JOURNAL OF MATHEMATICAL PSYCHOLOGY 25, 119–162 (1982)

An Experimental and Theoretical Investigation of the Constant-Ratio Rule and Other Models of Visual Letter Confusion

JAMES T. TOWNSEND AND DOUGLAS E. LANDON

Purdue University

The constant-ratio rule (CRR) and four interpretations of R. D. Luce's (In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1). New York: Wiley, 1963) similarity choice model (SCM) were tested using an alphabetic confusion paradigm. Four stimulus conditions were employed that varied in set size (three, four or five stimulus elements) and set constituency (block letters: *A, E, X*; *F, H, X*; *A, E, F, H*; *A, E, F, H, X*), and were presented to each subject in independent blocks. The four interpretations of the SCM were generated by constraining one, both, or neither of its similarity and bias parameter sets to be invariant in across-stimulus set model predictions. The strictest interpretation of the SCM (both the similarity and bias parameters constrained), shown to be a special case of the CRR, and the CRR produced nearly equivalent across-set predictions that provided a reasonable first approximation to the data. However, they proved inferior to the least strict SCM (neither the similarity nor bias parameters were constrained; the common interpretation of the SCM in visual confusion). Additionally, the least strict SCM was compared to J. T. Townsend's (*Perception and Psychophysics*, 1971, **9**, 40–50, 449–454) overlap model, the all-or-none model (J. T. Townsend, *Journal of Mathematical Psychology*, 1978, **18**, 25–38), and a modified version of L. H. Nakatani's (*Journal of Mathematical Psychology*, 1972, **9**, 104–127) confusion-choice model. Both the least strict SCM and confusion-choice models produced nearly equivalent within stimulus set predictions that were superior to the overlap and all-or-none within-set predictions. Measurement conditions related to model structure and equivalence relations among the models, many of them new, were examined and compared with the statistical fit results of the investigation.

How people go about the recognition of symbolic patterns, especially letters in their language, has long been of interest to psychologists, both in its own right and as a subprocess of reading. Although the recognition process is still far from well understood, it is perhaps fair to say that progress has been made, at least in so far as being able to predict letter accuracy and confusions in experiments employing degraded displays (e.g., Townsend, 1971a, 1971b), response time deadline procedures (Pachella, Smith, & Stanovich, 1978), and subjects who are not entirely versed in the alphabet (e.g., children, as in Gibson, Osser, Shiff, & Smith, Note 1) is concerned.

Portions of this study were presented at the 13th Annual Mathematical Psychology Meeting, Madison, Wis., 1980. Support for this research was provided by NSF Grant BNS 7920298 to the first author. Send requests for reprints to James T. Townsend, Department of Psychological Sciences, Purdue University, West Lafayette, Ind. 47907.

0022-2496/82/020119–44$02.00/0

A particularly useful device for the analysis of the recognition process has been the so-called confusion matrix, with one of the first applications to letter perception being that of Gibson *et al.* (Note 1). In a typical experiment, a subject is presented a letter in a very brief display (on the order of less than 100 msec in most cases) and is asked to identify the presented letter by verbal response or by pushing an appropriate button. The data acquired by such a procedure can be analyzed by arraying it as a confusion matrix. Usually, the rows of the matrix represent the stimulus letters and the columns the responses. The columns are ordered to correspond to the rows so that row i and column i designate the proper response-to-stimulus assignment (e.g., in a whole alphabet experiment, response A with stimulus A, etc.). Each cell in the matrix will then contain the proportion of times the subject responded with j given the presentation of stimulus i, with the negative diagonal cells containing the proportion correct for each stimulus.

The theoretical approaches to letter recognition that are oriented around mathematical modeling can be roughly broken down into those models based explicitly on the notion of feature processing (e.g., Geyer & DeWald, 1973; Townsend, Hu, & Ashby, 1981) and those which are somewhat of a more descriptive nature (e.g., Townsend, 1971a, 1971b, 1978). In fact, the former have been referred to as "substantive feature models" because of the relatively detailed delineation of the models in terms of feature mechanisms, while the others have been called "descriptive models" (Townsend & Landon, 1982; Townsend & Ashby, Note 2). Although the dynamic processing notions of the feature models are usually quite explicit, the descriptive models vary in the degree to which they have been interpreted in terms of processing concepts.

A third mathematical approach, as yet largely untried but not independent of the approaches just described, may be referred to as the "measurement approach." The measurement approach (in our sense) could be said to be interested in the types of constraints on confusion matrices that are implied by various models of confusion (e.g., Holman, 1979), or that are simply related to some characterization of confusion. In the former case, it is sometimes possible to find constraints on the patterns of confusion in the matrices that are non-parametrically predicted to hold by a model or class of models. Or, in the latter case, one might start with a set of constraints and define a class of models by those constraints.

One measurement constraint of considerable importance is the constant-ratio rule (Clarke, 1957; Luce, 1959; hereafter referred to as CRR[1]). It has been employed or tested in other contexts (e.g., auditory recognition; Clarke, 1957) but not, to our knowledge, in visual letter recognition (Anderson, Note 3, employed visually presented monosyllables). The CRR can be tested in its own right (several tests are cited below), but may also be associated with a particular interpretation of what we call the similarity choice model (SCM; see Luce, 1963). In addition to testing the CRR in letter recognition, one of the primary goals of the present investigation is to test several interpretations of the SCM (each based on varying degrees of association

[1] Appendix A contains a glossary of the abbreviations which are used in the paper to save space.

with the CRR measurement constraint) against each other, as well as testing the SCM against other descriptive models of confusion. The emphasis on the SCM derives from its consistently good performance in predicting the patterns of confusions in recognition relative to other models over a number of studies (several are cited below). In addition, where feasible, implication and equivalence relations are developed among models in order to enhance understanding of their logical structure and where it is possible and not possible to test them. Finally, certain other measurement constraints that have been discovered are examined in the data.

Before proceeding, it will be helpful to outline some notation that will be used throughout the paper. The lowercase letters $i, j, k, \ldots$ will denote elements from the sets $\mathcal{S} = \{1, 2, 3, \ldots, s\}$, $\mathcal{M} = \{1, 2, 3, \ldots, m\}$, $\mathcal{Q} = \{1, 2, 3, \ldots, q\}$, where $\mathcal{S}$, $\mathcal{M}$, and $\mathcal{Q}$ denote sets of stimuli comprising confusion matrices S, M, and Q, respectively. Specifically, M will represent a finite $m \times m$ confusion matrix containing the m stimuli of $\mathcal{M}$, where $\mathcal{M}$ contains all of the stimuli employed in the experiment. M will be considered the master matrix. S will represent a finite $s \times s$ confusion matrix such that $\mathcal{S} \subseteq \mathcal{M}$. That is, S will be a confusion matrix obtained with a subset of the stimuli used in the master matrix M and will be referred to as a subset matrix. Q will represent any general finite $q \times q$ confusion matrix. The notation M and S will be used when the distinction between a master matrix and a subset matrix is important, whereas Q will be used in the general case where the distinction is not important. The notation $P_S(j \mid i)$, $P_M(j \mid i)$, and $P_Q(j \mid i)$ will represent the probability that stimulus j is confused for stimulus i in a confusion matrix S, M, and Q, respectively. Other notation will be introduced as needed.

I. The Constant-Ratio Rule and Similarity Choice Model

Clarke (1957) introduced a method of predicting the patterns of confusions in a subset of stimuli from known confusions in a subset of stimuli from known confusions in a master set. He termed this method the CRR because it assumes that the ratio between any two entries in a row of the subset matrix is equal to the ratio between the corresponding entries in the master matrix.

DEFINITION 1. For a subset matrix S, the CRR hypothesis is defined as

$$P_S(j \mid i) = \frac{P_M(j \mid i)}{\sum_{k \in \mathcal{S}} P_M(k \mid i)}, \quad p(j \mid i) > 0, \qquad \text{for} \quad i, j \in \mathcal{S}. \tag{1}$$

Several tests of the CRR have been conducted, and most have concluded that the CRR does make satisfactory predictions with choice data (Anderson, Note 3; Clarke, 1957, 1959; Clarke & Anderson, 1957; Egan, 1957, Note 4; Hodge, 1967; Hodge & Pollack, 1962; Pollack & Decker, 1960). However, there is some question as to the validity of the results of these past studies due to what was at the time a lack of any statistical test of the deviations of the results from the predictions of the CRR.

Morgan (1974) pointed out this problem and proposed that a likelihood ratio test of equality between two transition matrices (Hilton, 1971; Wilks, 1938) be utilized to statistically test CRR predictions. Morgan (1974) applied his likelihood ratio test to the data reported by Clarke (1957) and Egan (1957) and found their data to depart significantly from the predictions of the CRR. Both Clarke (1957) and Egan (1957) presented their data pooled over subjects and reported that the CRR held with their data. Morgan (1974) suggested that the discrepancy between his conclusion and the conclusions of Clarke and Egan might have been due to the use of pooled data, and could possibly be cleared up through the use of individual subject data and analysis.

Luce (1959) has shown that the CRR can be derived from a general theory of choice behavior composed of the two-part choice axiom. Let $\mathscr{A}$ be a set of choice objects and $\mathscr{S}$ a subset of $\mathscr{A}$. Part (i) of the choice axiom demands that the probability that $i \in \mathscr{S} \subset \mathscr{A}$ is selected from the total set $\mathscr{A}$ is equal to the probability that the selected element comes from the subset $\mathscr{S}$ times the probability that i is chosen when the only alternatives available are contained in $\mathscr{S}$. That is, letting $P(i; \mathscr{A})$ equal the overall probability of selecting i from a set $\mathscr{A}$, then part (i) of the choice axiom states that

$$P(i; \mathscr{A}) = P(\mathscr{S}; \mathscr{A})\, P(i; \mathscr{S}). \tag{2}$$

Part (ii) of the choice axiom states that if an alternative i is always chosen over an alternative j in set $\mathscr{A}$, then j may be deleted from set $\mathscr{A}$ when considering choices from that set. If $P(j; i)$ represents the probability that j is selected from set $\{i, j\}$, then, for $i, j \in \mathscr{A}$ and $P(j; i) = 0$, part (ii) of the choice axiom states that

$$P(i; \mathscr{A}) = P(i; \mathscr{A} - \{i\}).$$

In the situations considered in this paper, imperfect discrimination will be assumed so that part (ii) will not be required. In confusion terms, $P(j \mid i) > 0$ for all $i, j \in \mathscr{A}$. Observe that our notation for choice in confusion experiments is more specialized than that used here for the choice axiom to take into account the stimulus presented as well as the choice set (stimulus alphabet) available.

Part (i) is equivalent to the CRR (see Luce, 1959, p. 9), and rests on a version of the assumption of independence from irrelevant alternatives, the idea being that alternatives that should be irrelevant to the choice situation are in fact irrelevant. Luce (1959, pp. 23–27) has shown further that his general choice theory can be expressed by the formula

$$P(i; \mathscr{A}) = \frac{V(i)}{\sum_{j \in \mathscr{A}} V(j)}, \tag{3}$$

where $V(i)$ represents the constant scale value of alternative i.

Luce (1963) has also proposed a variation of the choice theory (Eq. (3)) that can be used for non-binary psychophysical analyses (e.g., confusion matrices) in which he multiplicatively decomposes the V ratio scale into an η ratio scale and a β ratio scale.

This model has sometimes been referred to as the "biased choice model." We prefer the name "similarity choice model" (SCM, as noted) to differentiate it from certain other choice models of Luce. For instance, Luce (1959) presents a model for psychophysical tasks which includes separate terms for the stimulus and response bias, but the stimulus terms do not permit unique similarity between two stimuli.

DEFINITION 2. For any single finite $q \times q$ confusion matrix Q, the SCM predicts the entries in Q as

$$P_Q(j \mid i) = \frac{\eta_{ij}\beta_j}{\sum_{k \in \mathscr{Q}} \eta_{ik}\beta_k}, \qquad \text{for all } i, j \in \mathscr{Q}, \tag{4}$$

with the constraints that $0 \leqslant \eta_{ij} \leqslant 1$, $0 \leqslant \beta_j \leqslant 1$, $\eta_{ii} = 1$, and $\eta_{ij} = \eta_{ji}$.[2] The SCM requires that $[(q(q+1))/2] - 1$ parameters be estimated. The η scale is interpreted as a stimulus similarity measure (η_{ij} represents the similarity between stimulus i and stimulus j) and the β scale is interpreted as a response bias measure.

Equation (4) is fairly well known, and has been widely investigated. For example, Shipley and Luce (1964) had some success in fitting Eq. (4) to data from discriminations of different weights, while Townsend (1971a, 1971b) and Townsend and Ashby (Note 2) found that Eq. (4) performed better than the other tested models in fitting alphabetic confusion matrices. More recently, Lupker (1979) found acceptable fits of Eq. (4) to his eight confusion matrices.

The SCM (Eq. (4)) need not be associated with the CRR (Eq. (1)). That is, Eq. (4) may be employed to predict a confusion matrix without necessarily assuming the CRR, which is an across-matrix predictor. If, however, it is assumed that the η's and β's of Eq. (4) are invariant from the master to the subset matrices, then the CRR is automatically predicted. We term this the strong similarity choice model (SSCM), while the variant without the CRR will be termed the weak similarity choice model (WSCM).

DEFINITION 3. (a) Two confusion matrices S and M satisfy the SSCM if there exist parameters $0 \leqslant \eta_{ij} \leqslant 1$, $0 \leqslant \beta_j \leqslant 1$, $\eta_{ii} = 1$, $\eta_{ij} = \eta_{ji}$, such that

$$P_M(j \mid i) = \frac{\eta_{ij}\beta_j}{\sum_{k \in M} \eta_{ik}\beta_k}, \qquad \text{for} \quad i, j \in M \tag{5}$$

and

$$P_S(j \mid i) = \frac{\eta_{ij}\beta_j}{\sum_{k \in S} \eta_{ik}\beta_k}, \qquad \text{for} \quad i, j \in S.$$

[2] It is easy to show that the $\eta_{ii} = 1$ condition is equivalent to simply requiring that $\eta_{ii} = \eta_{jj}$. The constraint that $\beta_j \leqslant 1$ is not substantive; in fact, one may require that $\sum \beta_j = 1$ without loss of generality.

SSCM requires estimation of one set of parameters for both matrices. Typically, if $S \subseteq M$ this entails estimation of $[m(m+1)/2] - 1$ parameters.

(b) Any matrix S satisfies WSCM if there exist parameters η_{ij} and β_j constrained as in part (a) such that

$$P_S(j \mid i) = \frac{\eta_{ij}\beta_j}{\sum_{k \in S} \eta_{ik}\beta_k}, \qquad \text{for} \quad i, j \in S. \tag{6}$$

The WSCM requires estimation of the $[s(s+1)/2] - 1$ parameters for each such matrix.

