Psychological Review
2004, Vol. 111, No. 1, 80–93

0033-295X/04/$12.00 DOI: 10.1037/0033-295X.111.1.80

Modeling the Effects of Choice-Set Size on the Processing of Letters and Words

Jeffrey N. Rouder
University of Missouri—Columbia

Letters and words are better identified when there are fewer available choices. How do readers use choice-set restrictions? By analyzing new experimental data and previously reported data, the author shows that Bayes theorem-based models overestimate readers' use of choice-set restrictions. This result is discordant with choice-similarity models such as R. D. Luce's (1963a) similarity choice model, G. Keren and S. Baggen's (1981) letter recognition model, and D. W. Massaro and G. C. Oden's (1979) fuzzy logical model of perception. Other models posit that choice restrictions affect accuracy only by improving guessing (e.g., J. L. McClelland & D. E. Rumelhart's, 1981, interactive activation model). It is shown that these models underestimate readers' use of choice-set restrictions. Restriction of choice set does improve perception of letters and words, but not optimally. Decision models that may be able to explain this phenomenon are discussed.

The psychological processes underlying letter and word identification have intrigued psychologists for over a century (e.g., Cattell, 1886). One of the most studied phenomenon in letter and word recognition concerns the role of context. Context can readily provide constraints on the number of plausible choices in an identification problem. For example, the word fragment *so_a* provides a great amount of constraint on the possible completions of the missing letter (it can be only *d* or *f* in English). In this article, I address the central question of how well people can condition their perception and subsequent decisions about letters and words when the number of available choices is manipulated. People are notoriously poor at conditioning their decisions on the available choices in macroscopic decisions such as disease diagnosis or job candidate evaluation (e.g., Phillips & Edwards, 1966; Tversky & Kahneman, 1990). It, however, is not known to what degree they can condition automatic processes, such as letter and word identification, on available choices. As will be shown, people are not as adept at conditioning as would be expected from the law of conditional probability. They, however, are better than they would be in the worst-case scenario of using choice-set restrictions only when guessing. Because people's ability is in between these two extremes, this ability is termed *somewhat-efficient conditioning*. Most extant models of letter and word recognition are challenged by somewhat-efficient conditioning.

Figure 1 shows an example of proper or ideal conditioning. Suppose a game show contestant is asked a question about a local beer in a small Midwestern city. The contestant is unsure of the correct answer but is able to assign probability values to the four choices as indicated in Figure 1. The game show host then eliminates Choices A and C. The law of conditional probability provides an ideal means by which contestants should revise their beliefs. Accordingly, probabilities are normalized by the available choices. In the example, the probabilities of the remaining choices, B and D, sum to .20; hence, the probabilities are normalized by this amount. These beliefs can be conceptualized as performance indices—that is, if contestants have a high belief in a correct answer, they are likely to answer correctly. The correct answer in this case is B. Performance in the four-choice condition is quite low (.19) but is much improved (.95) in the two-choice case. Proper conditioning is useful as an ideal-observer model. It provides a means of relating performance across conditions with different numbers of choices.

The goal is to assess people's ability to condition; to do so, I asked participants to identify letters with varying numbers of choices. Under normal viewing conditions, people are quite good at reading letters, and performance is at or near ceiling. But when the stimulus is presented under impoverished viewing conditions, either as a low-level illuminant or followed by a patterned mask, a substantial number of confusions occur. In this case, comparing identification under conditions with different numbers of choices provides a suitable test. For example, Rouder (2001) presented letters in either a six-choice condition or a two-choice condition. In the six-choice condition, participants identified the letters *Q*, *W*, *E*, *R*, *T*, and *Y*; in the two-choice condition, participants identified the letters *W* and *E*. The critical question concerns the change in performance in identifying *W* and *E* in these two conditions.

Formal models of identification often specify bottom-up perceptual effects, top-down bias effects, and an explicit mechanism of incorporating choice-set size. One approach to evaluate the veracity of these models is to stress parameter invariances. In each model, there are parameters that represent the bottom-up, perceptual effect of the presented letter. Choice-set-size manipulations are top-down in nature. Estimates of bottom-up perceptual param-

This research was supported by National Science Foundation Grant BCS-9917561 and Missouri Research Board Grant 00-77. The frequency matrices from the analyses may be obtained from the author. I am indebted to Tonya Cottrell for running the reported experiment and to James Townsend for several influential discussions. I thank Jonathan W. King, Paul Speckman, Nelson Cowan, and William Batchelder for helpful comments.

Correspondence concerning this article should be addressed to Jeffrey N. Rouder, Department of Psychological Sciences, University of Missouri—Columbia, 210 McAlester Hall, Columbia, MO 65211. E-mail: jeff@banta.psyc.missouri.edu

Q. Which of the following is brewed and served at The Flatbranch Brewery in Columbia, Missouri?

FOUR CHOICES

A) Ant Eater Lager	B) Oil Change Stout
C) Tiger Ale	D) Chancellor's Reserve

BELIEFS

A) .20	B) .19
C) .60	D) .01

TWO CHOICES

B) Oil Change Stout

D) Chancellor's Reserve

REVISED BELIEFS

B) .95

D) .05

Figure 1. An example of the law of conditional probability.

eters should not vary significantly across different choice-set-size conditions. In this article, I show that most models, including Massaro and Oden's fuzzy logical model of perception (FLMP; Massaro & Hary, 1986; Massaro & Oden, 1979; Oden, 1979) and McClellend and Rumelhart's (1981) interactive activation model (IAM), fail to predict perceptual parameter invariance with manipulation of choice-set size. In these models, perceptual parameters change systematically with the number of available choices, indicating flaws in the postulated decision mechanisms.

Formal models of letter recognition tend to embed one of two basic decision models: a threshold model or a choice-similarity model (see Townsend & Landon, 1983). Threshold models assume that processes occur in an all-or-none fashion and encompass models such as the high-threshold model, the double high-threshold model (Egan, 1975), and the low-threshold model (Luce, 1963b). Choice-similarity models (Luce, 1959, 1963a; Shepard, 1957) assume that identification processes are graded and driven by the similarity between the stimulus and the available responses. More substantively motivated models, such as those from Keren and Baggen (1981), Loomis (1982), Lupker (1979), Massaro and Oden (1979), and McClelland and Rumelhart (1981), incorporate one of these two general decision model classes. Hence, broad tests of the decision models are tests of these substantive models as well. Before discussing the models and tests, I present two benchmark data sets.

Benchmark Data Sets

In the following sections, formal models are fit to two different sets of data from letter-identification tasks. The first set is from Rouder (2001), who asked participants to identify briefly presented and subsequently masked letters. The task was absolute identification—a single letter was presented on a trial and each letter had a corresponding unique response. The conditions were previewed above; Rouder presented letters *Q*, *W*, *E*, *R*, *T*, and *Y* in the six-choice condition and letters *W* and *E* in the two-choice condition. The choice set was manipulated across blocks (a block lasted for 50 trials). Participants observed 16 blocks (8 for each condition) in a single session. Fifteen people participated in the experiment.

The second set comes from Townsend and Landon (1982). Their participants also identified letters in an absolute identification task. The letters, which were composed of line segments, were also briefly presented and subsequently masked. In the five-choice condition, participants observed the letters *A*, *E*, *F*, *H*, and *X*. There were two different three-choice conditions; each choice set in a three-choice condition was a subset of the five-choice condition choice set. Each participant observed stimuli for 16 sessions. Choice-set conditions were held constant for an entire session. Four people participated in Townsend and Landon's experiment.

Choice-Similarity Models of Letter Identification

Constant Ratio Rule

Choice-similarity models are, for the purposes of this article, those models that rely on normalization to describe how choice behavior depends on the set of available choices. Normalization refers to the fact that the activation or evidence for a response is divided by the activation or evidence for all available responses. The first choice-similarity model is the constant ratio rule (Clarke, 1957). A model is a constant ratio rule model if P_{r_i,s_j}, the probability of the *i*th response to the *j*th stimulus, can be expressed as

$$P_{r_i,s_j} = \frac{S_{i,j}}{\sum_k S_{k,j}}, \tag{1}$$

where the sum in the denominator is over all available choices and the *S*s are nonnegative and do not vary with choice set. The law of conditional probability, demonstrated in Figure 1, is a constant ratio rule. Equation 1 yields the correct conditioned probabilities when the *S*s are the relevant marginal probabilities. The model displays a testable property for which it is named—the ratios of response probabilities are constant across different choice-set sizes. For example, in Figure 1, the ratio between the probabilities for Choices B and D in the four-choice case is 19 (.19/.01). When the choices are reduced, the ratio is still preserved (e.g., .95/.05 = 19). Clarke (1957) tested and accepted the constant ratio rule with his data on phoneme confusions. More recently, Morgan (1974), Townsend and Landon (1982), and Takane and Shibayama (1992) have provided more stringent statistical tests of the constant ratio rule. On the basis of these tests, they rejected the constant ratio rule, but the reason for this failure is not clear.

