically follows a χ^2 distribution with degrees of freedom equal to the difference in the number of free parameters between M_1 and M_2 . In the present context, if M_1 assumes PI or PS but M_2 does not and the likelihood ratio test rejects the null hypothesis, we can conclude that PI or PS does not hold for those data.

The likelihood ratio test is useful when models are nested with respect to one another, but as can be seen in Figure 6, many models are not nested (e.g., PS_B , PI is not nested in PS, one ρ). In this case, the choice of which model is better depends on whether they have the same number of parameters or not. If two models have the same number of parameters (e.g., PS_A , PI and PS_B , PI), the one with the largest likelihood is preferred. If nonnested models differ in number of parameters, a measure of goodness of fit that takes this difference into account is AIC. Each model has an associated AIC that is a function of its likelihood, L , and number of parameters, v , $AIC = -2 \ln L + 2v$. One computes an AIC for each model in question and chooses the one with the smallest AIC as the "best" one.

(Manuscript received October 20, 1997;
revision accepted for publication June 28, 2000.)