Equation (6) describes Eq. (4) when Eq. (4) is separately fit to the available confusion matrices, even though one or more of the available confusion matrices are comprised of stimuli that represent a subset of a larger confusion matrix. Equation (6) is also the common interpretation of Eq. (4) that most researchers have used when employing the SCM in their studies, although the distinction between the weak and strong versions of Eq. (4) was not specified. However, some work has been done investigating the invariance of the η's and variations of the β's in the SCM when motivational aspects of the experiment were changed (Townsend & Ashby, Note 2).

It is interesting to note that the SSCM is a special case of the CRR. The latter may hold without the former holding, but not vice versa. This is demonstrated in the following theorem.

THEOREM 1. *The SSCM implies the CRR, but the CRR does not imply the SSCM.*

Proof. (a) SSCM $\Rightarrow$ CRR is obvious by the definition of the SSCM.

(b) CRR $\not\Rightarrow$ SSCM. First note that the predictions of the CRR are independent across rows in the sense that logically it can hold for some rows but not for others (see Eq. (1)). Let $V_M(i, j)$ be the scale value of response j given stimulus i in the master matrix M, where in general, $V_M(i, j) \neq V_M(j, i)$. Rewriting the scale values of Eq. (3) in terms of a certain row of Eq. (1), the confusion probabilities of a subset matrix S may be expressed as

$$P_S(j \mid i) = \frac{V_M(i, j)}{\sum_{k \in \mathscr{S}} V_M(i, k)},$$

where $0 < V_M(i, j) < +\infty$. From this representation, it is easy to ascertain that the CRR puts no constraints whatsoever on the master matrix. That is, the representation by Eq. (3) is always possible for a single matrix. This is of course not true for the SSCM in its form of WSCM for a single matrix. It will be shown later that the WSCM puts several constraints on a single matrix (cf. Definition 8 and Theorem 2). ∎

Equation (3) (and the CRR) acquires its strength (in the SSCM) from the invariance constraints in going from the master matrix to the submatrices. It is also apparent that the CRR does not imply the WSCM and conversely because the CRR assumes a row independence in the confusion matrix, which the WSCM does not, and the WSCM assumes that $\eta_{ij} = \eta_{ji}$, which the CRR does not.

The SSCM expressed by Eq. (5) is a very strong interpretation of Eq. (4) and the CRR in the sense that both the η and β parameter estimates in an experiment are constrained to be invariant between the master matrix and the predicted subset matrices. In other words, only one set of parameters may be estimated from a master confusion matrix containing all of the stimuli used in the experiment, and then the appropriate η and β parameter values necessary for a specific subset drawn from the master set are used to find the model predictions of the confusions that should be obtained when that subset of stimuli is independently presented to the subject.

Models that lie between the SSCM and the WSCM can be obtained by constraining either the η or the β parameter estimates rather than both, as in the SSCM. In the former case, the η parameter estimates obtained from a fit of Eq. (4) to the master matrix are constrained to be constant across all subsets drawn from the master set, while the β parameters are considered unconstrained and are reestimated for each subset matrix. This model will be referred to as the SSCM(β) (i.e., the unconstrained parameter appears in the parentheses). The β parameters can also be constrained, while the η parameters are reestimated for each subset matrix. This model will be called SSCM(η).

DEFINITION 4. (a) Two confusion M and S matrices satisfy the SSCM(β) if there exist a single set of similarity parameters η_{ij} which applies to both matrices and two sets of bias parameters β_{jM} and β_{jS} which apply to the separate matrices. If $S \subseteq M$ then denote the η_{ij} parameters as η_{ijM} so that

$$P_M(j \mid i) = \frac{\eta_{ijM}\beta_{jM}}{\sum_{k\in M} \eta_{ikM}\beta_{kM}}, \qquad \text{for} \quad i, j \in M,$$
$$P_S(j \mid i) = \frac{\eta_{ijM}\beta_{jS}}{\sum_{k\in S} \eta_{ikM}\beta_{kS}}, \qquad \text{for} \quad i, j \in S, \tag{7}$$

where η_{ij} and β_j obey the constraints of Definitions 1 and 2. It is then required that $m(m-1)/2$ similarity parameters be estimated that apply to M and (any) $S \subseteq M$ and an additional $(s-1)+(m-1)$ bias parameters be estimated for S and M resulting in a total of $[m(m+1)/2] - 2 + s$ parameter estimates.

(b) Two confusion matrices M and S satisfy the SSCM(η) if there exist a single set of bias parameters β_j which applies to both matrices and two sets of similarity parameters η_{ijM} and η_{ijS} which apply to the separate matrices. If $S \subseteq M$ then denote the β_j parameters as β_{jM} so that

$$P_M(j \mid i) = \frac{\eta_{ijM}\beta_{jM}}{\sum_{k\in M} \eta_{ikM}\beta_{kM}}, \qquad \text{for} \quad i, j \in M$$

and

$$P_S(j \mid i) = \frac{\eta_{ijS}\beta_{jM}}{\sum_{k\in S} \eta_{ikS}\beta_{kM}}, \qquad \text{for} \quad i, j \in S \tag{8}$$

with the usual constraints on η_{ij}, β_k. It is then required that $m-1$ bias parameters be estimated that apply to M and (any) $S \subseteq M$ and $s(s-1)/2 + m(m-1)/2$ similarity parameters be estimated for the separate matrices, resulting in a total of $[m(m+1)/2] + s(s-1)/2 - 1$ parameters.

Although the CRR has been found to perform fairly well, the previous tests of the CRR were not very stringent in that they lacked strict statistical tests of CRR predictions. Additionally, although Eq. (4) has been tested, the distinction between the weak and the strong versions was not made. The fact that the SSCM is predicted by the CRR, and that the SSCM, SSCM(η), and SSCM(β) are special cases of the WSCM, gives rise to the question of how these models will compare in an appropriate test.

II. Other Current Models of Confusion

The SSCM and the WSCM can be compared to other standard models of confusion. The other models to be examined in this study are the overlap model (OVLP; Townsend, 1971a), the all-or-none model (AON; Townsend, 1978), and a specialized version of Nakatani's (1972) confusion-choice model (NCC). The OVLP is a special case of the WSCM within the context of a single confusion matrix (Pachella *et al.*, 1978), although not across two or more confusion matrices produced by manipulation of stimulus or response variables. The AON is a useful, if simple, vehicle for providing a base comparison in that it assumes a complete lack of effect of inter-stimulus similarity (Townsend, 1971a, 1971b) as well as row by column multiplicativity in off-diagonal matrix cells (Townsend, 1978). The NCC is a general version of the OVLP that has the same number of parameters as the WSCM. Neither the OVLP, the AON, nor the NCC predicts the CRR. As mathematical and psychological treatments of all three models may be found elsewhere (e.g., Nakatani, 1972; Pachalla *et al.*, 1978; Townsend, 1971a, 1971b, 1978; Townsend & Ashby, Note 2), only a brief explanation of each model will be offered.

The simplest interpretation of the AON is as a strict template matching perceptual process. Upon presentation of the stimulus, the subject either recognizes the stimulus perfectly (which occurs with probability p_i), or with probability $1 - p_i$ the subject is thrown into a null information state in which he/she can guess the correct stimulus with probability h_i. The probability p_i is interpreted as a sensory parameter that indicates the degree to which stimulus i is perfectly recognized. The probability h_j is interpreted as a bias parameter for stimulus j indicative of subject preference for responding j. The AON is formally presented in Definition 5.

Definition 5. For any single finite $q \times q$ confusion matrix Q, the AON predicts the entries in Q as

$$\begin{aligned} P_Q(j \mid i) &= (1 - p_i)h_j, && \text{for} \quad i, j \in \mathcal{Q}, \quad i \neq j, \\ &= p_i + (1 - p_i)h_i, && \text{for} \quad i, j \in \mathcal{Q}, \quad i = j, \end{aligned} \tag{9}$$

with the constraints that $0 \leqslant p_i \leqslant 1$, $0 \leqslant h_j \leqslant 1$, and $\sum_j h_j = 1$. The AON requires that $2q - 1$ parameters be estimated.

The OVLP model assumes that either perfect information or two-way partial information is acquired by the subject at each stimulus presentation. With probability ξ_{ii} stimulus i is recognized perfectly and the correct response is given. With probability $1 - \xi_{ii}$ the subject enters some pairwise confusion state in which the subject is unsure as to which of the two letters in the pairwise state was actually presented. ξ_{ij} represents the probability that the pairwise confusion state is between stimulus i and stimulus j. When in the pairwise confusion state, the subject responds with j according to the ratio of the biases of the two stimuli involved in the pairwise confusion. The OVLP is formally presented in Definition 6.

DEFINITION 6. For any single finite $q \times q$ confusion matrix Q, the OVLP predicts the entries in Q as

$$\begin{aligned} P_Q(j \mid i) &= \xi_{ij} \times [g_j/(g_j + g_i)], && \text{for } i, j \in \mathscr{Q}, i \neq j, \\ &= \xi_{ii} + \left\{ \sum_{k=1, k \neq i}^{q} \xi_{ik} \times [g_i/(g_i + g_k)] \right\}, && \text{for } i, j \in \mathscr{Q}, i = j, \end{aligned} \tag{10}$$

with the constraints that $0 \leqslant \xi_{ij} \leqslant 1$, $\xi_{ij} = \xi_{ji}$, and $\sum_j \xi_{ij} = 1$. The OVLP requires that $[q(q+1)/2] - 1$ parameters be estimated. It may also be assumed that $0 \leqslant g_j \leqslant 1$ and $\sum_j g_j = 1$ without a loss of generality.

The probability ξ_{ij} is interpreted as a similarity parameter indicative of the degree to which stimulus i and stimulus j are likely to occur together in a pairwise confusion. The probability g_j is interpreted as a bias parameter that indicates the subject's preference for responding with j.

The OVLP was employed in the present study in order that simple pairwise interstimulus relationships may be contrasted with the pure guessing strategy of the AON and the stimulus similarity relationships of the WSCM and the SSCM. An additional comparison can be made between the OVLP and the WSCM by employing a perceptual model that allows for larger confusion states than just pairwise states (i.e., three-way and four-way confusions).[3] A specialized version of Nakatani's (1972) confusion-choice model (NCC, as noted) was developed for this purpose as it can predict large confusion states without an overabundance of parameters. It is particularly useful in this study as the specialized version has the same number of parameters as the WSCM and OVLP ((i.e., $[(q(q+1))/2] - 1$) and has not yet been tested against the AON, OVLP, and choice models.

[3] Pachella *et al.* (1978) have developed a model they refer to as the informed guessing model. Although it allows for a total confusion state along with pairwise confusion states, it does not allow for the possible three- and four-way confusions that could be found in the stimulus sets that are used in the present study.

The special case of the NCC derived for the present study is essentially a generalized (logically, not mathematically) version of the OVLP. Upon stimulus presentation, the subject enters an n-way confusion state with some probability, and then selects his/her response from that set. Let $\mathscr{C}$ be the class of all possible confusion states, where $\mathscr{C}_k \in \mathscr{C}$ represents one possible confusion state with n acceptable responses out of the possible q stimuli. That is, each $\mathscr{C}_k$ represents a set of possible alternatives in $\mathscr{Q}$. The probability that $\mathscr{C}_k$ is the confusion state that arises from presentation of stimulus i is denoted π_{ik}. Each probability π_{ik} is obtained by a multiplicative combination of acceptance probabilities, denoted ω_{ij}, where ω_{ij} represents the probability that j enters the confusion state (set) when stimulus i has been presented. Once the subject enters a particular n-way confusion state, $\mathscr{C}_k$, response j is selected from the state with probability λ_{kj}, which represents the ratio of the biases of the stimuli contained in the n-way confusion state. The NCC is formalized in Definition 7.

DEFINITION 7. (a) The acceptance matrix, A, is defined as a finite $q \times 2^q$ matrix of probabilities π_{ik}, where π_{ik} is defined as follows for a finite stimulus set $\mathscr{Q}$:

$$\begin{aligned} \pi_{ik} &= P(\mathscr{C}_k = \{i, j, ...\} \mid i) \\ &= \prod_{j \in \mathscr{C}_k} \omega_{ij} \prod_{l \notin \mathscr{C}_k} (1 - \omega_{il}), \qquad \mathscr{C}_k \subseteq \mathscr{Q}, \end{aligned} \tag{11}$$

with the constraints that $0 \leqslant \omega_{ij} = \omega_{ji} \leqslant 1$ and $\omega_{ii} = \omega_{jj} = 1$. The constraint that $\omega_{ii} = 1$ will set the probability $\pi_{ik} = 0$ whenever the presented stimulus i is not included in the confusion set $\mathscr{C}_k$.

(b) The decision matrix, D, is defined as a finite $2^q \times q$ matrix of probabilities λ_{kj}, where each λ_{kj} is defined as follows for a finite stimulus set $\mathscr{Q}$:

$$\begin{aligned} \lambda_{kj} &= P(j \mid \mathscr{C}_k = \{i, j, ...\}) \\ &= \delta_j \Big/ \sum_{L \in \mathscr{C}_k} \delta_L, \qquad j \in \mathscr{C}_k, \quad \mathscr{C}_k \subseteq \mathscr{Q}, \end{aligned} \tag{12}$$

with the constraints that $0 \leqslant \delta_j \leqslant 1$ and $\sum_{j \in \mathscr{Q}} \delta_j = 1$. When $j \notin \mathscr{C}_k$, $\lambda_{kj} = 0$.

(c) For any single finite $q \times q$ confusion matrix Q, the NCC predicts the entries in Q as

$$\begin{aligned} Q &= AD \\ &\Rightarrow P_Q(j \mid i) = \sum_k \pi_{ik} \lambda_{kj}. \end{aligned} \tag{13}$$

The index k in the summation of Eq. (13) ranges over the number of possible confusion sets, $\mathscr{C}_k$, in a finite stimulus set $\mathscr{Q}$, which is 2^q. As noted, the NCC requires that $[q(q+1)/2]-1$ parameters be estimated.