To test the constant ratio rule with absolute identification data, one can simply divide the response proportions. For example, in Rouder's (2001) data, the critical ratio is the proportion of *W* responses divided by the proportion of *E* responses. There are two such ratios, one for when *W* is presented and one for when *E* is presented; these two ratios need not be equal. The question is whether these two ratios vary depending on the number of available choices. To assess this question, I plotted the logarithms of the ratios in the left panel of Figure 2. The points denoted by 1 are from the proportion of *W* responses divided by the proportion of *E*

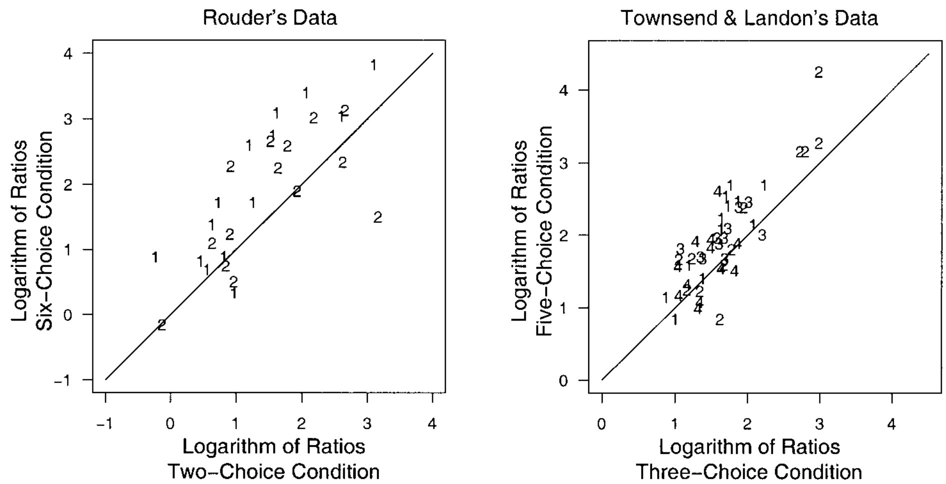

Figure 2. Test of the constant ratio rule. Left: Ratios from Rouder's (2001) data. Right: Ratios from Townsend and Landon's (1982) data. Ratios tended to be greater for the six- and five-choice conditions than for the two- and three-choice conditions, respectively, indicating worse performance with fewer numbers of choices than would have been expected under the constant ratio rule. In the left panel, the points denoted by 1 are from the proportion of *W* responses divided by the proportion of *E* responses when the letter *W* was presented, and the points denoted by 2 are the proportion of *E* responses divided by the proportion of *W* responses when the letter *E* was presented. In the right panel, the points denoted by numerals 1–4 represent each of the 4 participants in Townsend and Landon's experiment.

responses when the letter *W* was presented. The points denoted by 2 are the proportion of *E* responses divided by the proportion of *W* responses when the letter *E* was presented. With this convention, larger ratios are associated with better performance. The *x*-coordinate of the point is the logarithm of the ratio for the two-choice condition, and the *y*-coordinate is the logarithm of the same ratio for the six-choice condition. There are 14 such points,[1] 1 for each participant.

If the data obey the constant ratio rule, then the logarithms of the ratios should not vary with the number of choices, and the plotted points should cluster around the diagonal. However, most of these points (21 out of 28) are above the diagonal. Therefore, the constant ratio rule does not hold. Performance was worse in the two-choice condition than predicted by the constant ratio rule. The interpretation is that participants did not fully condition their identification on the reduced number of choices.

The same analysis was performed with Townsend and Landon's (1982) data. Altogether, there are six letter pairs (*AE*, *AX*, *EX*, *FH*, *FX*, and *HX*) with 2 ratios per pair. This arrangement resulted in 12 ratios per participant and 48 ratios over all 4 participants. The right panel of Figure 2 shows the logarithm of these ratios for the three-choice[2] and five-choice conditions. The points are denoted with numerals from 1 to 4 for each of the 4 participants. As can be seen, most of the points (39 out of 48) lie above the diagonal. This result indicates that participants did worse with fewer alternatives than would be expected from the constant ratio rule. The same conclusion is reached as with Rouder's (2001) data—participants were not fully efficient in conditioning their identification on the choice-set restriction.

The constant ratio rule serves as a decision mechanism in substantive models. One example is FLMP (Massaro, 1998; Massaro & Hary, 1986; Massaro & Oden, 1979; Oden, 1979). FLMP is a model of how different sources of information are combined to produce identification decisions. According to FLMP, participants assess the overall match—the degree to which critical features are present or absent in a target stimulus. Response probabilities are given by Equation 1, with overall match being denoted by *S*. The overall match represents perceptual processes and should be constant across choice-set-size manipulations. Hence, the model is formally equivalent to the constant ratio rule and is challenged by the present analysis of Rouder's (2001) and Townsend and Landon's (1982) data.

Another example of a substantive model that uses the constant ratio rule is Keren and Baggen's (1981) recognition model. This model, which is based on Tversky's (1977) set-theoretic approach to similarity, does not assume letter similarity is symmetric: For example, *F* might be more similar to *E* than *E* is to *F*. The probability of response is given by Equation 1, with similarity denoted by *S*. Because similarity is a perceptual parameter that should be invariant to the choice-set-size manipulations, Keren and Baggen's model is challenged by the present analysis.

It is important to put these challenges in context. FLMP is designed to explain how letters are represented and how multiple sources of information may be integrated in making letter decisions. The decision component is one facet of the model; the

[1] Fifteen people participated in Rouder's (2001) experiment. One is excluded because the ratio is undefined in the six-choice condition due to no error responses. This omission does not weaken any of the claims made subsequently. Had a single error response occurred, then the six-choice ratio would have been larger than the two-choice ratio for this participant.

[2] There were two different three-choice conditions: one with letters *A*, *E*, and *X* and another with letters *F*, *H*, and *X*.

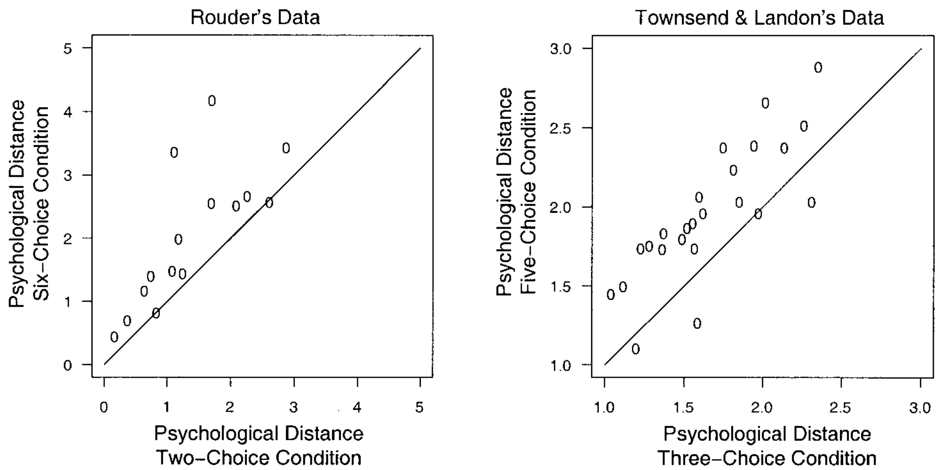

Figure 3. Similarity choice model (SCM) psychological distance estimates. Left: Distances from Rouder's (2001) data. Right: Distances from Townsend and Landon's (1982) data. Distances tended to be greater for the six- and five-choice conditions than for the two- and three-choice conditions, respectively, indicating worse performance with fewer numbers of choices than would have been expected under SCM. The circles denote the distances from each participant on a specific letter pair. Adapted from "Absolute Identification With Simple and Complex Stimuli," by J. N. Rouder, 2001, *Psychological Science, 12*, p. 320. Copyright 2001 by Blackwell. Adapted with permission.

current analysis is a test of this facet and is not about the rules for representing letters or integrating multiple sources of information. Likewise, Keren and Baggen's (1981) model explains how similarity can be asymmetric. The negation of the decision component is not a negation of this explanation of asymmetry.

Similarity Choice Model

As pointed out by Luce (1959, 1963a), the constant ratio rule may fail for fairly mundane reasons. In particular, there may be idiosyncratic response biases. To account for these biases, Luce's (1963a) similarity choice model (SCM) uses a weakened form of the constant ratio rule. In SCM, the probability of response r_i to stimulus s_j is given by

$$P_{r_i,s_j} = \frac{\eta_{i,j}\beta_i}{\sum_k \eta_{k,j}\beta_k}. \tag{2}$$

In Equation 2, $\eta_{i,j}$ is the similarity between stimulus s_i and stimulus s_j ($0 \leq \eta_{i,j} \leq 1$). Similarity is assumed to reflect the perceptibility of letters. Letters that are more similar are more confusable. The β parameters reflect response biases ($0 \leq \beta_i \leq 1$, $\sum_j \beta_j = 1$). To make the model identifiable and testable, it is assumed that similarity is symmetric (e.g., $\eta_{i,j} = \eta_{j,i}$). If it is further assumed that the similarity of any stimulus to itself is 1 and that the triangle inequality condition ($\eta_{i,j} \leq \eta_{i,k}\, \eta_{k,j}$, $\forall i, j, k$) holds, then it is possible to interpret $d_{i,j} = -\log(\eta_{i,j})$ as the psychological distance between stimuli i and j (Shepard, 1957). If the distance between two stimuli is great, then the letters are not easily confused. Psychological distance is a measure of the perceptual discriminability, and its interpretation is similar to that of the d' statistic in the theory of signal detection.