A comparison of the above specialized version of the NCC with the original formulation (Nakatani, (972) will reveal that the probabilities ω_{ij} are being assumed to be free parameters that will be estimated. In the original formulation, the ω_{ij} probabilities are defined (not estimated) relative to a Euclidean multidimensional space. This space is occupied by sets of multidimensional Gaussian distributions (or spheres), with the center point of these spheres representative of a stimulus/response. A boundary lies within each of these spheres that defines an acceptance region. Response j is an acceptable reponse given that an observation falls within the acceptance region of the sphere whose center point represents the location of stimulus/response j. Therefore, ω_{ij} represents the probability that an observation will fall within the acceptance region of the sphere of response j given that stimulus i was presented. In the present version of Nakatani's model, the constraint that $\omega_{ij}=\omega_{ji}$ implies that the acceptance regions have equal diameter and variance. The constraint that $\omega_{ii}=1$ further implies that the variance of the multidimensional Gaussian distribution in the sphere surrounding stimulus/response i is small relative to the diameter of the acceptance region to ensure that response i is acceptable given that stimulus i was presented. The free parameters in the original model are the point coordinates in an L-dimensional Euclidean space and the bias probabilities, while the specialized version employed here begins on a more macroscopic level. This permits estimation and testing to be comparable to that of the other models investigated here.

III. Equivalence and Implication Relations and Measurement Constraints

Some knowledge has been gained concerning the equivalence relationships among the AON, OVLP, and WSCM. "Equivalence" in this discussion will be with respect to a single confusion matrix rather than across confusion matrices generated empirically by manipulation of experimental variables (such as stimulus related or response related variables), or theoretically by varying the parameter values of a model (see, e.g., Townsend, 1978). All of the models considered here are distinct if it is assumed that certain classes of parameters change with certain experimental manipulations while leaving the others unchanged. Within this context, it has been shown through parameter mappings from one model to another that AON and OVLP are special cases of the WSCM when the WSCM is restricted to the confusion data space of the AON and OVLP (Townsend, 1978; Pachella *et al.*, 1978). Therefore, if the AON and OVLP perfectly predict a confusion matrix, then the WSCM will perfectly fit the confusion matrix.

However, the converse need not be true as the AON and OVLP do not fill out the WSCM confusion space.

Although equivalence relations are important, a related important concern for the confusion models revolves around the necessary and sufficient measurement conditions that must exist in order for a confusion data space to be fit perfectly by the models. These conditions place restrictions on the admissable data so that there is quantitative assurance that the models will fit the data. By quantitative assurance we mean that, as long as the data meet certain requirements, parameter estimation is not needed in order to know that the model will be able to perfectly predict the data. Further, the identification of necessary and sufficient measurement constraints for the models sometimes permits model equivalences to be investigated in a way other than parameter mappings. If two models share the same necessary and sufficient constraints, they are equivalent. We have been able to determine nine measurement conditions that, in one way or another, the various models require in order to restrict the data in the confusion matrices and allow perfect predictive capability. These will be listed below and then discussed.[4]

It is pertinent to mention that other types of equivalence than that pursued here may be investigated. For example, van Santen and Bamber (1981) have shown that a large class of confusion models (which includes the SCM) may be represented as possessing an infinite number of sensory confusion states, but not as finite state models. They refer to the type of equivalence we are pursuing here as "data equivalence."

DEFINITION 8. For any single finite $q \times q$ confusion matrix Q, let $P_{ij} = P_Q(j \mid i)$ represent the probability that response j is confused for stimulus i in Q.[5] For $i, j, k, l \in \mathscr{Q}$, any, all, or none of the following conditions may hold in Q:

(a) bias transitivity (Holman, 1979)

$$P_{ji} \geqslant P_{ij} \text{ and } P_{kj} \geqslant P_{jk} \Rightarrow P_{ki} \geqslant P_{ik}, \qquad i \neq j \neq k,$$

(b) bias substitutability (Holman, 1979)

$$P_{ki} \geqslant P_{ik} \text{ and } P_{li} \geqslant P_{il} \Rightarrow P_{ji} \geqslant P_{lk} \text{ or } P_{ij} \leqslant P_{kl}, \qquad i \neq j \neq k \neq l,$$

(c) quasisymmetry (Caussinus, 1966)

$$P_{ij}P_{jk}P_{ki} = P_{ik}P_{kj}P_{ji}, \qquad i \neq j \neq k,$$

(d) diagonal maximization (Schulze, personal communication to the first author)

$$P_{ii}P_{jj} \geqslant P_{ij}P_{ji}, \qquad i \neq j,$$

(e) modified diagonal maximization (Schulze, Note 5)

$$P_{ii} > (P_{ji}P_{ik})/P_{jk}, \qquad i \neq j \neq k,$$

[4] These conditions have been obtained from several sources. Citations are made when appropriate.

[5] Notation is being simplified for this section since the conditions being discussed are applicable within a single confusion matrix only. That is, if condition (a), for example, holds in a master matrix, it does not necessarily hold in a subset matrix.

(f) symmetric cells inequality (Townsend, 1971a)

$$P_{ij} + P_{ji} \leqslant 1, \qquad i \neq j,$$

(g) column constraint (Townsend, 1971a)

$$P_{ii} \geqslant \sum_{\substack{j \neq i \\ = 1}}^{q} P_{ji},$$

(h) ratio constraint, two independent types (Townsend, 1978),

(1) P_{ik}/P_{jk} is constant over k, $\quad i, j \neq k$,

(2) P_{ij}/P_{ik} is constant over i, $\quad j, k \neq i$.

THEOREM 2. *The WSCM (i.e., Eq. (4))*[6] *holds if and only if* (i) *quasisymmetry and* (*ii*) *diagonal maximization hold.*

Proof. The necessity of (i) and (ii) can easily be proven by the reader. Sufficiency proceeds as follows:

Define η_{ij}, β_j as follows:

$$\eta_{ij} = \left(\frac{P_{ij}P_{ji}}{P_{ii}P_{jj}}\right)^{1/2}, \beta_j = \left(\frac{P_{jj}P_{mj}}{P_{jm}P_{mm}}\right)^{1/2}, \qquad i, j, m \in \mathscr{Q},$$

where m is abitrary but fixed. Then, by the similarity choice formula (i.e., Eq. (4)), it can be investigated whether

$$P_{ij} \overset{?}{=} \frac{\eta_{ij}\beta_j}{\sum_{k=1}^{q} \eta_{ik}\beta_k} = \frac{(P_{ij}P_{ji}/P_{ii}P_{jj})^{1/2}(P_{jj}P_{mj}/P_{jm}P_{mm})^{1/2}}{\sum_{k=1}^{q}(P_{ik}P_{ki}/P_{ii}P_{kk})^{1/2}(P_{kk}P_{mk}/P_{km}P_{mm})^{1/2}}.$$

By cancellation and noting from (i) that

$$\frac{P_{mj}P_{ji}}{P_{jm}P_{ij}} = \frac{P_{mi}}{P_{im}},$$

and dividing the numerator and denominator by the square root of the latter, we get

$$P_{ij} \overset{?}{=} \frac{P_{ij}}{\sum_{k=1}^{q}(P_{ik}P_{ki}P_{mk}P_{im}/P_{km}P_{mi})^{1/2}}.$$

Since, from (i),

$$\left(\frac{P_{ki}P_{im}P_{mk}}{P_{km}P_{mi}}\right)^{1/2} = (P_{ik})^{1/2},$$

[6] As noted, the equivalence relations and measurement conditions are being applied within the context of a single confusion matrix only. In this case, the additional notation of Definition 3(b) is not needed, and therefore WSCM and SCM denote the same model (Eq. (4)).

the expression becomes

$$\frac{P_{ij}}{\sum_{k=1}^{q} P_{ik}} = P_{ij},$$

and so there is indeed an identity. The remaining task is to check the parameter constraints. By definition, $\eta_{ii} = \eta_{jj} = 1$ and $\eta_{ij} = \eta_{ji}$. Finally,

$$\eta_{ij} = \left(\frac{P_{ij}P_{ji}}{P_{ii}P_{jj}}\right)^{1/2} \leqslant 1$$

by (ii). ∎

THEOREM 3 (From Schulze, personal communication to first author). *The OVLP (i.e., Definition* 6) *holds if and only if* (i) *quasisymmetry,* (ii) *symmetric cells inequality, and* (iii) *the column constraint hold.*

Proof. Let $G_{ij} = g_j/(g_j + g_i)$.

(a) Necessity.

Condition (i),

$$\begin{aligned} P_{ij}P_{jk}P_{ki} &= \xi_{ij}G_{ij} \cdot \xi_{jk}G_{jk} \cdot \xi_{ki}G_{ki} \\ &= \xi_{ji}G_{ji} \cdot \xi_{kj}G_{kj} \cdot \xi_{ik}G_{ik} \\ &= P_{ji}P_{kj}P_{ik} = P_{ik}P_{kj}P_{ji}, \end{aligned}$$

by virtue of $\xi_{ij} = \xi_{ji}$ and shifting the g's.

Condition (ii),

$$\begin{aligned} P_{ij} + P_{ji} &= \xi_{ij}G_{ij} + \xi_{ji}G_{ji} \\ &= \xi_{ij} \leqslant 1, \end{aligned}$$

by virtue of $\xi_{ij} = \xi_{ji}$ and $G_{ij} + G_{ji} = 1$.

Condition (iii),

$$\begin{aligned} P_{ii} = \xi_{ii} + \sum_{k \neq i}^{q} \xi_{ik}G_{ki} &\geqslant \sum_{k \neq i}^{q} \xi_{ik}G_{ki} \\ &= \sum_{k \neq i}^{q} \xi_{ki}G_{ki} = \sum_{k \neq i}^{q} P_{ki}, \end{aligned}$$

by virtue of $\xi_{ik} = \xi_{ki}$.

(b) Sufficiency.

Suppose conditions (i), (ii), and (iii) hold. Now define

$$\xi_{ij} = P_{ij} + P_{ji}; \qquad \xi_{ii} = P_{ii} - \sum_{k \neq i}^{q} P_{ki},$$

and $g_i = P_{ki}/P_{ik}$ for arbitrary but fixed $k \in \mathscr{Q}$. Clearly, by (ii) and (iii), $0 \leqslant \xi_{ij} \leqslant 1$ for $i, j \in \mathscr{Q}$. Further, for any $i \neq j$,

$$
\begin{aligned}
P_{ij} &= (P_{ij}P_{kj}P_{ik} + P_{ki}P_{ij}P_{jk})/(P_{kj}P_{ik} + P_{ki}P_{jk}) \\
&= (P_{ij}P_{kj}P_{ik} + P_{kj}P_{ji}P_{ik})/(P_{kj}P_{ik} + P_{ki}P_{jk}) \qquad \text{[from (i)]} \\
&= (P_{ij} + P_{ji}) \frac{P_{kj}/P_{jk}}{(P_{ki}/P_{ik}) + (P_{kj}/P_{jk})} \\
&= \xi_{ij} \times g_j/(g_i + g_j),
\end{aligned}
$$

as required when $i \neq j$. The same strategy may be used to show that

$$P_{ii} = \xi_{ii} + \sum_{k \neq i}^{q} \xi_{ik} \times g_i/(g_i + g_k),$$

and it will be left to the reader. ∎

The AON (i.e., Definition 5) is probably the most well known model of the multiplicative type (see Townsend, 1978). Townsend (1978) has shown that confusion data must follow quasisymmetry and the ratio constraints (both types 1 and 2) in order to be perfectly fit by a multiplicative model, while Schulze (Note 5) has shown that the ratio constraints and modified diagonal maximization are necessary and sufficient measurement conditions for the AON. Additionally, since the AON and OVLP are special cases of the WSCM, then diagonal maximization holds for these models also.

Holman (1979) has shown that the WSCM follows properties that he terms row and column biases. Although these are unimportant for the present discussion, what is important is that he shows that bias substitutability and bias transitivity follow from row and column biases. Since the WSCM has these properties, then (qualitatively speaking) both bias substitutability and bias transitivity should be observable in confusion matrices the model fits well.

The NCC is a member of the general class of models known as sophisticated guessing models. Actually, the general form of the sophisticated guessing model is given by Eq. (13) in Definition 7, where π_{ik} represents the probability a confusion state $\mathscr{C}_k$ arises from presentation of stimulus i, and λ_{kj} represents the probability that response j is given when confusion state $\mathscr{C}_k$ is centered. If Eq. (13) is taken without further constraints (i.e., ignoring parts a and b of Definition 7), it can be shown that this general form of the sophisticated guessing model requires that $3q2^{q-1} - 2q - 2^q + 1$ parameters be estimated. This is a rather large number when compared to the degrees of freedom in a confusion matrix (i.e., $q(q-1)$). Therefore, additional assumptions are used in various specific instances of the general sophisticated guessing model to make the specific models tractable and without a preponderance of parameters. It seems that three common assumptions are the following (e.g., Broadbent, 1967; Pachella *et al.*, 1978):

(1) $\pi_{ik} = 0$ if $i \notin \mathscr{C}_k$,

(2) $\pi_{ik} = \pi_{jk}$ if $i, j \in \mathscr{C}_k$, for all $\mathscr{C}_k \in \mathscr{C}$,

$$\begin{aligned}(3)\quad \lambda_{kj} &= g_j/\textstyle\sum_{k\in\mathscr{C}_k} g_k, & j &\in \mathscr{C}_k\,.\\ &= g_j, & \mathscr{C}_k &= \varnothing\\ &= 0, & j &\notin \mathscr{C}_k \neq \varnothing\end{aligned}$$

where $0 \leqslant g_j \leqslant 1$ and $\sum_{k=1}^{q} g_k = 1$.

It is clear from Definition 7 that the NCC satisfies conditions (1) and (3) but not (2). The OVLP is an instance of the sophisticated guessing model satisfying conditions (1), (2), and (3). From Definition 5 it can be seen that the AON is an instance of the sophisticated guessing model where there are $q+2$ confusion states, $\mathscr{C}_k = \{k\}$ and $\mathscr{C}_{\varnothing} = \{$the empty set$\}$, and conditions (1) and (3) are satisfied.

Theorem 4 establishes a more formal relationship between the WSCM and the sophisticated guessing models.

THEOREM 4. *Any sophisticated guessing model satisfying restrictions* (1), (2), *and* (3) *is a special case of the WSCM.*[7]

Proof. The proof proceeds by way of showing that such a sophisticated guessing model satisfies (i) quasisymmetry and (ii) diagonal maximization (cf. Theorem 2). Because of restrictions (1) and (3), the only confusion states relating to P_{ij} that have nonzero probability contain both i and j. Let π_{ij} be the probability that confusion state $\mathscr{C}_{ij} = \{i, j,...\}$ arises from the presentation of stimulus i. Let λ_j be the probability that response j is given when confusion state $\mathscr{C}_{ij}$ is entered. In general, there are $2^{q-2}\mathscr{C}_{ij}$'s and different λ_j's for each distinct $\mathscr{C}_{ij}$.

(1) Quasisymmetry:

$$P_{ij}P_{jk}P_{ki} \stackrel{?}{=} P_{ik}P_{kj}P_{ji}, \qquad i \neq j \neq k.$$

In terms of the general sophisticated guessing model (cf. Eq. (13)), this may be expressed as

$$\sum_{\mathscr{C}_{ij}} \pi_{ij}\lambda_j \cdot \sum_{\mathscr{C}_{jk}} \pi_{jk}\lambda_k \cdot \sum_{\mathscr{C}_{ki}} \pi_{ki}\lambda_i \stackrel{?}{=} \sum_{\mathscr{C}_{ik}} \pi_{ik}\lambda_k \cdot \sum_{\mathscr{C}_{kj}} \pi_{kj}\lambda_j \cdot \sum_{\mathscr{C}_{ji}} \pi_{ji}\lambda_i.$$

Because of restrictions (1) and (3) the sets $\{\mathscr{C}_{ij}\}$ and $\{\mathscr{C}_{ji}\}$ are the same. From restriction (3), the proposed identity becomes

$$\begin{aligned}&\sum_{\mathscr{C}_{ij}} \pi_{ij} \frac{g_j}{\sum_{m\in\mathscr{C}_{ij}} g_m} \cdot \sum_{\mathscr{C}_{jk}} \pi_{jk} \frac{g_k}{\sum_{m\in\mathscr{C}_{jk}} g_m} \cdot \sum_{\mathscr{C}_{ki}} \pi_{ki} \frac{g_i}{\sum_{m\in\mathscr{C}_{ki}} g_m}\\ &\quad \stackrel{?}{=} \sum_{\mathscr{C}_{ik}} \pi_{ik} \frac{g_k}{\sum_{m\in\mathscr{C}_{ik}} g_m} \cdot \sum_{\mathscr{C}_{kj}} \pi_{kj} \frac{g_j}{\sum_{m\in\mathscr{C}_{kj}} g_m} \cdot \sum_{\mathscr{C}_{ji}} \pi_{ji} \frac{g_i}{\sum_{m\in\mathscr{C}_{ji}} g_m}.\end{aligned}$$

[7] Smith (1980) presents a sophisticated guessing model he terms the symmetric sophisticated guessing model. This model follows restrictions (1), (2), and (3). Although the model has more parameters than there are degrees of freedom in the data, Smith shows that the model is still testable via the quasisymmetry measurement constraint and the constraint that $P_{ij} \leqslant P_{jj}$. Smith presented parametric mappings between this symmetric sophisticated guessing model and the WSCM that suggest the result shown by Theorem 4 (which was developed independently by the authors).

From restriction (2), $\pi_{ij} = \pi_{ji}$, and the fact that the g's in the numerators are independent of the confusion states $\mathscr{C}_{..}$, both sides of the proposed identity can be expressed as

$$\sum_{\mathscr{C}_{ij}} \sum_{\mathscr{C}_{jk}} \sum_{\mathscr{C}_{ki}} \pi_{ij}\pi_{jk}\pi_{ki} \frac{(g_j\, g_k\, g_i)}{\sum_{m\in\mathscr{C}_{ij}} g_m \sum_{m\in\mathscr{C}_{jk}} g_m \sum_{m\in\mathscr{C}_{ki}} g_m}.$$

Quasisymmetry is thereby satisfied.

(2) Diagonal Maximization:

$$P_{ii}P_{jj} \overset{?}{\geqslant} P_{ij}P_{ji}, \qquad i \neq j.$$

Noting that $\mathscr{C}_i = \{i,...\}$ (i.e., any confusion state containing at least stimulus i), from the general sophisticated guessing model (cf. Eq. (13)), diagonal maximization may be written as (using restriction (1)),

$$\sum_{\mathscr{C}_i} \pi_{ii}\lambda_i \cdot \sum_{\mathscr{C}_j} \pi_{jj}\lambda_j \overset{?}{\geqslant} \sum_{\mathscr{C}_{ij}} \pi_{ij}\lambda_j \cdot \sum_{\mathscr{C}_{ji}} \pi_{ji}\lambda_i.$$

Note that the set of states $\{\mathscr{C}_i\}$ can be broken down into a union of the set states $\{\mathscr{C}_{ij}\}$, those confusion states containing both stimuli i and j, and the set of states $\{\mathscr{C}_{i\bar{j}}\}$, those confusion states containing stimulus i but not stimulus j. That is,

$$\{\mathscr{C}_i\} = \{\mathscr{C}_{ij}\} \cup \{\mathscr{C}_{i\bar{j}}\}.$$

The above proposed inequality can be rewritten as

$$\left[\sum_{\mathscr{C}_{ij}} \pi_{ij}\lambda_i + \sum_{\mathscr{C}_{i\bar{j}}} \pi_{i\bar{j}}\lambda_i\right] \cdot \left[\sum_{\mathscr{C}_{ji}} \pi_{ji}\lambda_j + \sum_{\mathscr{C}_{j\bar{i}}} \pi_{j\bar{i}}\lambda_j\right] \overset{?}{\geqslant} \sum_{\mathscr{C}_{ij}} \pi_{ij}\lambda_j \cdot \sum_{\mathscr{C}_{ij}} \pi_{ji}\lambda_i.$$

From restriction (3), this becomes

$$\left[\sum_{\mathscr{C}_{ij}} \pi_{ij} \frac{g_i}{\sum_{m\in\mathscr{C}_{ij}} g_m} + \sum_{\mathscr{C}_{i\bar{j}}} \pi_{i\bar{j}} \frac{g_i}{\sum_{m\in\mathscr{C}_{i\bar{j}}} g_m}\right]$$
$$\cdot \left[\sum_{\mathscr{C}_{ji}} \pi_{ji} \frac{g_j}{\sum_{m\in\mathscr{C}_{ji}} g_m} + \sum_{\mathscr{C}_{j\bar{i}}} \pi_{j\bar{i}} \frac{g_j}{\sum_{m\in\mathscr{C}_{j\bar{i}}} g_m}\right]$$
$$\overset{?}{\geqslant} \sum_{\mathscr{C}_{ij}} \pi_{ij} \frac{g_j}{\sum_{m\in\mathscr{C}_{ij}} g_m} \cdot \sum_{\mathscr{C}_{ji}} \pi_{ji} \frac{g_i}{\sum_{m\in\mathscr{C}_{ji}} g_m}.$$

Since g_i and g_j cancel, and all terms in the expression are nonnegative, the inequality is proven. ∎

Part (1) of Theorem 4 succeeds because of all three of the restrictions being placed on the general sophisticated guessing model. From Definition 7(a), it can be seen that the NCC is *not* constrained by restriction (2). (I.e., $\pi_{ij} \neq \pi_{ji}$. You can enter the same confusion set with different probabilities depending upon which stimulus of that confusion set was presented.)

Although restriction (2) is used in the proof of quasisymmetry (see part (1), Theorem 4), the lack of restriction (2) in the NCC leaves its relationship with quasisymmetry still undetermined. That the NCC does not imply quasisymmetry will be shown through the following example. For simplicity, assume that a three-element stimulus set, $\{i, j, k\}$, is presented to the subject. Quasisymmetry predicts that the following condition will be observed in the resultant 3×3 confusion matrix (cf. Definition 8(c)):

$$\left(\frac{P_{ij}}{P_{ji}}\right) \cdot \left(\frac{P_{jk}}{P_{kj}}\right) \cdot \left(\frac{P_{ki}}{P_{ik}}\right) = 1.$$

Using the first ratio as an example, the left-hand side of the above equation can be rewritten in terms of the NCC parameters as follows (again let $\mathscr{C}_{ij}$ = confusion set given stimulus i and containing j as well as i):

$$\frac{P_{ij}}{P_{ji}} = \frac{\sum_{\mathscr{C}_{ij}} \pi_{ij} \lambda_j}{\sum_{\mathscr{C}_{ji}} \pi_{ji} \lambda_i} = \frac{\sum_{\mathscr{C}_{ij}} [\prod_{1 \in \mathscr{C}_{ij}} \omega_{i1} \prod_{m \notin \mathscr{C}_{ij}} (1 - \omega_{im})\, g_j / \sum_{n \in \mathscr{C}_{ij}} g_n]}{\sum_{\mathscr{C}_{ji}} [\prod_{1 \in \mathscr{C}_{ji}} \omega_{j1} \prod_{m \notin \mathscr{C}_{ji}} (1 - \omega_{jm})\, g_i / \sum_{n \in \mathscr{C}_{ji}} g_n]}.$$

The summation in both the numerator and denominator ranges over those subsets of the three-element stimulus set that certain at least stimuli i and j. There are only two such subsets, $\{i, j\}$ and $\{i, j, k\}$, and so the ratio can be expanded as

$$\frac{g_j}{g_i} \cdot \left[\frac{\omega_{ii}\omega_{ij}(1 - \omega_{ik})/(g_i + g_j) + \omega_{ii}\omega_{ij}\omega_{ik}/(g_i + g_j + g_k)}{\omega_{jj}\omega_{ji}(1 - \omega_{jk})/(g_i + g_j) + \omega_{jj}\omega_{ji}\omega_{jk}/(g_i + g_j + g_k)}\right].$$

Since $\omega_{ii} = \omega_{jj} = 1$, $\omega_{ji} = \omega_{ij}$, and $g_i + g_j + g_k = 1$, this can be simplified to

$$\frac{g_j}{g_i} \cdot \left[\frac{(1 - \omega_{ik}) + \omega_{ik}(g_i + g_j)}{(1 - \omega_{jk}) + \omega_{jk}(g_i + g_j)}\right].$$

Rewriting P_{jk}/P_{kj} and P_{ki}/P_{ik} in terms of the NCC parameters as was done above for P_{ij}/P_{ji}, and cancelling the g's, quasisymmetry implies that

$$\frac{(1 - \omega_{ik}) + \omega_{ik}(g_i + g_j)}{(1 - \omega_{jk}) + \omega_{jk}(g_i + g_j)} \cdot \frac{(1 - \omega_{ji}) + \omega_{ji}(g_j + g_k)}{(1 - \omega_{ki}) + \omega_{ki}(g_j + g_k)} \cdot \frac{(1 - \omega_{kj}) + \omega_{kj}(g_k + g_i)}{(1 - \omega_{ij}) + \omega_{ij}(g_k + g_i)} = 1.$$

As an example, letting $\omega_{ij} = \omega_{ji} = .80$, $\omega_{ik} = \omega_{ki} = .70$, $\omega_{jk} = \omega_{kj} = .60$, $g_i = .50$, $g_j = .30$, and $g_k = .20$, and simplifying, yields:

$$\frac{.42312}{.43472} \neq 1.$$

A similar strategy can be used for stimulus sets with more than three elements, although the equations become more complex due to the increase in the number of the relevant subsets.

It is obvious from the above discussion that quasisymmetry is an important condition, entering into the structure of a number of models, and as a test between models. However, we have found it difficult to test, due in part to the very small numbers that often result from the multiplication of the off-diagonal proportions. Additionally, it follows from the above discussion that since the NCC does not imply quasisymmetry, it is not a special case of the WSCM. However, we have been unable to develop an analytic proof showing that the WSCM is, or is not, a special case of the NCC. We have done some computer simulations indicating that confusion matrices generated by certain extreme patterns of WSCM parameter values are not fit well by the NCC, suggesting that the WSCM is not a special case of the NCC. However, these results are tentative and will not be presented.

One last point to note is that part (2) of Theorem 4 depends only upon restrictions (1) and (3), which are constraints of the NCC (cf. Definition 7). Therefore, diagonal maximization is implied by the NCC.

To summarize, Fig. 1 presents a diagram of the relationships between models and between models and measurement conditions established by the theorems and discussion in this section and Appendix B. In addition, certain additional obvious results are included in Fig. 1, such as the fact that NCC does not imply OVLP (because it need not satisfy the column constraint). It may be that neither AON nor OVLP implies NCC because of the multinomial form of the latter but this has not been proven. It is also easy to show that OVLP implies diagonal maximization but this is not shown in Fig. 1 to prevent clutter.

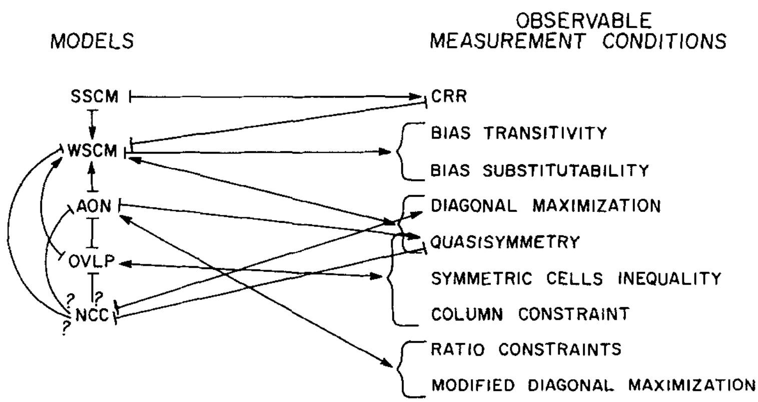

FIG. 1. A summary of implication (→) and failure of implication (—ı) relations between models and measurement conditions. A question mark denotes an unestablished relationship.

IV. Empirical Investigation of the Confusion Models

Empirical investigations of perceptual models can proceed in several ways. One often used method of testing is to construct a stimulus set and examine how well the models predict the results derived from that stimulus set. Other stimulus sets can then be developed to test the generality or limitations of the models across the various types of stimulus sets. The present perceptual models, along with the CRR, have been previously tested in this manner, numerous citations having been presented earlier in the paper.

A different method of testing these types of models is to systematically examine their structure, which includes both their theoretical underpinnings and the interpretations of their parameters. That is, a single well defined stimulus set (not necessarily "realistic," but hopefully so) can be constructed and systematic subsets of this set examined to determine how far and in what ways the models can be "pushed" with respect to their measurement constraints and the invariances of their parameters across the different subsets. The results of such a test would help to formulate perceptual processing mechanisms based on the models' structures, an important concern in the case of the SCM where no accepted processing interpretation yet exists.

The models presented in this paper have not yet been thoroughly investigated in this manner. Thus, the SCM can be tested with respect to the invariance of its parameters across context manipulations, various across-set predictor versions of the SCM having been constructed with respect to the CRR in Section I of this paper. Parameter invariances can also be investigated in the confusion models presented in Section II. Additionally, the measurement constraints of the models established in Section III can be examined both within and across sets. The present experiment was designed to generate data that could be used explicitly for such tests as proposed here, there being little, if any, data extant in the literature suitable for the purpose. Moreover, the SCM and CRR tests are particularly important because: (a) SCM has overall fared better than competing models, (b) it is therefore of substantial interest to learn whether behavior that satisfies WSCM might be generated by a stronger deeper principle which implies the CRR version of "independence of irrelevant alternatives." The experiment is *not* simply an attempt to determine if a visual stimulus set can be discovered that fails to satisfy CRR.