SCM is considered the leading model of letter identification because it repeatedly fits identification data better than any competitor model (e.g., Smith, 1992; Townsend & Landon, 1982). SCM also passes an invariance test with regard to response bias. Townsend and Ashby (1982) manipulated the payoffs for various responses and found that the perceptual parameters, $d_{i,j}$, were fairly invariant.

The question is whether the perceptual parameters, $d_{i,j}$, change with choice-set size. They should not. Although invariance of perceptual parameters is a critical test, no such invariance is sought with response bias. Response bias may vary across choice-set size for mundane reasons; for example, a participant may favor the left-most response key and this key may vary for different choice sets. Figure 3 shows the distances from Rouder's (2001) and Townsend and Landon's (1982) data.[3] In the left panel, the x-coordinate of each point is the distance between E and W in the two-choice condition; the y-coordinate is the same distance in the six-choice condition. There are 15 points—one for each participant. As can be seen, 14 of the 15 points lie above the diagonal, indicating a greater distance between E and W in the six-choice condition than in the two-choice condition. The right panel shows distances between letter pairs from Townsend and Landon's (1982) data. The x-coordinate of each point is the distance between two letters (the pairs were *AE*, *AX*, *EX*, *FH*, *FX*, and *HX*) in a

[3] Parameter estimation is done by minimizing the chi-square goodness-of-fit statistic with repeated application of the simplex algorithm (Nelder & Mead, 1965). The simplex algorithm is a local minimization routine and may settle on local minima. To help find a global minimum, an iterative form of simplex was constructed. The best-fitting parameters from one cycle were perturbed and then used as the starting value for the next cycle. Cycling continued until 500 such cycles resulted in the same minimum. The amount of perturbation of parameter values between cycles increased each time the routine yielded the same minimum of the previous cycle.

three-choice condition; the *y*-coordinate is the same distance in the five-choice condition. The points correspond to distances from each participant's performance on a specific pair of letters. Once again, most of the points (20 out of 24) lie above the diagonal, indicating greater distance between letters in the five-choice condition than in the three-choice condition. Participants did worse with fewer choices than predicted by SCM.

Chi-square goodness-of-fit test statistics were computed for each of Rouder's (2001) participants in the six-choice conditions (there are no degrees of freedom for a test in the two-choice condition) and for each of Townsend and Landon's (1982) participants in the five- and three-choice conditions. Of the resulting 27 tests, none of the associated chi-square statistics exceed its respective .05 criterion. This good fit is characteristic of the previous literature. SCM fits well when only a single condition is fit, but it fails when parameter invariance is sought across choice-set size manipulation. This failure effectively negates any support from good fits to individual confusion matrices.

The failings of the constant ratio rule and SCM most likely stem from their shared structural assumption that response activation is normalized by the activation of all available choices. Normalization is too efficient; it predicts better performance with fewer choices than what was observed. The interpretation offered here, inefficient conditioning, is itself relative. The results can also be restated in terms of capacity. Given the performance in the two- and three-choice conditions, performance is better than predicted in the six- and five-choice conditions, respectively. Performance increases with an increase in the number of stimuli—that is, it is "super capacity." The direction of the structural failure shown here can be interpreted either as super capacity or inefficient conditioning. The latter is retained with the caveat that both are mutually compatible.

Threshold Models

All-or-None Model

In this section, threshold models are presented and fit. For the purposes of this article, threshold models posit that identification is mediated by discrete states. The effect of manipulating choice-set size is to change accuracy when the participant is in a guessing state. Such a limited use of choice-set information produces inefficient conditioning. The simplest threshold model is the high-threshold one. For absolute identification, it is assumed that the participant either detects the stimulus (in which case they respond correctly) or guesses. When guessing, participants choose responses at random with no influence of the presented stimulus. Townsend (1971a, 1971b) termed this model the all-or-none model and showed that it provides a fairly good account of his letter recognition data with few parameters. The all-or-none model is given by

$$P_{r_i,s_j} = \begin{cases} D_j + (1 - D_j)g_i, & i = j. \\ g_i, & i \neq j. \end{cases} \tag{3}$$

Parameter D_j denotes the probability that the participant detects the stimulus s_j when it is presented ($0 \leq D_j \leq 1$). Parameter g_i denotes the probability that the participant produces response r_i when guessing ($0 \leq g_i \leq 1$, $\Sigma_i\, g_i = 1$). In this model, the perceptual effects of the stimuli are represented in the detection parameters. Therefore, these parameters should be invariant to choice-set-size manipulations.

The left panel of Figure 4 shows the detection parameters for Rouder's (2001) data; the right panel shows the same for Townsend and Landon's (1982) data. The *x*-axis of each point denotes the detection parameter in the condition with a small number of choices (two choices for Rouder's set; three choices for

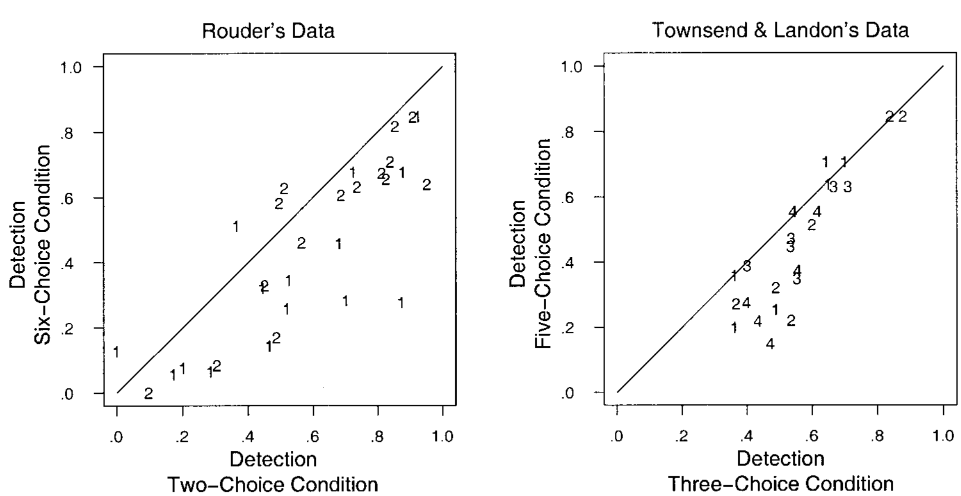

Figure 4. Test of the all-or-none model. Left: Detection estimates from Rouder's (2001) data. Right: Detection estimates from Townsend and Landon's (1982) data. Detection tended to be greater for the two- and three-choice conditions than for the six- and five-choice conditions, respectively, indicating better perception with fewer choices. In the left panel, points denoted 1 and 2 are each participant's estimates of the *W* and *E* detection parameters, respectively. In the right panel, the label of the points corresponds to the participant number.

Townsend & Landon's set). The *y*-axis of each point denotes the detection parameters in the condition with a larger number of choices. For Rouder's data, there are 30 points. Points denoted 1 and 2 are each participant's estimates of the *W* and *E* detection parameters,[4] respectively. As can be seen, most of these points (26 out of 30) lie below the diagonal, indicating better detection with fewer choices. For Townsend and Landon's data, the label of the points corresponds to the participant number. Once again, the majority of the points (20 out of 24) lie below the diagonal indicating better detection with fewer choices. Chi-square goodness-of-fit tests were performed on both data sets. The all-or-none model fit well for Rouder's data and could not be rejected for any of the 15 participants. But, it failed at the .05 level for all 4 participants in Townsend and Landon's five-choice condition: $\chi^2(11, N \approx 1{,}200) = 42.0$, 76.1, 28.8, and 80.5, respectively. I discuss a possible reason for the difference in fit below. Regardless of fit, the model fails invariance in the same direction for both data sets, and this failure is a serious challenge to the all-or-none model.

The direction of the failure is opposite to that obtained with choice-similarity models. According to the choice-similarity models, perception is worse than predicted with fewer choices; however, according to the all-or-none model, perception is better than predicted with fewer choices. The choice-similarity models assume that people are very efficient at conditioning perception on a choice-set restriction, whereas the threshold models assume that people are quite poor at it—they use the choice-set reduction only in the guessing state and not in perception. The analyses reveal that people are intermediary in their ability to condition on a choice-set restriction. They are not as efficient as predicted by normalization but are more efficient than they would be by simply restricting guesses to available choices. Therefore, people are somewhat efficient in letter identification.

General-Threshold Model

The all-or-none model of the previous section is a high-threshold model. Yet, researchers often postulate low-threshold processes on models as well (Luce, 1963b). Low-threshold processes correspond to perceiving a letter that was not presented. The plausibility of such processes is enhanced in letter-identification tasks with patterned masks. Sometimes the mask is suggestive of a letter (for example, the mask $ is suggestive of the letter *S*). Participants may detect the letter as any one in a set or fail to detect it at all. For example, if the letter *E* is presented, there is some probability that the participant will detect it as an *E*. There is also some probability that the participant will detect it as an *F* or an *X*. Let $D_{i,j}$ denote the probability of detecting stimulus *j* as *i*. In this case, the response is r_i. Failure to detect the stimulus is possible, and in this case, the participant guesses. When guessing, the participant produces response r_i with probability g_i. This model[5] is termed the *general-threshold model* (GTM) and is expressed as

$$P_{r_i,s_j} = D_{i,j} + g_i\left(1 - \sum_k D_{k,j}\right). \tag{4}$$

In GTM, the perceptual parameters are the detection parameters, $D_{i,j}$. Therefore, the invariance of these parameters is assessed.

Unfortunately, there are more parameters in the model than degrees of freedom in the data. To make the model identifiable, restrictions are placed on $D_{i,j}$. There are a number of options. For simplicity, symmetry was assumed ($D_{i,j} = D_{j,i}$) and was used in fitting GTM. Figure 5 shows the correct-item detection parameter estimates.[6] The left panel of Figure 5 shows the estimates for Rouder's (2001) data (the points labeled 1 and 2 denote correct-item detection for *W* and *E*, respectively). Detection was better with fewer choices. The model yielded a mediocre fit (it failed at the .05 level for 6 of 15 participants). Wrong-item detection estimates were always near zero, indicating no evidence for false detection. For Rouder's data, GTM offers no advantage over the all-or-none model.

The right panel of Figure 5 shows the same analysis for Townsend and Landon's (1982) data. For the five-choice condition, the detection symmetry assumption, $D_{i,j} = D_{j,i}$, is sufficient and yields a testable model.[7] GTM fits for each participant were quite acceptable and each of the corresponding chi-square test statistics was below its .05 criterion. This good fit is in contrast to the poor fit for the all-or-none model. For Townsend and Landon's data, GTM has nonzero wrong-item detection. Although there is no evidence for false detection in Rouder's data, there is such evidence for Townsend and Landon's. The difference is probably due to differences in font and masks across these studies. Nonetheless, the same conclusion can be reached regarding parameter invariance—detection was better with fewer choices than with more choices.

---

[4] The model has $2n - 1$ parameters, and the data have $n(n - 1)$ degrees of freedom, where n is the number of choices. A necessary condition for identifiability is that the degrees of freedom in the data be equal to or greater than the number of model parameters. For the case in which there are two choices, there are three parameters in the model but only two degrees of freedom in the data. Hence, the model is not identifiable in this case. To overcome these difficulties, I conducted two different analyses. In the first analysis, the detection of *E* in the two-choice case was fixed and set equal to that from the six-choice case. This left the detection of *W* as a free parameter and its invariance to be tested. In the second analysis, the detection of *W* in the two-choice case was fixed and set equal to the detection of *W* in the six-choice case. This left the detection of *E* as a free parameter and its invariance to be tested. In both analyses, it is assumed that one of the detection parameters was invariance across the choice-set-size manipulation. If the free detection parameter varies across choice condition (within error), then the detection-parameter invariance can be rejected.

[5] I am grateful to William Batchelder, who suggested I consider this model.

[6] Fitting the general-threshold model to the two-choice condition is complicated as there are four parameters for only two degrees of freedom. Consequently, some of the parameters in the two-choice condition were fixed and set equal to their estimates in the six-choice condition. In one analysis, detection parameters $D_{W,W}$ and $D_{E,W}$ were set equal to their six-choice estimates, and the detection parameter $D_{E,E}$ was a free parameter. Alternatively, in the other analysis, detection parameters $D_{E,E}$ and $D_{E,W}$ were set equal to their six-choice estimates, and the detection parameter $D_{W,W}$ was a free parameter. In both cases, response bias was free to vary across conditions.

[7] To estimate detection in the three-choice conditions, the wrong-item detection parameters ($D_{i,j}$, $i \neq j$) were fixed and set equal to their estimates in the five-choice condition.

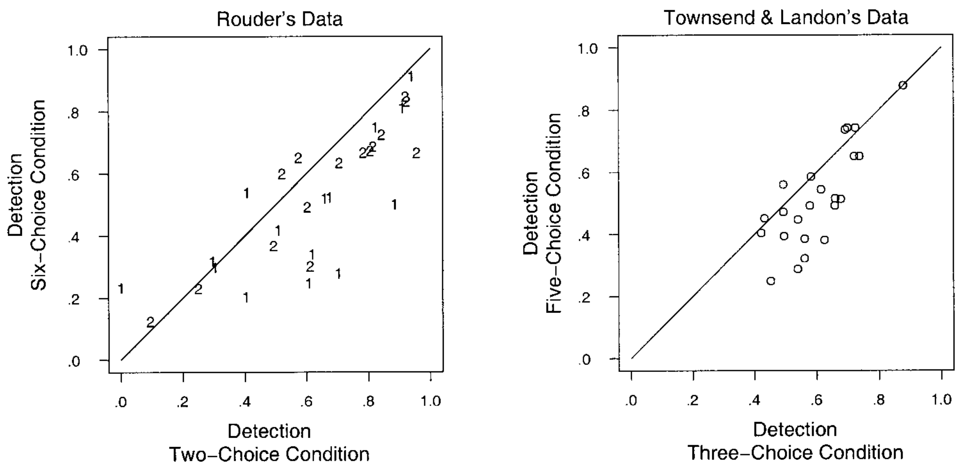

Figure 5. Test of the general-threshold model. Left: Detection estimates from Rouder's (2001) data. Right: Detection estimates from Townsend and Landon's (1982) data. Detection tended to be greater for the two- and three-choice conditions than for the six- and five-choice conditions, respectively, indicating better perception with fewer choices. In the left panel, the points labeled 1 and 2 denote correct-item detection for *W* and *E*, respectively. In the right panel, circles denote correct-item detection for specific letters.

IAM

Perhaps the best known model of letter and word identification is McClelland and Rumelhart's (1981) IAM. In this model, letters and words are each represented by nodes. Activation flows from feature nodes to letter nodes to word nodes and then back to letter nodes. The model has been inordinately successful in accounting for word-recognition phenomena (see Rumelhart & McClelland, 1982), and its core assumptions are used in current word-recognition models, for example, Grainger and Jacob's (1996) multiple readout model.

The decision mechanism of IAM is a general-threshold model. To identify letters, readers monitor the letter nodes. The activation of a certain letter node is determined by both the presented stimulus sequence and the activations of other letter and word nodes. The dynamics of the model are fairly complex and vary depending on the viewing conditions. For the case of letter recognition, IAM computes a naming probability for each letter. The basic idea is that at a predetermined time, one of the letter nodes has highest activation and hence is named. In cases in which the participant is forced to choose from a reduced number of alternatives, either the node with the highest activation is a member of the set of choices or it is not. If it is a valid choice, then the response corresponding to the winning letter node is produced. If the winning node is not a choice, then the participant guesses at random. For example, suppose a participant is presented the letter *E* and given the choice set of *F*, *Q*, and *X*. On some proportion of the trials, the node corresponding to *F* is going to have the highest activation because it is similar to *E*. In these cases, the participant will choose equally often between all three alternatives even though *F* is more similar to *E*. This inefficiency is necessary for McClelland and Rumelhart (1981) to fit both naming and forced-choice identification paradigms. Uninformed or pure guessing is implemented in other models as well (e.g., Anderson & Lebiere, 1998; Wagenmakers, Zeelenberg, Schooler, & Raaijmakers, 2000).

To explore the behavior of IAM with respect to choice-set size, I implemented the feature and letter layers of the model in a C program. The features, letters, and the connection between features and letters were exactly the same as those used by McClelland and Rumelhart (1981; Figure 6 shows the feature-to-letter mappings used in the model; these are from Rumelhart & Siple, 1974). The word layer was not programmed as the inputs were single letters that could not activate word nodes. All of the parameters related to the model dynamics were very similar to those used by McClelland and Rumelhart. The only free parameter is the length of time (in cycles) that the stimulus is presented before masking. Figure 7 shows the model's accu-

Figure 6. Letters introduced by Rumelhart and Siple (1974) and used by McClelland and Rumelhart (1981). The bottom character depicts the segments used in forming the letters. Adapted from "An Interactive Activation Model of Context Effects in Letter Perception: I. An Account of Basic Findings," by J. L. McClelland and D. E. Rumelhart, 1981, *Psychological Review, 88,* p. 383. Copyright 1981 by the American Psychological Association. Adapted with permission of the author.