With this motivation in mind, the stimulus set selected for the present study was chosen on the basis of being able to provide a straightforward manipulation of set context and size as well as being identical in form to a previous study (Townsend & Ashby, Note 2). Five block letters (A, E, F, H, X) were constructed of equal length line segments with the aid of a square template, with the letter X comprising the diagonals of the template (see Fig. 2). The letters A, E, F, H were selected due to their high overlap of common line segments and angles in this font in contrast to the letter X. By adding and subtracting the letters A, E, F, H and the letter X, a context manipulation can be achieved between various subsets of the five letters in independent presentations to the subject of the various subsets of the letters (i.e.,

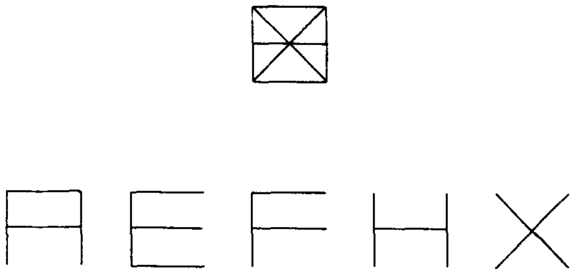

FIG. 2. Block letters employed as stimuli.

independent blocks of presentations, each block consisting of a different subset of letters). The confusion matrices so generated can then be used to compare the performances of the models both within and across the various confusion matrices.

Four different sets of the five letters (A, E, F, H, X) were employed as independent stimulus sets in the present study. One set served as the master set from which the CRR and the SSCM predictions were obtained, and consisted of all five letters. A four letter subset (A, E, F, H) was employed in order to examine the relationships between these letters and the effects of these relationships on model predictions. Two three letter subsets, one consisting of the letters, A, E, X and the other of the letters F, H, X, were also employed as independent stimulus sets. The A, E, X subset was used to determine the effects of letter dissimilarity on model predictions. A previous study (Townsend & Ashby, Note 2) has shown that the letters A, H are highly similar, as are E, F, in the font being used here. In order to create a dissimilar subset, the letters A, E, X were selected. The dissimilarity of the letter X was assumed. The F, H, X set was used as a three letter control for comparison with the A, E, X set. A previous study (Townsend, 1971a), using a font fairly similar to the one used here, suggested that the letters F, H, X are confused with each other to about the same degree, and presumably more than the letters A, E, X. As will be seen, the latter supposition turned out to be questionable in the present experiment.

Debreu (1960) suggested a criticism of the choice axiom that concerned the assumption of independence from irrelevant alternatives as expressed by Eq. (2). Debreu pointed out that adding alternatives to a choice set that are very similar to one of the alternatives already present in the choice set would probably affect the relative choice probabilities of the similar alternatives, and not those of the dissimilar alternatives. In other words, dependencies may exist between alternatives in a choice set, and these dependencies may be selectively altered by the appropriate choice of alternatives for addition to or deletion from the offered choice set. This criticism bears a strong resemblance to the differences exhibited between the stimulus sets being used in the present study, although the sets were not constructed specifically with Debreu's criticism in mind. Even though Debreu's criticism was more concerned with the general area of human preference, recent theoretical work in perception (e.g., Krumhansl, 1978) indicates that the specific criticism of Debreu may have some application to certain psychophysical areas of research.

The CRR is unable to predict reference data obtained with sets constructed in the manner that was outlined above (e.g., Becker, DeGroot, & Marschak, 1963; Tversky, 1972), and may not be able to predict psychophysical data of a similar nature. As shown by Theorem 1, the SSCM is a special case of the CRR and so should perform no better than the CRR. The WSCM may be able to predict the data quite well, as it can be fit to the obtained confusion matrices separately without any violations of its parameter constraints. Neither the SSCM(η) nor the SSCM(β) implies the CRR and should predict the subsets somewhat better than the SSCM and somewhat worse than the WSCM.

Method

Subjects

The subjects were four Purdue University students who participated for credit in an experimental psychology course. Each subject was individually tested in daily 1-hr sessions for 16 days, excluding Sundays. Before the experiment proper commenced, each subject underwent a calibration and practice period of 4 days in which each subject's average accuracy of correct responses was stabilized at between .55 and .65. Neither the calibration data nor the practice data were included in the analyses presented for the present study.

Apparatus

A Gerbrands two-field tachistoscope (model T-2B) was used to present the stimuli.

The five stimulus letters (A, E, F, H, X) were drawn in black ink with the aid of a square template, and were presented on 5×8 white index cards, one letter per card. The letters A, E, F, H were constructed of equal length line segments (the length of a side of the square template), with the two line segments of the letter X comprising the diagonals of the square template (see Fig. 2).

A pre- and post-stimulus fixation field was described by a set of four dots arranged at the corners of a square with the letter appearing in the center. The four dots were present at all times except for the intervals of stimulus presentation. The fixation field on any one side subtended a visual angle of about 4°, and a letter subtended a visual angle of about $\frac{1}{2}$°. The luminance for the pre- and post-stimulus fixation field was held at 35.63 cd/m^2 and the luminance for the stimulus field was held at 33.00 cd/m^2.

Procedure

Each experimental session consisted of four blocks of presentations. One block consisted of the letters A, E, F, H, X (master set). Another block consisted of the letters A, E, X. A third block consisted of the letters F, H, X. A fourth block consisted of the letters A, E, F, H. The four different blocks of presentations were counterbalanced across sessions with respect to presentation order so that each different block appeared four times in each of the four within-session serial positions. The letters for each block were randomly presented for a total of 15 trials per letter per block in which it appeared. There was a total of 225 trials per session.

When summed over blocks, each letter appeared 45 times per session. When summed over sessions, each letter appeared 240 times for each block in which it occurred.

The subject was informed of the block presentation order before each session began, and was reminded of the letters in each block prior to their presentation. The presentation rate was regulated by the experimenter under the constraint that no 2 trials occur within 8 sec of each other. Responses were given verbally and consisted of the subject naming the letter he or she thought had been presented on that trial. Stimuli and responses were recorded by hand on appropriate record keeping sheets. Prior to each experimental session, the subject was given 15 practice trials consisting of three presentations each of the five letters, randomly ordered.

Results and Discussion

(i) *Analysis of the Constant-Ratio Rule and Choice Model*

Tables 1–4 contain the obtained and predicted response proportions for the four individual subjects. The four sets presented in this study appear as groups of their respective stimuli along the ordinates of Tables 1–4. Each confusion cell in Tables 1–4 has been quartered. The obtained response proportions appear in the upper left quadrant of each cell. Predicted response proportions for the WSCM, SSCM, and CRR appear in the upper right, lower right, and lower left quadrants of each cell, respectively. A line has been drawn in the confusion cells not applicable to a specific matrix. The WSCM predictions were of course elicited by completely re-estimating the η_{ij}'s and β_j's for each matrix. SSCM predictions on the other hand, were obtained by estimating the η_{ij}'s and β_j in the master matrix and then carrying them over to predict the subset matrices. These estimates and the fit statistics were provided by Chandler's χ^2 STEPIT (1959) program. Similarly, CRR generated entries for Table 4 by Eq. (1). That is, each appropriate cell of the master matrix M was normalized by the sum of the appropriate cells in the same row, thus forming a prediction entry for the subset matrix S.[8] Under the stated procedure no CRR predicted proportion existed for the master matrix (A, E, F, H, X). The WSCM and SSCM, likewise, produce equivalent predicted response proportions in all cases for the master matrix, and are reported in the lower right quadrant (SSCM) of the master matrix (i.e., the top 5×5 matrix) in each of Tables 1–4. A guide for the identification of the confusion cell entries has been included in a footnote beneath each table.

[8] In order to save space, the parameter estimates of all of the models fit to the data, via STEPIT, have been omitted. Certain other numerical results have also been omitted. The parameter estimates and other numerical results that are discussed but not reported are available from the authors upon request.

TABLE 1

Obtained and Predicted Response Proportions: Subject *D.X.*

Stimulus sets	Response: *A*[a]	*E*	*F*	*H*	*X*
A	.442 .442	.083 .089	.121 .115	.225 .223	.129 .131
E	.129 .124	.450 .450	.241 .249	.088 .088	.091 .089
F	.117 .123	.196 .189	.425 .425	.079 .082	.183 .181
H	.159 .160	.046 .045	.058 .055	.633 .633	.104 .107
X	.033 .031	.012 .015	.038 .040	.038 .035	.879 .879
A	.446 .444 .507 .507	.108 .129 .096 .104	.175 .185 .139 .127	.271 .242 .258 .262	—
E	.204 .184 .142 .138	.433 .429 .495 .495	.275 .250 .267 .268	.088 .137 .096 .099	—
F	.217 .206 .143 .149	.171 .195 .240 .236	.483 .482 .520 .512	.129 .117 .097 .103	—
H	.107 .196 .177 .179	.113 .077 .051 .051	.070 .084 .065 .060	.650 .643 .707 .710	—
A	.620 .621 .675 .667	.217 .220 .127 .137	—	—	.163 .159 .198 .196
E	.208 .205 .193 .189	.667 .667 .670 .673	—	—	.125 .128 .137 .138
X	.046 .049 .036 .033	.046 .043 .014 .017	—	—	.908 .908 .950 .950
F	—	—	.671 .671 .618 .609	.196 .188 .115 .122	.133 .141 .267 .269
H	—	—	.108 .116 .073 .068	.763 .762 .796 .801	.129 .122 .131 .131
X	—	—	.054 .047 .039 .041	.058 .066 .039 .036	.888 .887 .922 .923

[a] Values represent

OBT = obtained response proportions, WSCM = weak similarity choice model predicted proportions,
CRR = constant-ratio rule predicted proportions, SSCM = strong similarity choice model predicted proportions.

TABLE 2

Obtained and Predicted Response Proportions: Subject *M. X.*

Stimulus set	Response A^{a}	F	F	H	X
A	.371	.138	.154	.258	.079
	.370	.133	.138	.265	.094
E	.129	.417	.208	.104	.142
	.133	.417	.207	.098	.145
F	.142	.238	.421	.087	.112
	.159	.238	.419	.092	.092
H	.242	.079	.075	.525	.079
	.235	.086	.071	.525	.083
X	.100	.137	.046	.092	.625
	.086	.133	.073	.087	.621
A	.471 .470	.121 .136	.129 .137	.279 .257	—
	.403 .414	.149 .149	.167 .149	.281 .288	
E	.183 .169	.463 .462	.233 .242	.121 .127	—
	.151 .159	.485 .493	.243 .234	.121 .114	
F	.192 .184	.267 .258	.429 .429	.112 .129	—
	.160 .182	.267 .267	.474 .452	.099 .099	
H	.192 .215	.091 .085	.096 .081	.621 .619	—
	.263 .263	.086 .098	.081 .074	.570 .565	
A	.683 .683	.183 .174	—	—	.134 .143
	.631 .620	.234 .223			.135 .157
E	.217 .226	.621 .621	—	—	.162 .153
	.188 .194	.606 .601			.206 .205
X	.162 .153	.117 .126	—	—	.721 .721
	.116 .105	.159 .158			.725 .737
F	—	—	.600 .599	.212 .228	.188 .173
			.678 .692	.141 .152	.181 .156
H	—	—	.163 .148	.725 .724	.112 .128
			.110 .101	.773 .772	.117 .127
X	—	—	.137 .153	.188 .173	.675 .674
			.060 .091	.120 .110	.820 .799

[a] Values represent

OBT = obtained response proportions, WSCM = weak similarity choice model predicted proportions,

CRR = constant-ratio rule predicted proportions, SSCM = strong similarity choice model predicted proportions.

TABLE 3

Obtained and Predicted Response Proportions: Subject $G.X.$

Stimulus set	Response: A^a	E	F	H	X
A	.504 .502	.092 .088	.158 .172	.175 .152	.071 .086
E	.088 .91	.575 .573	.208 .189	.058 .077	.071 .070
F	.146 .132	.121 .142	.562 .561	.092 .082	.079 .083
H	.138 .162	.096 .080	.104 .114	.558 .555	.104 .089
X	.092 .079	.063 .063	.104 .100	.058 .077	.683 .681
A	.571 .571 .543 .544	.117 .130 .099 .098	.121 .121 .170 .186	.191 .178 .188 .172	—
E	.129 .117 .094 .098	.613 .612 .619 .611	.175 .178 .224 .205	.083 .093 .063 .086	—
F	.108 .108 .158 .141	.179 .176 .131 .154	.608 .608 .611 .615	.104 .108 .100 .090	—
H	.150 .164 .154 .175	.104 .095 .107 .086	.117 .112 .116 .121	.629 .629 .623 .617	—
A	.692 .692 .756 .740	.179 .181 .138 .133	—	—	.129 .127 .106 .127
E	.146 .143 .119 .125	.725 .725 .784 .776	—	—	.129 .132 .097 .099
X	.087 .090 .109 .095	.121 .118 .075 .078	—	—	.792 .792 .816 .827
F	—	—	.646 .645 .767 .771	.221 .207 .125 .113	.133 .148 .108 .116
H	—	—	.175 .190 .136 .146	.696 .695 .728 .742	.129 .115 .136 .112
X	—	—	.150 .136 .123 .119	.100 .115 .069 .089	.750 .749 .808 .792

[a] Values represent
OBT = obtained response proportions, WSCM = weak similarity choice model predicted proportions,
CRR = constant-ratio rule predicted proportions, SSCM = strong similarity choice model predicted proportions.

TABLE 4

Obtained and Predicted Response Proportions: Subject $A.X.$

Stimulus set	Response: A^a	E	F	H	X
A	.429 .428	.088 .100	.179 .167	.254 .252	.050 .053
E	.150 .138	.471 .469	.242 .227	.079 .089	.058 .077
F	.146 .159	.142 .156	.425 .424	.183 .166	.104 .095
H	.113 .114	.038 .029	.058 .079	.742 .738	.050 .040
X	.050 .047	.062 .048	.079 .088	.067 .078	.742 .739
A	.517 .517 .452 .457	.070 .073 .092 .104	.171 .166 .188 .174	.242 .244 .268 .265	—
E	.129 .127 .159 .150	.463 .461 .500 .506	.300 .279 .257 .243	.108 .133 .084 .101	—
F	.158 .163 .163 .178	.137 .159 .158 .172	.513 .511 .474 .468	.192 .167 .205 .182	—
H	.142 .138 .118 .119	.062 .043 .040 .031	.067 .096 .061 .080	.729 .723 .781 .770	—
A	.700 .700 .757 .740	.213 .207 .155 .168	—	—	.087 .093 .088 .092
E	.258 .264 .221 .204	.621 .621 .693 .688	—	—	.121 .115 .086 .108
X	.129 .124 .059 .058	.117 .122 .073 .056	—	—	.754 .754 .868 .886
F	—	—	.613 .613 .597 .622	.223 .234 .257 .241	.154 .153 .146 .137
H	—	—	.142 .141 .068 .089	.775 .775 .872 .862	.083 .084 .059 .049
X	—	—	.142 .142 .089 .095	.129 .129 .075 .092	.729 .729 .836 .813

[a] Values represent

OBT = obtained response proportions, WSCM = weak similarity choice model predicted proportions,

CRR = constant-ratio rule predicted proportions, SSCM = strong similarity choise model predicted proportions.