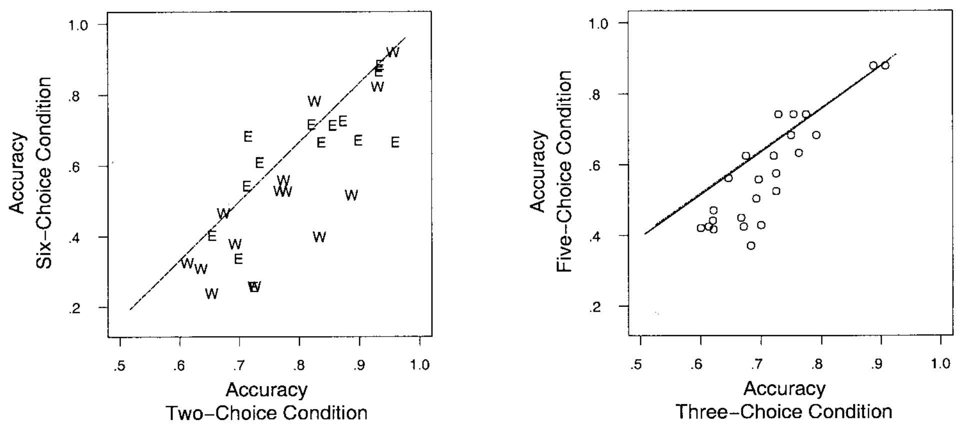

Figure 7. Response accuracy data and interactive activation model (IAM) predictions. The lines show the IAM accuracy predictions for different set sizes. The letters (E and W) are accuracies from Rouder's (2001) data (left), and the circles are accuracies from Townsend and Landon's (1982) data (right).

racy predictions for the case in which choice-set size is manipulated. The solid lines are IAM predictions. The *x*-axis value of each point on the line corresponds to the predicted probability correct for fewer choices (two choices for Rouder's, 2001, data; three choices for Townsend & Landon's, 1982); the *y*-axis value is the same for the condition with more choices (six choices for Rouder's, 2001, data; five choices for Townsend & Landon's, 1982). The trajectory of the line is traced out by varying the stimulus duration parameter. The advantage of this technique for comparing data with the predictions is that the free parameter (stimulus duration) need not be estimated for each participant. This simplified approach could not be used in fitting the previous models because there were several free parameters rather than a single one.

As can be seen, most of the points lie below the line, indicating that IAM predicts better performance in the condition with more choices than what was obtained. This result challenges the decision mechanism in IAM. As with the evaluation of previous substantive models, it is important to put this challenge in perspective. IAM explains how letters and words are identified and how information on one level affects that on another. The negation of the decision rule is not a challenge to these more central aims.

The preceding analyses reveal that people are somewhat efficient in using choice-set restrictions with performance intermediate to the two extremes. This intermediate level is fairly difficult to model. Before discussing possible approaches that can predict somewhat-efficient conditioning, I explore the scope of the phenomenon. The goal of the experiment below is to assess the efficiency of conditioning for word stimuli when the numbers of choices are varied. Participants named briefly presented and subsequently masked four-letter target words. Target words were presented in either a two-choice condition or a naming condition. There are about 1,200 four-letter words in the Kucera and Francis (1967) corpus with frequencies above 2 per million. Hence, the naming condition can be considered a choice from about 1,200 alternatives.

Experiment: Choice-Set Size and Word Stimuli

Method

Participants. Twelve University of Missouri undergraduate students served as participants in exchange for course credit in an introductory psychology class.

Materials. The stimuli were pairs of four-letter words. Members of a pair were matched for frequency (Kucera & Francis, 1967, word-frequency norms). Words of a pair may have had letters in common, but they never occurred in the same position in the words. For example, *bolt* and *echo* form a pair. They have the same Kucera–Francis frequency and they share the letter *o* in common. The letter *o*, however, occurs at different positions. Emotionally charged words, such as *rape*, as well as swear words were not used. With these words eliminated, there were still 1,159 four-letter words in the Kucera and Francis (1967) corpus.

Apparatus. Participants sat in front of a Pentium or Pentium II PC. Stimuli were presented on a Dell 17-in CRT with a custom-written set of C-language routines run under MS-DOS. The monitor was set to a 60-hz refresh rate.

Design. The experiment was a 3 $\times$ 2 within-subject design. The first factor is condition. There was a naming condition and two different two-choice conditions. In the prestimulus cue condition, the participant was presented the two choices before the target word was presented. In the poststimulus cue condition, the participant was presented the two choices after the target word was presented. The other factor was whether the first or second member of the pair served as the target. Choice condition (three levels) was crossed with target (two levels) and participants in a Latin-square design.

Procedure. The structure of trials is shown in Figure 8. Each trial started with the participant depressing the space bar. The sequence was shown and the participant attempted to name the target word. The experimenter sat behind the participant and recorded the response on a code sheet. The participant was permitted to say "no" if they perceived no information about the stimulus. But, participants were instructed to produce word responses as often as possible, even if they had low confidence in their answers. The experimenter recorded the response and then asked the participant to proceed to the next trial.

Before the test phase, participants partook in a 31-trial calibration phase to determine individualized stimulus durations. In the calibration phase, participants observed 6 practice trials and 31 calibration trials. Stimulus duration for the practice trials was 100 ms, and stimulus duration for the

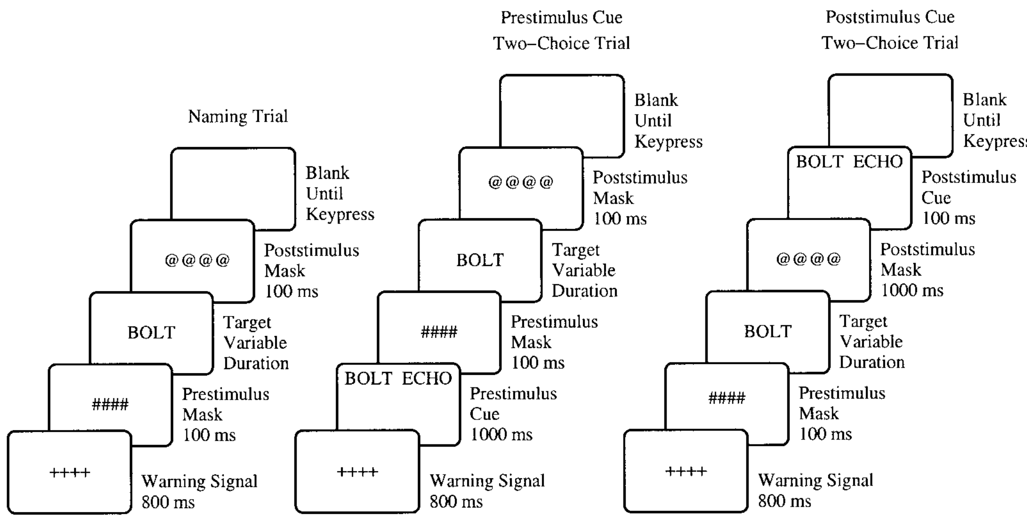

Figure 8. The structure of trials in the different choice conditions.

calibration trials was 50 ms. At the end of the calibration phase, the experimenter received the number of correct responses. If participants made fewer than 4 errors, then the stimulus duration was set to 33 ms; if they made between 4 and 10 errors, then stimulus duration was set to 50 ms; and if they made more than 10 errors, then the stimulus duration was set to 67 ms.

Immediately after calibration, the test phase began. Participants first observed 12 practice trials. Afterward, they observed the 294 trials that made up the critical list. The experiment was run at a leisurely pace; participants and experimenters took breaks whenever needed. The session lasted about 45 min.

Results

Target-naming accuracy served as the dependent variable. A strict criterion was used to decide if a reported word was correct. The left panel of Figure 9 shows the accuracy results in the two-choice condition as a function of those in the naming condition. The right panel shows the same data after a logit transformation (e.g., $l = \log[p/(1 - p)]$). The logit transform provides a different view of the data that stresses differences near ceiling.

All-or-none model fits. It is fairly straightforward to fit an all-or-none model given by the equation $P_N = D + (1 - D)/N$. The dashed line in Figure 9 shows the predictions for the correct two-choice naming probability ($N = 2$) as a function of the correct naming probability ($N = 1{,}159$). The line was made by systematically changing the value of detection, D. As can be seen, all points in the panels lie above this line, indicating that participants' relative efficiency is better than that predicted by the all-or-none model.

SCM model fits. Producing predictions for SCM is more complicated. Unlike the previous application in which full confusion

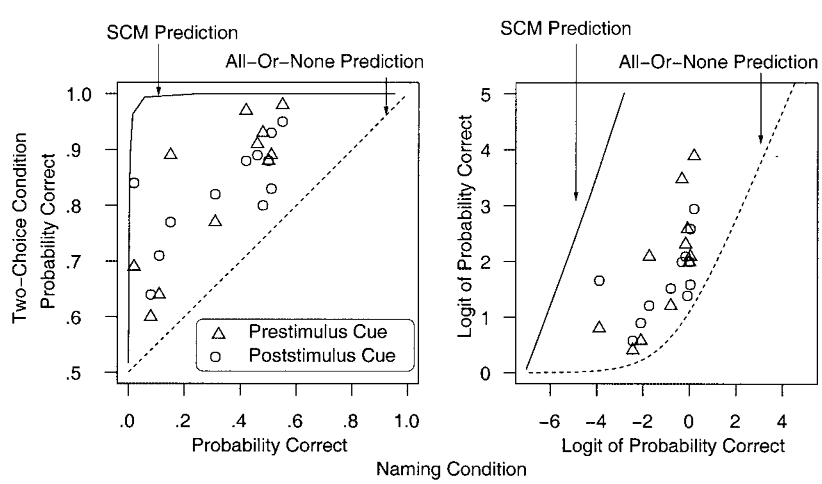

Figure 9. Results and model predictions for the experiment. Left: Accuracies in the two-choice condition as a function of those in the naming condition. Right: The same after a logit transformation. SCM = similarity choice model.

matrices were collected, the design of the current study provides only overall accuracies. The confusion matrices were essential in the preceding analyses because they provided a means of estimating the similarity between pairs of items. As an alternative, the similarity between words was calculated using the Rumelhart and Siple (1974) feature construction of letters (see Figure 6). Each letter is defined by 14 features and each four-letter word is defined by the 56 features that make up individual letters. The similarity between two words was defined as the proportion of these 56 features that matched and is denoted with $v_{i,j}$. With this technique, the font-based similarity between all pairs of four-letter words in the Kucera and Francis (1967) corpus was computed. To derive predictions, the following SCM model was used:

$$P_{r_i,s_j} = \frac{\nu_{i,j}^{\alpha}\beta_i}{\sum_k \nu_{k,j}^{\alpha}\beta_k}. \tag{5}$$

The exponent α reflects viewing condition such as the illuminant of the target or its duration before masking. As viewing condition improves, α increases. To derive predictions, values of α were varied. For each value of α, correct response probabilities for the two-choice and naming trials were computed. The solid lines in Figure 9 show these predictions. All of the data points lie below the line, indicating that the relative efficiency of the data is not as great as that predicted by SCM.

Discussion

The effects of manipulating choice-set size on word identification are similar to those on letter identification. Participants are somewhat efficient in conditioning on a choice-set restriction. They are not as efficient as normalization would imply but are more efficient than they would be by simply using the choice-set restriction during guessing.

Models of Somewhat-Efficient Conditioning

The main point of the preceding analyses is that neither threshold nor choice-similarity models do a good job of explaining how letter and word identification decisions change with choice set. Choice-similarity models condition too efficiently; threshold models are not efficient enough. These failures are fundamental and relate to the structural assumptions of the models. In the following section, a few modeling approaches are presented that hold the possibility of explaining the somewhat-efficient conditioning effect.

Simple Mixture Model

One tactic in accounting for the data is to make the choice-similarity models less efficient. Nosofsky (1991) assumed that behavior is the mixture of two states: one in which the stimulus is encoded and one in which it is not. In paradigms in which stimuli are masked, the possibility exists that participants may miss entirely the stimulus presentation on some proportion of the trials. Because the stimuli are not encoded on these trials, the similarity of the stimuli cannot affect the response. Responses on these trials reflect a simple guessing process. On the trials in which the stimulus is encoded, performance is governed by SCM; errors reflect confusions of the stimulus with similar stimuli. Overall, performance is a mixture between a simple guessing process and SCM. Nosofsky's mixture model can be expressed as

$$P_{r_i,s_j} = \frac{D\eta_{s_i,s_j}\beta_{r_i}}{\sum_k \eta_{s_k,s_j}\beta_{r_k}} + \frac{1-D}{N}, \tag{6}$$

where D is the probability that the participant encodes the stimulus. This is a proper generalization of SCM (SCM results if $D = 1$) with one additional parameter. To fit the mixture model, the parameter D was fixed across choice conditions. This is reasonable as the probability of encoding a stimulus should not depend on the number of choices. The similarities were free to vary across choice condition as were response biases. Rouder (2001) reported the fit of the mixture model to his data. The encoding parameter, D, was greater than .995 for 8 of the 15 participants. For the other 7, the estimate varied from .50 to .96. When D was near 1.0, the estimates of distance were very close to the SCM estimates. But even when D was not near 1.0, the distance estimates were greater in the six-choice case than in the two-choice case. Two conclusions may be reached: (a) In the context of the mixture model, most errors are driven by similarity rather than random guessing, and (b) when fitted to letter-identification data, the mixture model overpredicts the efficiency of conditioning.

The Nosofsky (1991) mixture model has not previously been fit to Townsend and Landon's (1982) data. To do so, I fit the different choice conditions jointly. Similarity and response bias were free to vary across conditions, but the encoding parameter, D, was not. The results were quite concordant with those from fitting the mixture model to Rouder's (2001) data. For 3 of the 4 participants, estimated values of D were greater than .995, and the distance estimates were close to the SCM estimates. For the remaining participant, the 6 five-choice distances were greater than their corresponding three-choice condition distances. Hence, the same conclusion is reached: When fit, the mixture model is too efficient to account for the data.

Trial-by-Trial Process Variability

An alternative method of making the SCM less efficient is to add trial-by-trial variability to the parameters (e.g., Van Zandt & Ratcliff, 1995). To add trial-by-trial variability, psychological distance was decomposed into two components: one that represents the font-based psychological distance between two letters and one that represents the effects of viewing condition. For example, in most fonts, the psychological distance between E and F is always smaller than that between E and X, and this holds regardless of viewing condition. The degradation or enhancement of viewing condition is assumed to affect all distances uniformly:

$$d_{i,j} = \alpha\delta_{i,j}, \tag{7}$$

where δ is the font-related distance and $\alpha \geq 0$ reflects the viewing condition. Large values of α correspond to good viewing conditions; small values of α correspond to poor viewing conditions. Incorporating the distances defined in Equation 7 into SCM yields Equation 5 with font-based similarity given by $v_{i,j} = \exp(-\delta_{i,j})$.

To generalize SCM to include trial-by-trial variability, I assumed that the parameter α varies from trial to trial. Variation in α reflects variation in attention. When participants pay attention, the value of α is large, and when they do not, the value is small. This generalization of SCM is termed the *variable similarity choice model* (vSCM). Response probabilities are given by

$$P_{r_i,s_j} = \mathrm{E}_\alpha\left[\frac{\nu^\alpha_{s_i,s_j}\beta_{r_i}}{\sum_k \nu^\alpha_{s_k,s_j}\beta_{r_k}}\right], \tag{8}$$

where E_α is the expectation operator with respect to α. If we assume that α is a continuous random variable with density function f, then the model can be expressed as

$$P_{r_i,s_j} = \int_0^\infty \left[\frac{\nu^\alpha_{s_i,s_j}\beta_{r_i}}{\sum_k \nu^\alpha_{s_k,s_j}\beta_{r_k}}\right] f(\alpha)d\alpha. \tag{9}$$

The variability of α determines the relative ability of the model to condition on a choice-set restriction. When variability is small, the model is able to condition efficiently on reduced numbers of choices. But when it is great, conditioning is inefficient and performance approaches that of threshold models. Figure 10 illustrates this property; it shows two-choice performance as a function of letter naming (26-choice performance). In this figure, font-based similarity, $v_{i,j}$, is the proportion of features that matched in the Rumelhart and Siple (1974) letters. Because α can never be negative, it is convenient to use the logarithm scale (log α). Variability in parameter α was implemented by distributing log α as a normal random variable. The three different curves in Figure 10 correspond to three different values of variance. The different points on the curves correspond to different values of the mean. As the mean of log α was increased, performance in both the two-choice and naming condition increased.[8] As the variability of the exponent increases, the relative performance in the two-choice condition becomes progressively worse. That is, conditioning becomes less efficient.

To assess whether process variability is feasible, I fit the model of Equation 9 to Rouder's (2001) data. For computational convenience in fitting the model to data, mass was placed on only two values of α:

$$f(\log\alpha) = \begin{cases} 1/2, & \text{if } \log\alpha = \gamma. \\ 1/2, & \text{if } \log\alpha = -\gamma. \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

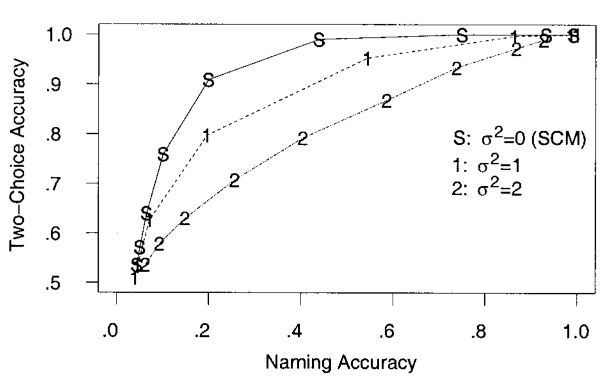

Figure 10. Effect of variability of α on efficiency in the variable similarity choice model (vSCM). As α becomes more variable, performance in the two-choice condition is degraded relative to that in the naming condition.