A non-parametric test of the obtained response proportions and the CRR and SSCM predicted response proportions is given in the upper half of Table 5 (the lower half will be explained below). This method of analysis was employed in some of the earlier studies of the CRR (e.g., Clarke, 1957), prior to Morgan's (1974) development of his likelihood ratio test for CRR predictions. The first row in the upper half of Table 5 reports the mean of the obtained response proportions of correct responses (i.e., entries on the negative diagonal) in the various matrices. The second row reports the mean of the absolute difference scores between the obtained and predicted proportions in the various matrices. The third row reports the percentage of the absolute difference scores that exceeded .10 (a criterion suggested by Clarke, 1957). A second criterion of .05 is reported in the fourth row due to the 0% entry in the third row of the A, E, F, H subset matrix for both the CRR and SSCM. The individual subject results were combined for the non-parametric analysis in the upper half of Table 5.

The non-parametric analyses of the CRR and SSCM presented in Table 5 appear to support the predictive capability of both the CRR and SSCM according to the criteria outlined by the earlier studies of the CRR. In order to provide more stringent

TABLE 5

Analysis of CRR and SSCM Predictions

	Non-parametric tests[a]					
	CCR			SSCM		
Test[a]	A, E, X[b]	F, H, X	A, E, F, H	A, E, X	F, H, X	A, E, F, H
Mean proportion correct	.709	.711	.540	.709	.711	.540
Mean difference	.039	.051	.028	.038	.050	.028
Percentage differences >.10	2.8	11.1	0.0	2.8	11.1	0.0
Percentage differences >.05	30.6	47.2	20.3	25.0	36.1	20.3
	χ^2 Goodness of fit tests[c]					
	CRR			SSCM		
Subject	A, E, X $df = 6$	F, H, X $df = 6$	A, E, F, H $df = 12$	A, E, X $df = 6$	F, H, X $df = 6$	A, E, F, H $df = 12$
$A. X.$	42.39*	47.17*	16.19	51.74*	28.17*	24.12
$G. X.$	18.81*	31.40*	19.85	14.03*	35.79*	15.69
$M. X.$	14.30*	58.94*	17.57	17.29*	42.81*	13.45
$D. X.$	39.47*	38.85*	51.58*	27.72*	40.14*	51.03*

[a] See text for explanation of the measures.
[b] The letters represent the confusion matrix containing those letters.
* $p < .05$.

TABLE 6

Morgan's Likelihood Ratio Test for CRR Predictions

Subject	Matrix: A, E, X $df = 6$	F, H, X $df = 6$	A, E, F, H $df = 12$
$A. X.$	14.41*	15.78*	7.70
$G. X.$	7.00	11.57	9.78
$M. X.$	6.14	18.10*	8.67
$D. X.$	10.42	16.71*	18.73
Group	12.49	40.61*	9.81

* $p < .05$.

tests, some stricter parametric analyses were conducted for both the CRR and SSCM predictions. The likelihood ratio test of CRR predictions proposed by Morgan (1974) was conducted on both the individual subject and grouped data. (The grouped data were obtained by summing the numbers of confusions in corresponding confusion matrix cells across subjects. The resultant grouped data represented response proportions out of 960 total for each row in each of the four confusion matrices.) The results are presented in Table 6, and support the non-parametric analyses of the CRR (upper half of Table 5) in that the A, E, X and A, E, F, H subset matrices were predicted well, but the F, H, X subset matrix was not. These results suggest that Morgan's (1974) rejection of the CRR might have been due to the use of grouped data. However, in the present study the grouped data for the A, E, X and A, E, F, H subset matrices were predicted fairly well according to the likelihood ratio test. The discrepancies between these results and those of Morgan (1974) could be due to differences in modality (Morgan's data were derived from auditory confusions) and design (see Clarke, 1957, and Egan, 1957).

SSCM, as well as CRR, was then tested according to a χ^2 test of deviations. The SSCM test was based on the STEPIT-estimated parameters whereas the theoretical entries of CRR for a subset matrix S simply came, as noted above, directly from M. Although it appears possible to provide a likelihood ratio test for SSCM, the χ^2 test was chosen for purposes of comparison with the other models below as well as with past studies. The lower half of Table 5 presents χ^2 goodness of fit values for the CRR and SSCM predicted subset matrices. The χ^2 values in the lower half of Table 5 indicate that neither the CRR nor the SSCM predicted the A, E, X and F, H, X subset matrices with a high degree of accuracy. Both the CRR and SSCM fitted excellently, with the exception of $D.X.$, in the A, E, F, H subset matrices.[9]

[9] First, on face value on might expect, if $L =$ Morgan's likelihood ratio value, that $-2 \log L$ should about equal our χ^2 values, on the basis of large sample statistical theory. It is true that these values must ultimately approach one another for very large samples. However, the likelihood ratio in essence forms an estimate of the "true" subset matrix using both S and M under the null hypothesis. The χ^2 test, on the

The pattern of results presented in Table 5 suggests that both the CRR and SSCM provided a good first, but not second, order approximation in the prediction of the A, E, X and F, H, X subset matrices with F, H, X being especially suspect. The CRR and SSCM are apparently unable to predict the patterns of the results obtained in the A, E, X and F, H, X subset matrices with high accuracy as they both predict that the confusions for the missing stimuli (i.e., A, E or F, H) would be evenly distributed across the confusions with the remaining stimuli. This follows from the assumption of independence from irrelevant alternatives that underlies both models. In the present study, this assumption suggests that subtracting letters from the master set should not have any effect on the relative recognition probabilities between the remaining letters, whether they are presented in the context of the master set or in a subset by themselves. This result was obtained with the A, E, F, H subset, and to a moderate extent with the A, E, X subset, but not at all with the F, H, X subset.

What happened can be seen by a visual inspection of Tables 1–4. The confusions among the four letters A, E, F, H in the master matrix tended to concentrate into the two letters A, E and F, H in their respective subset matrices rather than spreading evenly across all three letters in the subset matrices as predicted by the CRR and SSCM. Both the CRR and SSCM tended to underestimate the $P(E \mid A)$ and $P(A \mid E)$ confusion entries in the A, E, X subset matrix and the $P(F \mid H)$ and $P(H \mid F)$ confusion entries in the F, H, X subset matrix. That the CRR and SSCM also underestimated the $P(j \mid X)$ (for $j = A, E, F, H$) entries in the A, E, X and F, H, X subset matrices suggests that the letter X was quite dissimilar to all of the letters A, E, F, H in this study. (This was not assumed *a priori*. Recall that an earlier study suggested that F, H might be more similar to X than would A, E.) This analysis indicates that both the CRR and SSCM are unable to compensate for the failures inherent in the strong assumptions underlying their development, specifically the assumpion of the independence from irrelevant alternatives. The lack of high precision of prediction in the subset matrices containing a highly dissimilar object (here, the letter X) may indicate some relevance of Debreu's (1960) criticisms of the choice axiom that were mentioned earlier.

One could disagree with the above analysis as there is a confounding of changes in the number of stimuli in the subsets with changes in the letter constituency of those subsets. It may be argued that the change in the number of stimuli could be responsible for the poorer performance of the CRR and SSCM in the 3×3 subset matrix. That is, reducing the set size from 5 to 4 is a smaller change in number than reducing the set size from 5 to 3. For this to be a plausible explanation, the CRR and SSCM should predict the 4×4 subset matrices better than the 3×3 subset matrices and

other hand, is testing the hypothesis that S is composed of *exactly* the values estimated from M. Because of the variance of both M and S, the sample χ^2 value test is going to have greater variance than $-2 \log L$ and will be conservative with respect to accepting the CRR hypothesis. In any case, the basic outcome was the same for both tests. Second, the fact that SSCM sometimes fits a little better than CRR is not really a paradox. SSCM is indeed a special case of CRR in that if SSCM fits both matrices perfectly so must CRR but not vice versa. However, if SSCM does not fit M perfectly, then it is quite possible for the S predictions to be better than those from CRR.

worse than the 5×5 master matrix. The χ^2 goodness of fit values in Tables 5 and 7 can be compared to see if this is the case. While all χ^2 fit values of the 4×4 subset matrix reported in Table 5 are larger than the corresponding values in Table 7 (the SSCM master matrix χ^2 values are where the M row intersects the WSCM column in Table 7),[10] the 3×3 χ^2 fit values in Table 5 fail to exceed the corresponding 4×4 χ^2 values 7 out of 16 times, or 44%. This result, plus analyses to be presented later, make an explanation based solely on changes in the number of stimuli rather than changes in set constituency unlikely.

Additional insight as to the failure of the SSCM to highly predict the A, E, X and F, H, X subset matrices can be obtained by an examination of the SSCM(η) and the SSCM(β) models. These models were applied by estimating both parameters from M and then re-estimating the flexible parameter (η and β, respectively) in S. This analysis indicated the SSCM(η) did somewhat better than the SSCM(β) in predicting the three subset matrices (7 significant χ^2 fit values, 4 at the .01 level, vs 9 dignificant χ^2 fit values, 7 at the .01 level, respectively, over a total of 12 χ^2 fits; cf. Table 5).[11] However, a more revealing analysis is obtained from the fact that the SSCM, SSCM(η), SSCM(β), and WSCM are nested versions of the same model. (I.e., Eq. (4). This is readily apparent from Definitions 2–4.) As such, it would be of interest to see if any of the χ^2 fits of the SSCM, SSCM(η), SSCM(β), and WSCM are significantly different from each other (using a nested χ^2 test). This information can then be used to help determine whether the η or β measure is more sensitive to constraint, and is thus making the larger contribution to the poor performance of the SSCM. This analysis indicated that the SSCM(η) produced more significant improvements over the SSCM in predicting the three submatrices than did the SSCM(β). Further the WSCM produced fewer significant improvements over the SSCM(η) than it did over the SSCM(β). [Out of a total of 12 χ^2 differences, for the SSCM–SSCM(η) comparison there were 9 significant χ^2 differences, 8 at the .01 level; for the SSCM–SSCM(β) comparison there were 7 significant χ^2 differences, 4 at the .01 level; for the SSCM(η)–WSCM comparison there were 6 significant χ^2 differences, 3 at the .01 level; and for the SSCM(β)–WSCM comparison there were 8 significant χ^2 differences, all at the .01 level.] These findings indicate that the change in the alphabet from the master set to the subsets is disturbing both the sensory-similarity structure and the bias structure of the confusion matrices, but the sensory-similarity structure more so than the bias structure. An examination of the estimated η's for the WSCM revealed that similarity (as measured by η) increased an average of 69% of the time (within each subject, 12 comparisons were possible) when descending from the master to the subset matrices. The β's, on the other hand, showed no systematic variation.

In order to learn to what degree the improvement of the WSCM over the SSCM was due to the minimization of the χ^2 over all of the matrices in the WSCM, the

[10] The CRR will return χ^2 values of 0 if "fit" to the master matrix as it uses the probabilities in the master matrix for predictions.

[11] A rejection level of .05 was used for significance calculations. A note is included for the number of significant values that could also be rejected at the .01 level.

SSCM was simultaneously fit to all four matrices within each subject. Although the overall χ^2's were, as expected, improved over what they had been when the SSCM parameters were simply taken from the master matrix, they still were not as good as when the WSCM was employed in separate fits.

Overall, we can characterize the results with the present class of choice models (and the CRR) as establishing that the CRR holds to a reasonable first approximation. Further, the SSCM performed about as well as the CRR, even though it is a special case of the CRR. The SSCM also has the advantage of possessing interpretable parameter structure. However, the change in the alphabet in the subsets brought about an alteration in the similarity structure of the resultant confusion matrices so that a high degree of prediction was possible only with the WSCM, fit as it was to each separate confusion matrix.

(ii) *Analysis of the WSCM, OVLP, AON, and NCC Models*

Table 7 presents χ^2 goodness of fit values for the WSCM, OVLP, AON, and NCC

TABLE 7

STEPIT χ^2 Goodness of Fit Values: Best Fitting Model

Subject	Matrix[a]	NCC	WSCM	OVLP	AON
D.X.	*M*	.80	.71	6.71	77.24*
	A	.15	.18	.18	.18
	F	.82	.87	.87	.87
	C	12.77*	13.71*	14.02*	43.78*
M.X.	*M*	6.55	6.27	16.77*	81.60*
	A	.88	.82	.82	.82
	F	2.05	1.95	1.95	1.95
	C	3.04	3.44	3.89	53.91*
G.X.	*M*	8.90	8.80	26.71*	29.43*
	A	.08	.07	.07	.07
	F	2.07	2.10	2.10	2.10
	C	1.71	1.72	43.12*	21.90*
A.X.	*M*	6.33	7.69	9.60	42.65*
	A	.56	.28	1.65	.28
	F	.02	.003	.003	.003
	C	6.44	7.32	16.72*	40.45*
Degrees of freedom	*M*	6	6	6	11
	A	1	1	1	1
	F	1	1	1	1
	C	3	3	3	5

[a] $M = A, E, F, H, X$; $A = A, E, X$; $F = F, H, X$; $C = A, E, F, H$.

* $p < .05$.

in separate fits of these models to the data across subjects and matrices (Chandler's, 1959, χ^2 STEPIT program was used to obtain the model fits reported in this study). The bottom row in Table 7 gives the degrees of freedom for the four models across the four matrices.

The results presented in Table 7 indicate that the WSCM and NCC produced much better fits to the data in the master matrix and A, E, F, H subset matrix than did the OVLP and AON. However, there was an equality of model fits in both the A, E, X and F, H, X subset matrices for the WSCM, OVLP, and AON. The low χ^2 goodness of fit values indicate that the models were fitting these data nearly perfectly. It has been shown that the WSCM, OVLP, and AON will produce equivalent fits of 2×2 confusion matrices (i.e., all three of these models are multiplicative with respect to 2×2 confusion data; see Falmagne, 1972). This is not the case for 3×3 confusion data. It is also known that the AON and OVLP are special cases of the WSCM (see Sections II and III). It appears that the 3×3 confusion data obtained in this study did not fall in that part of the WSCM confusion data that would violate the AON or OVLP.[12] There is thus a range of 3×3 confusion data that the WSCM, OVLP, and AON fit equally well, as indicated by the results in Table 7, although it has not been determined just what this range is. Examples of 3×3 confusion data where the WSCM, OVLP, and AON are not equivalent are given in Appendix B.

The pattern of the χ^2 goodness of fit values for the AON and OVLP reported in Table 7 for the master matrix (M rows) and the A, E, F, H subset matrix (C rows) verifies earlier findings. Previous studies (Townsend, 1971a, 1971b; Townsend & Ashby, Note 2) have reported the same general pattern of the AON fitting poorly, the WSCM fitting quite well, and the OVLP fitting no better than the WSCM (depending on experimental circumstances). What is more surprising is the very good fits obtained with the specialized NCC, which has not heretofore been tested against alphabetic confusion data. In the light of this fact, the NCC will be examined more closely in a comparison with the various interpretations of the SCM that are being tested in this study.

The general version of the NCC (Definition 7) has already been compared with the general version of the SCM (i.e., the WSCM) in Table 7. Although the NCC does not imply the CRR, its ω (sensory confusion) and δ (response bias) parameters can be systematically constrained to obtain NCC parallels to the SSCM, SSCM(η), and SSCM(β). The NCC versions will be referred to by SNNC, SNCC(ω), and SNCC(δ), respectively (i.e., the free parameter appearing in the parentheses), with the general unconstrained version now referred to by WNCC.

[12] There was one exception to the 3×3 confusion data equivalence in the A, E, X subset matrix data for the subject $A.X.$ in the OVLP. This is indicative that there are some 3×3 confusion data in which the OVLP, at least, does not fit as well as the WSCM and AON. See Appendix B. Additionally, we calculated several parameter values using the parameter implication and equivalence mappings given by Townsend (1978) and Pachella *et al.* (1978) and found no violations for the 3×3 data. That is, the parameter estimates returned by STEPIT for the WSCM, OVLP, and AON fit the equivalence equations perfectly for the 3×3 data.

TABLE 8

χ^2 Goodness of Fit Test between SNCC Model Predicted Matrices and Obtained Matrices

	Matrix		
Subject	*A, E, X*	*F, H, X*	*A, E, F, H*
A.X.	61.60*	36.78*	22.70*
G.X.	27.87*	49.31*	15.96
M.X.	16.43*	63.27*	14.47
D.X.	35.41*	37.84*	50.72*

* $p < .05$.

The SNCC (i.e., the SSCM version) χ^2 goodness of fits are presented in Table 8. As can be seen, the SNCC model did not predict the A, E, X and F, H, X subset matrices with a high degree of precision, but did somewhat better in predicting the A, E, F, H subset marix. This is in general agreement with the results of the SSCM and CRR reported in Table 5. Nevetheless, it did not perform quite as well overall as the SSCM.

As was done with the SSCM(η) and SSCM(β), the SNCC(ω) and SNCC(δ) can be examined to see which measure, ω or δ, is the most sensitive to constraint. This analysis indicated that the SNCC(ω) did considerably better than the SNCC(δ) in predicting the three subset matrices (4 significant χ^2 fit values, 1 at the .01 level, and 9 significant χ^2 fit values, 8 at the .01 value, respectively, over a total of 12 χ^2 fits; cf. Table 8). The nested χ^2 differences were calculated for the SNCC, SNCC(ω), and SNCC(δ) [as was done for the SSCM, SSCM(η), and SSCM(β)], and the same pattern of results was obtained. [Out of a total of 12 χ^2 differences, for the SNCC–SNCC(ω) comparison there were 9 significant χ^2 differences, all at the .01 level; for the SNCC–SNCC(δ) comparison there were 5 significant χ^2 differences, 4 at the .01 level, and 4 cases where the SNCC(δ) did *worse* than the SNCC; for the SNCC(ω)–NCC comparison there were 3 significant χ^2 differences, 1 at the .01 level; and for the SNCC(δ)–NCC comparison there were 8 significant χ^2 differences, all at the .01 level.] Moreover, the sensory parameters (ω) in the SNCC increased 73% of the time (averaged across each subject, 12 comparisons per subject) between the master and the subset matrices, as η did in the WSCM analysis, denoting again a presumed increase in the relative similarity between pairs of letters (if it is assumed that ω is indicative of similarity as is η). The bias parameters, δ, also showed no systematic variations. Thus, the qualitative aspects here are precisely the same as those with the choice model analysis, if not a little stronger.

Because the overall fits of the AON and OVLP were poor, it did not seem worthwhile to analyze their measurement constraints in any detail. However, a casual inspection of Tables 1–4 suggests widespread violations of the column constraint and the ratio constraints (types 1 and 2) in the master matrix and A, E, F, H subset matrix for each subject.

Because the WNCC and WSCM predicted the four matrices quite well in this study, from Theorems 2 and 4 and Holman (1979), it should follow that the conditions of bias substitutability, bias transitivity, diagonal maximization, and quasisymmetry should not be grossly violated in the obtained data in Tables 1–4. Table 9 presents the percentages of the tests passed across each matrix and subject. Quasisymmetry was not included in Table 9. We found quasisymmetry to be difficult to test in any meaningful manner. This was partially due to the very small values it generates (by multiplying proportions), and partially due to the strict equality it requires. The bias substitutability test was not applicable to the A, E, X and F, H, X subset matrices as it requires a minimum of four stimuli. In most cases, a strong majority of the non-parametric tests held. Yet, the percentages presented in Table 9 indicate that the confusion data need not necessarily follow all of the non-parametric

TABLE 9

Measurement Tests of the Obtained Matrices; the Percentage of the Tests That Were Passed

	Matrix					
	A, E, F, H, X			A, E, F, H		
Subject	BT[a]	BS	DM	BT	BS	DM
D.X.	100	32	80	25	32	75
M.X.	57	40	80	100	68	75
G.X.	57	88	80	25	68	75
A.X.	61	67	80	100	100	75
No. tests	60	120	25	24	24	16

	Matrix					
	A, E, X			F, H, X		
Subject	BT	BS	DM	BT	BS	DM
D.X.	100		100	100		100
M.X.	100		100	100		100
G.X.	100		100	0		100
A.X.	100		100	100		100
No. tests	6	*NA*	9	6	*NA*	9

[a] BT = bias transitivity (Holman, 1979),
BS = bias substitutability (Holman, 1979),
DM = diagonal maximization.

requirements of these tests while being predicted reasonably well by the WNCC and WSCM. Of course, counting the numbers of violations is a rough measure which does not take into account the magnitude of the violation. That is, for our purposes, any magnitude greater than 0 was sufficient for counting a violation as having occurred. The problem is that no statistical interpretation of the number of violations of tests such as these yet exists. However, with a magnitude measure there is the equally difficult problem of determining just what magnitude of violation is excessive. At the least, more comparisons between qualitative measurement conditions and parametric fits, as in the present investigation, may help to achieve a more adequate understanding of the relationship between model representations and data structure.

V. Indications for Models of Visual Confusion

Inspection of the parameters of the WSCM and WNCC indicated that an increase of similarity parameter values occurred for the WSCM and WNCC in moving from the master matrix to the subset matrices, with some minor variations in the bias parameters on a more or less random basis. These results are in line with the major experimental manipulation being restricted to manipulations of the stimulus set rather than to motivational and learning variables.

Since the primary variables under consideration in the present study were set size and context (i.e., the specific stimuli presented in each block), reasons for the changes in the sensory parameter estimates necessary for the WSCM and WNCC to obtain the best fits across the four matrix conditions require consideration. It is important to note first that a simple limited capacity interpretation does not work. On the basis of limited capacity notions alone, the η_{ij}'s and ω_{ij}'s should decrease as set size decreases, indicating a sharpening of discrimination, in *opposition* to the observed results.

Consider first the WNCC. Recall that upon presentation of stimulus i, the processing mechanism enters some n-way confusion state $\mathscr{C}_k$ with probability π_{ik}, where each probability π_{ik} is determined by a product of the probability estimates ω_{ij} (Eq. (11)). A response is then selected from the available responses in the confusion state according to the ratio of the biases for each available response. Accordingly, the variation in ω_{ij} values might be accounted for by interpreting the ω measure in terms of Nakatani's (1972) original rationale. The increase in ω_{ij} as the number of stimuli decreases would then indicate that the subject may be relaxing his acceptance boundaries, or, in other words, increasing the acceptance regions for the possible responses in the subset. As there are fewer stimuli to confuse, then the subject could relax the acceptance boundaries because, even if all of the possible responses become acceptable and a guess is made, he/she will have a greater likelihood of getting a hit.

Next, although the SSCM demands invariance in the η parameter values, it may be that a processing intepretation of the WSCM would not be incompatible with a variation of η. Although there is as yet no uniformly accepted processing interpretation for the WSCM, some work has been done in this area. One of the more attractive candidates is the maximal match choice model (Townsend, Note 6; Townsend, Evans, & Hu, Note 7; van Santen & Bamber, 1981, independently arrived at this model). The model assumes that the observer selects that alternative pattern which exhibits the maximal comparison, or match, with the stimulus pattern. The match process is assumed to be random, utilizing the idea of Holman and Marley (cited in Luce & Suppes, 1965), and later Yellott (1977), showing how the double exponential distribution can lead to the ratio of strengths form (Eq. (4)).

First, the assumptions:

1. Let $\mathbf{X}_{ij}$ be a random variable expressing the matching of the input representation of stimulus i with the representation in memory of stimulus j. Assume that the output of the match is $\log \mathbf{X}_{ij}$.

2. Let β_j be a constant expressing the bias toward response j, and suppose that the output of the bias process is $\log \beta_j$.

3. Assume that the bias output and match output are added as shown in Fig. 3.

4. Assume that $\mathbf{Y}_{ij} = \log \mathbf{X}_{ij}$ is distributed as the double exponential with scale value $\log \eta_{ij}$. Then $\mathbf{Y}_{ij} + \log \beta_j$ is distributed as

$$\begin{aligned} P(\mathbf{Y}_{ij} + \log \beta_j \leqslant y_{ij} + \log \beta_j) &= F(y_{ij} + \log \beta_j) \\ &= \operatorname{Exp}[-\operatorname{Exp}(-y_{ij} + \log \eta_{ij} + \log \beta_j)], \end{aligned}$$

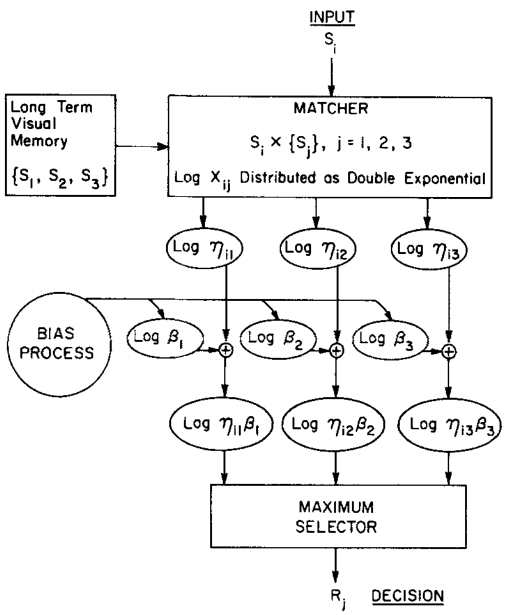

FIG. 3. Schema for the poosed SCM processing system for $N = 3$. S_i represents stimulus i and R_j represents response j. Note $\log \eta_{ij}$ ($j = 1, 2, 3$) represents the central parameter of the double exponential distribution, not the random output itself. β_j ($j = 1, 2, 3$) may be considered the mean of a bias distribution or simply a constant.

that is, as a double exponential with scale value $\log \eta_{ij} + \log \beta_j$. Figure 3 exhibits a flow diagram of this model.

We are now in a position to show that these combined assumptions result in the proper choice expression. The distribution on $\mathbf{Y}_{ij} = \log \mathbf{X}_{ij} + \log \beta_j$ may be expressed as

$$F(y_{ij} + \log \beta_j) = \mathrm{Exp}[-\mathrm{Exp}(-\log(x_{ij}/\eta_{ij}\beta_j))] = \mathrm{Exp}(-\eta_{ij}\beta_j/x_{ij}),$$
$$-\infty < y_{ij} < +\infty, \qquad 0 < x_{ij} < +\infty.$$

Now let $\mathbf{Z}_{ij} = 1/\mathbf{X}_{ij}$, so that

$$\begin{aligned} F(y_{ij} + \log \beta_j) &= F(\log x_{ij} + \log \beta_j) = F(\log z_{ij}^{-1} + \log \beta_j) \\ &= \mathrm{Exp}(-\eta_{ij}\beta_j z_{ij})|_{+\infty}^{z_{ij}} = \mathrm{Exp}(-\eta_{ij}\beta_j z_{ij}), \\ & \qquad 0 < z_{ij} < +\infty. \end{aligned}$$

This obviously results in (assuming a total of q stimuli and presentation of stimulus i):

$$\begin{aligned} P(\mathbf{X}_{ij} = \max_k \mathbf{X}_{ik}) &= P(\mathbf{Z}_{ij} = \min_k \mathbf{Z}_{ik}) \\ &= \int_0^{\infty} \eta_{ij}\beta_j \,\mathrm{Exp}(-\eta_{ij}\beta_j z) \prod_{\substack{k=1 \\ \neq j}}^{q} \mathrm{Exp}(-\eta_{ik}\beta_k z)\, dz \\ &= \eta_{ij}\beta_j \Big/ \sum_{k=1}^{q} \eta_{ik}\beta_k \end{aligned}$$

which is the formula sought.[13]

Within the context of the present study, the maximal match model would need to account for the increase in η_{ij} values as set size decreases. This could be accomplished by a shift in the location of the double exponential distribution of the $\log \mathbf{X}_{ij}$ towards $\log \mathbf{X}_{ii}$ (which has scale value 0 from the constraint that $\eta_{ii} = 1$).

One way in which this might occur is if the observer activates a filter or "template" for each potential pattern to be shown. It might take an amount of processing capacity to energize and maintain such filters. With a smaller alphabet, the observer could "loosen up" a little and permit a less finely tuned filter for each memory pattern. On the average, this would lead to some degradation in the pairwise matching process but performance would still improve overall due to the smaller number of patterns.

It is interesting that the proposed interpretations for the WSCM as well as the WNCC both rely on mechanisms that may be motivational yet affect what have traditionally been sensory parameters. It will take further experimental work to sort

[13] It is interesting that the WSCM has no reasonable representation along a continuum as it does here if, instead of permitting a maximal selector, it is supposed that a set of decision criteria separates the choice alternatives, except when $N = 2$ (Marley, 1971).

out these effects, not to mention more tests between the WSCM and the WNCC, both of which did a creditable job with the data obtained in the present study.