In this formulation, γ denotes the amount of variability in log α and determines the quality of conditioning. If $\gamma = 0$, then the model reduces to SCM. As γ increases, efficiency of conditioning is reduced and, in the limit, approaches that of threshold models.

The vSCM model was fit to Rouder's (2001) data in a nested approach. In the general model, γ was free to vary. In the nested model, γ was fixed to zero. In both cases, γ was held constant across the six- and two-choice conditions. There was only one parameter for the distance between W and E that was fixed to be the same across the six- and two-choice conditions. This last step is different from the previous modeling approach in which two separate distance parameters were used. The six- and two-choice cases, when fit jointly, yield 32 degrees of freedom in the data. There are 22 parameters in the general model yielding 10 degrees of freedom to assess the fit per participant. Over all 15 participants, the sum chi-square statistic is 145.9 (with 150 total degrees of freedom). This value is quite consistent with a well-fitting model. In the nested model ($\gamma = 0$), there are 21 parameters and 11 degrees of freedom to assess the fit per participant. Over all participants, the sum chi-square statistic is 177.6 (with 165 total degrees of freedom). Because the models are nested, the difference in chi-square serves as a test of whether $\gamma = 0$. This difference is distributed as a chi-square with 15 degrees of freedom (1 per participant). The observed difference, 31.8, is significant at the $p < .05$ level. The fact that this difference is significantly large implies that $\gamma > 0$. The conclusions are that (a) the vSCM model fits well and (b) variability is significantly larger than zero.

The preceding analysis demonstrates that vSCM can account for somewhat-efficient conditioning. But, vSCM makes a strong prediction about manipulations of stimulus duration and contrast. When stimulus duration or contrast is varied, only the parameter α can vary; similarity and bias must remain invariant. Although it is possible that these predictions may hold, there is some evidence against it. Townsend, Hu, and Kadlec (1988) noted that participants tend to weight certain features differently over the time course of perception. Their model follows a popular notion in letter perception that participants first identify low-frequency features (blobs) and, afterward, identify high-frequency features (e.g., Bouma, 1971; Lamb & Yund, 1993, 1996; Lupker, 1979; Navon, 1977). For example, X and O may be judged as relatively similar early in processing because they both have large extension. But, they may be judged relatively dissimilar later in processing as one has curvature whereas the other has an intersection point. Conversely, C and O may be judged dissimilar early on (O is a greater blob) but relatively similar later on in processing (both have curvature). Such dynamics are in contrast to the posited uniform changes in distance in vSCM. Townsend et al. documented nonuniform dynamics, but not for natural letters (instead, they used simple combinations of lines). It is an open question whether such dynamics occur in typical letter reading and identification.

General Recognition Theory Models

Many models in the literature assume that the perceptual effect of a presented stimulus can be represented in a multidimensional

[8] Naming performance is the mean over all 26 stimuli; two-choice performance is the mean over all 650 unique combination of a stimulus and an alternative.

psychological space. The effect is assumed to vary from trial to trial. To make a decision, the participant partitions the psychological space into contiguous regions with each region corresponding to a particular response. Because the percept is variable, there is some probability that the percept will fall in a region corresponding to an error response. If the distribution of the percept is a multivariate normal, then this approach reduces to the general recognition theory of Ashby and Townsend (1986); if the distribution is a univariate normal, then the approach further reduces to the theory of signal detection (Green & Swets, 1966/1974).

The main goal is to assess whether there are plausible models that can predict the somewhat-efficient conditioning. To assess efficiency, I have constructed the following statistic, called the *performance ratio*:

$$\xi = \sqrt{\frac{P_{i,i}P_{j,j}}{P_{i,j}P_{j,i}}}. \tag{11}$$

Overall, ξ is a measure of performance with larger values indicative of higher accuracy. To apply the performance ratio statistic to Rouder's (2001) data, I estimated ξ for the letters W and E in the two- and six-choice cases, respectively. If either the constant ratio rule or SCM holds, the value of ξ will be constant across choice conditions. The performance ratio is closely associated with the SCM model; it is the method-of-moments estimator for the inverse similarity ($1/\eta$; see Townsend, 1971a). The left panel of Figure 11 shows the performance-ratio relationship for Rouder's data (letter pairs W and E). The 15 Xs denote the empirical relationship for each of the 15 participants. The diagonal shows a constant ξ, the predicted relationship from SCM. All of the points, save one, are above this line, indicating that SCM overestimates participants' efficiency in using the choice-set restriction. The dashed line above the diagonal shows the predicted relationship from the all-or-none model with no guessing bias. This model underestimates participants' efficiency. The two solid lines, which show the

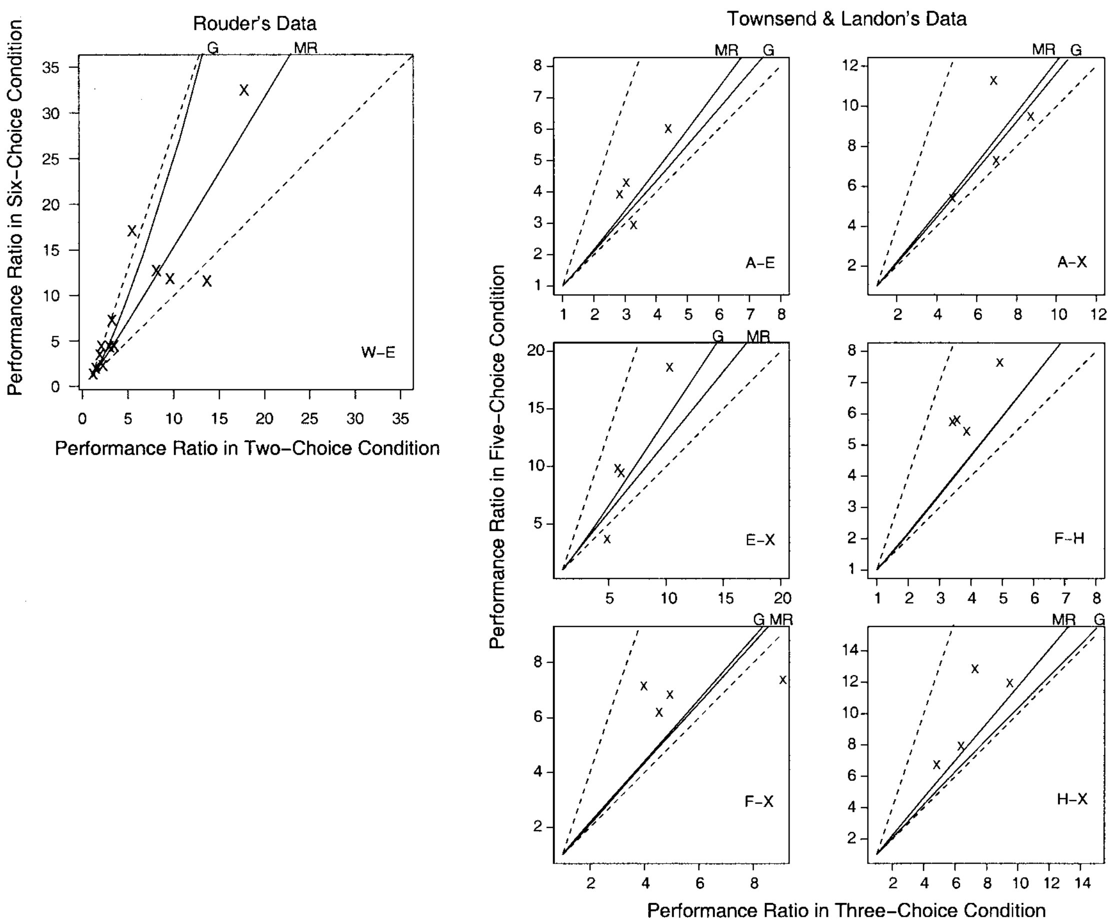

Figure 11. Efficiency of conditioning as shown through the performance ratio, ξ. In each panel, the x-axis indicates the value of ξ in the condition with fewer choices, and the y-axis indicates the value of ξ in the condition with more choices. Left: The results for Rouder's (2001) experiment. Right: The results for Townsend and Landon's (1982) experiment; each panel denotes a performance-ratio relationship for different letter pairs. Points denoted by Xs indicate empirical values, the dashed diagonals indicate predictions from the similarity choice model, the dashed lines above the diagonals indicate those from the all-or-none model, and the solid lines are predictions from general recognition models based on Gilmore et al.'s (1979) scaling (denoted with G) and on McClelland and Rumelhart's (1981) features (denoted with MR).

desired somewhat-efficient conditioning, are the predictions from the general recognition models discussed below. The six panels on the right show the same data and model predictions for Townsend and Landon's (1982) data set. Each panel denotes a performance-ratio relationship for different letter pairs. The 4 data points in each panel correspond to the 4 participants.