SUMMARY

The constant-ratio rule (CRR; Clarke, 1957) and four interpretatons of the similarity choice model (Luce, 1963) were tested in an across-matrix alphabetic confusion paradigm. The four matrix conditions varied with set size (three, four, or five stimulus elements), and set constituency (A, E, X; F, H, X; A, E, F, H; A, E, F, H, X). The strongest interpretation of the similarity choice model (SSCM) assumed an invariance of parameter estimates across the four matrix conditions, and was a special case of the CRR. The weak interpretation of the similarity choice model (WSCM) did not assume an invariance of parameter estimates across the four matrix conditions, and represented the commonly accepted interpretation of the similarity choice model (e.g., Pachella *et al.*, 1978; Townsend, 1971a). Additionally, the all-or-none (AON; Townsend, 1978), overlap (OVLP; Townsend, 1971a), and Nakatani (1972) confusion-choice (NCC) perceptual models were also examined in a comparison with the strong and weak versions of the similarity choice model.

The major conclusions of the study were as follows:

(1) The CRR and SSCM model both provided good first approximations in predicting the three subset matrices (A, E, X; F, H, X; A, E, F, H) from the master matrix (A, E, F, H, H). However, stricter statistical tests showed that, if a high degree of prediction precision is required, both the SSCM and the CRR are inadequate. The lack of a high degree of predictive precision of the CRR and SSCM appeared to be due to the varying set contexts across the four matrix conditions. An examination of two weaker versions of the SSCM model indicated that the failure of the model to predict accurately the data from the three subset matrices used in this study was due primarily to the restrictions placed on its η_{ij} similarity parameter estimates. The SSCM appears to be the preferable instrument as an overall first approximation because predictions were at least as good as the CRR and the SSCM is a stronger model. Further, the parameters of the SSCM permit the data to be broken down into bias and sensory confusion measures. However, the SSCM did not produce nearly as good a fit to the four matrix conditions as did the WSCM, even after the fact that the WSCM has fewer degrees of freedom was accounted for.

(2) Of the four perceptual models investigated in this study (AON; OVLP; NCC; WSCM), the WSCM and NCC produced by far the best fits to the data, and were essentially equal in their predictive power. The NCC results were found to closely parallel the interpretations of the WSCM. Psychological interpretations were offered for the sensory parameter changes that occurred in both models between the master matrix (A, E, F, H, X) and the three subset matrices. The results of this study indicate that the specialized version of the NCC analyzed in this study will have to be given more attention in future studies of visual perception, and tested further against the WSCM.

We conclude that although CRR (or SSCM) may be useful for general rough predictive purposes, it fails as an underlying principle of visual confusion; in particular as a probable generating seed behind the ratio of strengths form of WSCM. It might be mentioned in this context, that the SSCM or CRR predictions are as good or better than many predictions that are represented in the experimental literature as supporting a theory or model but that are not statistically tested.

Appendix A

Glossary of Abbreviations

AON	All-or-none model. See Definition 5.
CRR	Constant-ratio rule. See Definition 1.
OVLP	Overlap model. See Definition 6.
NCC	Nakatani's (1972) confusion-choice model, specialized version. See Definition 7.
WNCC	Weak Nakatani confusion-choice model. The ω and δ parameters are unconstrained.
SNCC	Strong Nakatani confusion-choice model. The NCC version of the SSCM. Both ω and δ parameters are constrained to not vary between matrices.
SNCC(δ)	Strong Nakatani confusion-choice model where the δ parameters are free and the ω parameters are constrained to not vary between matrices. The NCC version of the SSCM(β).
SNCC(ω)	Strong Nakatani confusion-choice model where the ω parameters are free and the δ parameters are constrained to not vary between matrices. The NCC version of the SSCM(η).
SCM	General similarity choice model. See Definition 2.
WSCM	Weak similarity choice model. See Definition 3(b).
SSCM	Strong similarity choice model. Both the η and β parameters are constrained to not vary between matrices. See Definition 3(a).
SSCM(β)	Strong similarity choice model where the β parameters are free and the η parameters are constrained to not vary between matrices. See Definition 4(a).
SSCM(η)	Strong similarity choice model where the η parameters are free and the β parameters are constrained to not vary between matrices. See Definition 4(b).

Appendix B

Although the range of the confusion space that would produce WSCM, AON, and OVLP equality in the prediction of 3×3 confusion data could not be determined, it is a rather easier task to show that there do exist 3×3 confusion data where an

TABLE A

A 3×3 Confusion Matrix Fit Well by the WSCM, but Not by the OVLP or AON

Stimulus	Response		
	i	j	k
i	.45	.08	.47
j	.06	.90	.04
k	.46	.05	.49

equality would not be produced. Theoretical results (e.g., Pachella *et al.*, 1978; Townsend, 1978), along with Section III, indicate that the WSCM, AON, and OVLP are related as follows: (1) the AON and OVLP are special cases of the WSCM; (2) the AON and OVLP are not special cases of each other; and (3) there is a range of confusion data that is fit well by all three models. Using this information, four types of confusion matrices can be examined: (1) where the WSCM fits well, but the AON and OVLP do not; (2) where the WSCM and OVLP fit well, but the AON does not; (3) where the WSCM and AON fit well, but the OVLP does not; and (4) where all three models fit well. Although these four types can be represented in any finite $q \times q$ confusion space, only 3×3 confusion matrix examples of the four types will be presented here. In fact, where the models "fit well" they fit perfectly, except for rounding error because they were used to generate the "data."

A confusion matrix representative of type (1) is obtained by letting $\eta_{ij} = \eta_{ji} = .11$, $\eta_{ik} = \eta_{ki} = .99$, $\eta_{jk} = \eta_{kj} = .06$, $\beta_i = .27$, $\beta_j = .44$, and $\beta_k = .29$ in the WSCM. Table A presents the (rounded off) confusion matrix produced by these parameter values. The matrix in Table A violates the column constraint (Definition 8(g)) necessary for the OVLP (cf. Theorem 2), and at least one instance of the modified diagonal maximization constraint (Definition 8(e)) necessary for the AON (Schulze, Note 5). Specifically, for the modified diagonal maximization constraint, $.45 \not> (.06 \times .47)/.04$.

The following 3×3 confusion matrix would be representative of type (2). Letting, for the WSCM, $\eta_{ij} = \eta_{ji} = .11$, $\eta_{ik} = \eta_{ki} = .64$, $\eta_{jk} = \eta_{kj} = .04$, $\beta_i = .28$, $\beta_j = .41$, $\beta_k = .31$, and letting, for the OVLP, $\xi_{ij} = \xi_{ji} = .16$, $\xi_{ik} = \xi_{ki} = .74$, $\xi_{jk} = \xi_{kj} = .06$,

TABLE B

A 3×3 Confusion Matrix Fit Well by the WSCM and OVLP, but Not by the AON

Stimulus	Response		
	i	j	k
i	.53	.09	.38
j	.07	.90	.03
k	.36	.03	.61

$g_i = .30$, $g_j = .36$, and $g_k = .34$, the (rounded) confusion matrix in Table B would be produced. The matrix in Table B violates at least one instance of the modified diagonal maximization constraint where $.53 \not> (.07 \times .38)/.03$.

A 3×3 confusion matrix representative of type (3) would be one where all of its cell entries are $\frac{1}{3}$. The parameter values for the AON would be $p_i = 0$, $h_j = \frac{1}{3}$ for all i, j and for the WSCM would be $\eta_{ij} = \eta_{ii} = 1$ and $\beta_j = \frac{1}{3}$ for all i, j. This matrix is trivial (and probably uninteresting), but nevertheless violates the column constraint necessary for the OVLP model.

A 3×3 confusion matrix representative of type (4) would be the obtained F, H, X matrix of subject $A.X.$ (see Table 4). This matrix was fit by all three models with a χ^2 value of .003 (see Table 7). The reason for the small but positive χ^2 value is most likely due to computational errors in the STEPIT computer estimation procedure (e.g., rounding errors).

References

BECKER, G. M., DEGROOT, M. H., & MARSCHAK, J. Probabilities of choices among very similar objects. *Behavioral Science*, 1963, **8**, 306–311.

BROADBENT, D. E. Word frequency effect and response bias. *Psychological Review*, 1967, **74**, 1–15.

CAUSSINUS, H. Contribution à l'analyse statistique des tableaux de corrélation. *Annales de la Faculté des Sciences de l'Université de Toulouse Pour les Sciences Mathématiques et les Sciences Physique*, 1966, **29**, 77–182.

CHANDLER, J. P. STEPIT; Finds local minima of a smooth function of several parameters. *Behavioral Science*, 1959, **14**, 81–82.

CLARKE, F. R. Constant-ratio rule for confusion matrices in speech communication. *Journal of the Acoustical Society of America*, 1957, **29**, 715–720.

CLARKE, F. R. Proportion of correct responses as a function of the number of stimulus alternatives. *Journal of the Acoustical Society of America*, 1959, **31**, 835.

CLARKE, F. R., & ANDERSON, C. D. Further tests of the constant-ratio rule in speech communication. *Journal of the Acoustical Society of America*, 1957, **29**, 1318–1320.

DEBREU, G. Review of R. D. Luce, Individual choice behavior: A theoretical analysis. *American Economic Review*, 1960, **50**, 186–188.

EGAN, J. P. Monitoring task in speech communication. *Journal of the Acoustical Society of America*, 1957, **29**, 482–489.

FALMAGNE, J. C. Biscalability of error matrices and all-or-none reaction time theories. *Journal of Mathematical Psychology*, 1972, **9**, 206–224.

GEYER, L. H., & DEWALD, C. G. Feature lists and confusion matrices. *Perception and Psychophysics*, 1973, **14**, 471–482.

HILTON, G. An algorithm for detecting differences between transition probability matrices. *Applied Statistics*, 1971, **20**, 80–85.

HODGE, M. H. Some further tests of the constant-ratio rule. *Perception and Psychophysics*, 1967, **2**, 429–437.

HODGE, M. H., & POLLACK, I. Confusion matrix analysis of single and multidimensional auditory displays. *Journal of Experimental Psychology*, 1962, **63**, 129–142.

HOLMAN, E. W. Monotonic models for asymmetric proximities. *Journal of Mathematical Psychology*, 1979, **20**, 1–20.

KRUMHANSL, C. L. Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review*, 1978, **85**, 445–463.

LUCE, R. D. *Individual choice behavior*. New York: Wiley, 1959.

LUCE, R. D. Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1). New York: Wiley, 1963.

LUCE, R. D., & SUPPES, P. Preference, utility and subjective probability. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 3). New York: Wiley, 1965.

LUPKER, S. J. On the nature of perceptual information during letter perception. *Perception and Psychophysics*, 1979, **25**, 303–312.

MARLEY, A. A. J. Conditions for the representation of absolute judgment and pair comparison isosensitivity curves by cumulative distributions. *Journal of Mathematical Psychology*, 1971, **8**, 554–590.

MORGAN, B. J. T. On Luce's choice axiom. *Journal of Mathematical Psychology*, 1974, **11**, 107–123.

NAKATANI, L. H. Confusion-choice model for multidimensional psychophysics. *Journal of Mathematical Psychology*, 1972, **9**, 104–127.

PACHELLA, R. G., SMITH, J. E. K., & STANOVICH, K. E. Qualitative error analysis and speeded classification. In N. J. Castellan & F. Restle (Eds.), *Cognitive theory* (Vol. 3). Hillsdale, N. J.: Erlbaum, 1978.

POLLACK, I., & DECKER, L. Consonant confusions and the constant-ratio rule. *Language and Speech*, 1960, **3**, 1–6.

VAN SANTEN, J. P. H., & BAMBER, D. Finite and infinite state confusion models. *Journal of Mathematical Psychology*, 1981, **24**(2), 101–111.

SHIPLEY, E. F., &LUCE, R. D. Discrimination among two- and three-element sets of weights. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford, Calif.: Standard Univ. Press, 1964.

SMITH, J. E. K. Models of identification. In R. S. Nickerson (Ed.), *Attention and performance VIII*. Hillsdale, N. J.: Erlbaum, 1980.

TOWNSEND, J. T. Theoretical analysis of an alphabetic confusion matrix. *Perception and Psychophysics*, 1971, **9**, 40–50. (a)

TOWNSEND, J. T. Alphabetic confusion: A test of models for individuals. *Perception and Psychophysics*, 1971, **9**, 449–454. (b)

TOWNSEND, J. T. A clarification of some current multiplicative confusion models. *Journal of Mathematical Psychology*, 1978, **18**, 25–38.

TOWNSEND, J. T., HU, G. G., & ASHBY, F. G. Perceptual sampling of orthogonal straight line features. *Psychological Research*, 1981, **43**(3), 259–276.

TOWNSEND, J. T., & LANDON, D. E. Mathematical models of recognition and confusion in psychology. *International Journal of Mathematical Social Sciences*, 1983, **1**(4), in press.

TVERSKY, A. Elimination by aspects: A theory of choice. *Psychological Review*, 1972, **79**, 281–299.

WILKS, S. S. The large sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics*, 1938, **9**, 60–62.

YELLOTT, J. I. The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution. *Journal of Mathematical Psychology*, 1977, **15**, 109–144.

REFERENCE NOTES

1. GIBSON, E. J., OSSER, H., SHIFF, W., & SMITH, J. An analysis of critical features of letters, tested by a confusion matrix. In *A basic research program on reading*. Cooperative Research Project No. 639, U. S. Office of Education, 1963.
2. TOWNSEND, J. T., & ASHBY, F. G. *Toward a theory of letter recognition: Testing of contemporary feature processing models*. Paper presented at the Midwestern Psychological Association Conference, Chicago, 1976. Incorporated in: An experimental test of contemporary mathematical models of letter recognition. *Journal of Experimental Psychology, Human Perception and Performance*, 1982, in press.

3. ANDERSON, C. D. *The constant-ratio rule as a predictor of confusions among various stimuli of brief exposure duration* (Technical Note AFCRC-TN-58-60). Hearing and Communications Laboratory, Indiana University, 1959.
4. EGAN, J. P. *Message repetition, operating characteristics, and confusion matrices in speech communication* (Technical Report 57-50). Hearing and Communications Laboratory, Indiana University, 1957.
5. SCHULZE, H. H. *Die Theoretische Analyse von Verwechslungsmatrizen in der Wahrnehmungspsychologie.* Unpublished manuscript, 1978.
6. TOWNSEND, J. T. *Mathematical models of visual character recognition.* Paper presented at the Fourth Annual Interdisciplinary Conference, Jackson Hole, Wyo., January 1979.
7. TOWNSEND, J. T., EVANS, R., & HU, G. *Some models in feature processing.* Paper presented at the Midwestern Psychological Association Conference, Chicago, 1979.

RECEIVED: August 12, 1981