The first model fit is a version of general recognition theory in which a letter is represented by a multidimensional normal distribution. The results of Gilmore, Hersh, Caramazza, and Griffin (1979) are helpful in fitting the model. Gilmore et al. estimated the distance between pairs of letters using SCM. These distances were then submitted to a multidimensional scaling routine. The researchers claimed that five dimensions are sufficient for an acceptably low level of stress in the analysis and reported the values for all letters on these five dimensions. Therefore, five-dimensional normals were used to represent letters with each centered on its five-dimensional coordinates from Gilmore et al.'s multidimensional scaling. For simplicity, it was assumed that the variance in each dimension was constant and there was no covariance across dimensions. Monte-Carlo simulation was used to make predictions. On each cycle, a sample was drawn from the five-dimensional distribution of the target letter. Response on a cycle was the letter whose center was closest to that cycles' five-dimensional sample (see Ashby & Maddox, 1993). The process was repeated for a million cycles per stimulus. For parametric predictions, the variance of the normal was varied from .2 to 150. As the variance became larger, the performance ratio became smaller. The solid line labeled *G* in each panel in Figure 11shows the relationship of the performance ratio as a function of the two different choice conditions. This line is in between the two dashed lines; hence, the model predicts somewhat-efficient conditioning.

A second model, based on the Rumelhart and Siple (1974) letters, was also fit. Each letter was represented by 14 binary features that were either absent or present. When a letter was presented, sensory noise was instantiated by randomly flipping the value of some of the features. The activation for each letter response was the dot product of the noisy-input feature vector and the feature vector for the response. The response with the greatest activation was chosen on each cycle. For parametric predictions, the probability that a feature had its value flipped was varied from .00075 to .86. The solid line labeled *MR* in each panel in Figure 11 shows the relationship of the performance ratio as a function of the two different choice conditions. This line is in between the two dashed lines; hence, the model predicts a form of somewhat-efficient conditioning. Overall, two different general recognition type models can at least crudely account for the somewhat-efficient conditioning effect.

Conclusion

The main goals of this study were to assess the degree to which participants use choice-set restrictions in letter identification and to use this assessment to test decision mechanisms in identification models. The result is that participants are somewhat efficient in their conditioning on choice-set restrictions—they are neither as efficient as they would be by using ideal conditioning nor as inefficient as they would be by simply using choice-set restrictions when guessing. This intermediate result held for both letter and word identification.

The challenge to theorists is to explain somewhat-efficient conditioning. Most researchers have proposed letter-identification theories that relied either on normalizing activation by the total activation of available choices (e.g., Keren & Baggen, 1981; Massaro & Oden, 1979) or on guessing among available alternatives when the identification process fails to recognize the stimulus as being in the available set (e.g., McClelland & Rumelhart, 1981). The analyses reported here indicate that participants' efficiency is between these two extremes. Hence, both approaches are insufficient.

In the last section, two different modeling approaches were proposed to account for the somewhat-efficient conditioning result. One approach, based on Luce's (1963a) SCM, assumes that the similarity between any two stimuli varies from trial to trial. This modified model, vSCM, can account for the somewhat-efficient conditioning aspect of the data. It remains to be seen whether this model can account for other aspects of the data with appropriate parameter invariances such as changes in stimulus duration, payoffs, and font. The results of Townsend et al. (1988) provide evidence against such invariances, but their experiments were not with natural letters. A second modeling approach is to assume that the letter percept itself randomly varies from trial to trial as is assumed in the theory of signal detection. Again, under appropriate conditions, this approach too can yield somewhat-efficient conditioning.

Historically, decision-theoretic models, such as SCM and threshold models, have been preferred to more substantive models because they provide better fits. In this article, it is shown that these good fits do not extend to the one manipulation these models were designed to account for—the manipulation of choice-set size. Hence, there is really little reason to prefer them to more substantive models that posit specific features and feature integration processes (e.g., Townsend et al., 1988). In the end, these more substantive models, which make strong commitments to representational issues, may encompass a more fruitful approach. One area of future synthesis may be to implement general recognition theory (Ashby & Townsend, 1986) in which letter representation is derived in a principled manner from a substantive theory.

References

Anderson, J. R., & Lebiere, C. (1998). *The atomic component of thought.* Mahwah, NJ: Erlbaum.

Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology, 37,* 372–400.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93,* 154–179.

Bouma, H. (1971). Visual recognition of isolated lower-case letters. *Vision Research, 11,* 459–474.

Cattell, J. M. (1886). The time taken up by cerebral operations. *Mind, 11,* 277–292.

Clarke, F. R. (1957). Constant ratio rule for confusion matrices in speech communications. *Journal of the Acoustical Society of America, 29,* 715–720.

Egan, J. P. (1975). *Signal detection theory and ROC analysis.* New York: Academic Press.

Gilmore, G. C., Hersh, H., Caramazza, A., & Griffin, J. (1979). Multidimensional letter similarity derived from recognition errors. *Perception & Psychophysics, 25,* 425–431.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review, 103,* 518–565.

Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics.* New York: Krieger, Huntington. (Original work published 1966)

Keren, G., & Baggen, S. (1981). Recognition models of alphanumeric characters. *Perception & Psychophysics, 29,* 234–246.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Lamb, M. R., & Yund, E. W. (1993). The role of spatial frequency in the processing of hierarchically organized stimuli. *Perception & Psychophysics, 54,* 773–784.

Lamb, M. R., & Yund, E. W. (1996). Spatial frequency and attention: Effects of level-, target-, and location-repetition on the processing of global and local forms. *Perception & Psychophysics, 58,* 363–373.

Loomis, J. (1982). Analysis of tactile and visual confusion matrice. *Perception & Psychophysics, 31,* 41–52.

Luce, R. D. (1959). *Individual choice behavior.* New York: Wiley.

Luce, R. D. (1963a). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 103–189). New York: Wiley.

Luce, R. D. (1963b). A threshold theory for simple detection experiments. *Psychological Review, 70,* 61–79.

Lupker, S. J. (1979). On the nature of perceptual information during letter perception. *Perception & Psychophysics, 25,* 303–312.

Massaro, D. W. (1998). Models for reading letters and words. In D. Scarborough & S. Sternberg (Eds.), *Methods, models, and conceptual issues: An invitation to cognitive science* (Vol. 4, pp. 301–364). Cambridge, MA: MIT Press.

Massaro, D. W., & Hary, J. M. (1986). Addressing issues in letter recognition. *Psychological Research, 48,* 123–132.

Massaro, D. W., & Oden, G. C. (1979). Integration of featural information in speech perception. *Psychological Review, 85,* 172–191.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review, 88,* 375–407.

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

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognition, 9,* 353–383.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal, 7,* 308–313.

Nosofsky, R. M. (1991). Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology, 23,* 94–140.

Oden, G. C. (1979). A fuzzy logical model of letter identification. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 336–352.

Phillips, L. D., & Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology, 72,* 346–354.

Rouder, J. N. (2001). Absolute identification with simple and complex stimuli. *Psychological Science, 12,* 318–322.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89,* 60–94.

Rumelhart, D. E., & Siple, P. (1974). Process of recognizing tachistoscopically presented words. *Psychological Review, 81,* 99–118.

Shepard, R. N. (1957). Stimulus and response generation: A stochastic model relating generalization to distance in a psychological space. *Psychometrika, 22,* 325–345.

Smith, J. E. K. (1992). Alternative biased choice models. *Mathematical Social Sciences, 23,* 199–219.

Takane, Y., & Shibayama, T. (1992). Structures in stimulus identification data. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 335–362). Hillsdale, NJ: Erlbaum.

Townsend, J. T. (1971a). Alphabetic confusion: A test for models of individuals. *Perception & Psychophysics, 9,* 449–454.

Townsend, J. T. (1971b). Theoretical analysis of an alphabetic confusion matrix. *Perception & Psychophysics, 9,* 40–50.

Townsend, J. T., & Ashby, F. G. (1982). Experimental test of contemporary mathematical models of visual letter recognition. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 834–864.

Townsend, J. T., Hu, G. G., & Kadlec, H. (1988). Feature sensitivity, bias, and interdependencies as a function of energy and payoffs. *Perception & Psychophysics, 43,* 575–591.

Townsend, J. T., & Landon, D. E. (1982). An experimental and theoretical investigation of the constant-ratio rule and other models of visual letter confusion. *Journal of Mathematical Psychology, 25,* 119–162.

Townsend, J. T., & Landon, D. E. (1983). Mathematical models of recognition and confusion in psychology. *Mathematical Social Sciences, 4,* 25–71.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327–352.

Tversky, A., & Kahneman, D. (1990). Judgements under uncertainty: Heuristics and biases. In P. K. Moser (Ed.), *Rationality in action: Contemporary approaches* (pp. 171–188). New York: Cambridge University Press.

Van Zandt, T., & Ratcliff, R. (1995). Statistical mimicking of response time distribution: Mixtures and parameter variability. *Psychonomic Bulletin & Review, 2,* 20–54.

Wagenmakers, E., Zeelenberg, R., Schooler, L. J., & Raaijmakers, J. G. W. (2000). A criterion-shift model of enhanced discriminability in perceptual identification: A note of the counter model. *Psychonomic Bulletin & Review, 7,* 662–667.

Received July 8, 2002
Revision received February 13, 2003
Accepted March 6, 2003 ■