Psychological Assessment
2002, Vol. 14, No. 2, 184–201

1040-3590/02/$5.00 DOI: 10.1037//1040-3590.14.2.184

Cognitive Psychometrics: Assessing Storage and Retrieval Deficits in Special Populations With Multinomial Processing Tree Models

David M. Riefer
California State University, San Bernardino

Bethany R. Knapp and William H. Batchelder
University of California, Irvine

Donald Bamber
Space and Naval Warfare Systems Center

Victor Manifold
Veterans Affairs Medical Center, St. Cloud, Minnesota

This article demonstrates how multinomial processing tree models can be used as assessment tools to measure cognitive deficits in clinical populations. This is illustrated with a model developed by W. H. Batchelder and D. M. Riefer (1980) that separately measures storage and retrieval processes in memory. The validity of the model is tested in 2 experiments, which show that presentation rate affects the storage of items (Experiment 1) and part-list cuing hurts item retrieval (Experiment 2). Experiments 3 and 4 examine 2 clinical populations: schizophrenics and alcoholics with organic brain damage. The model reveals that each group exhibits deficits in storage and retrieval, with the retrieval deficits being stronger and occurring more consistently over trials. Also, the alcoholics with organic brain damage show no improvement in retrieval over trials, although their storage improves at the same rate as a control group.

One way that abnormal behavior manifests itself in clinical patients is that they exhibit deficits in cognitive functioning. There are several different approaches for assessing cognitive deficits, but Batchelder and Riefer (1999; Batchelder, 1998) have advocated an approach based on formal models of cognition. The first step in this approach is to identify the population suspected of having a particular pattern of deficits and to conduct experiments on this population using standard, well-studied cognitive tasks. The tasks that are selected should be ones for which corresponding mathematical models have been developed and successfully validated in experiments conducted on normal populations. Estimates of the parameters for these statistical models provide a methodology for the measurement of latent cognitive subskills. Parameter estimates for a special population can be compared with estimates from suitable controls or with established baselines. Changes in the parameters of individuals can be measured across time or under different circumstances. Significant differences in particular parameters enable one to pinpoint the exact nature of the cognitive deficits.

David M. Riefer, Department of Psychology, California State University, San Bernardino; Bethany R. Knapp and William H. Batchelder, Department of Cognitive Sciences, University of California, Irvine; Donald Bamber, Space and Naval Warfare Systems Center, San Diego, California; Victor Manifold, Veterans Affairs Medical Center, St. Cloud, Minnesota.

The research conducted for Experiments 3 and 4 was supported by Veterans Administration medical research funds. Details on the method for incorporating order constraints into multinomial processing tree models can be obtained by contacting William H. Batchelder, Department of Cognitive Sciences, University of California, Irvine, California 92697. E-mail: whbatche@uci.edu

Correspondence concerning this article should be addressed to David M. Riefer, Department of Psychology, California State University, San Bernardino, 5500 University Parkway, San Bernardino, California 92407-2397. E-mail: driefer@csusb.edu

Batchelder (1998; Batchelder & Riefer, 1999) refers to this methodology as *cognitive psychometrics*, and it has several advantages over the more traditional approach of comparing manifest performance scores on specially constructed clinical assessment batteries. First, the model parameters are designed to measure different cognitive subskills separately, whereas performance measures from test batteries are the result of an unknown composite of cognitive processes. For example, a failure to recall an item from a memory test could be due to a deficit in attention, encoding, storage, retrieval, or any combination of these. Second, manifest performance measures, such as number of words recalled on a memory test, are based on performance scores aggregated over all test items. These aggregate measures often fail to capture the full information provided by the data, such as different types of error responses or the trial-to-trial pattern of responses on individual items.

Most researchers in the field of cognitive psychology accept that it is futile to try to provide process-pure, operational measures of underlying cognitive subskills. It is this recognition that has led to the frequent use of model-driven analyses of cognitive data, such as signal-detection models (MacMillan & Creelman, 1991), process-dissociation models (Jacoby, 1998), and Markov models of learning and memory (Greeno & Bjork, 1973). A large class of models for this purpose consists of multinomial processing tree (MPT) models, reviewed by Batchelder and Riefer (1999). MPT models subsume many of the other types of models, and there is a clear-cut approach to inferential statistics for MPT models, such as evaluating goodness of fit, parameter estimation, and hypothesis testing.

It is important to contrast the strategy of assessing deficits with mathematical models with the more general strategy of theoretical model building in cognitive psychology. There is a great effort today to develop cutting-edge models of cognition that are capable of understanding detailed data patterns over a wide range of

experimental tasks in normal populations. For example, in the area of recognition memory, there are a variety of global-matching models (see Clark & Gronlund, 1996, for a review) and dual-process models (e.g., Yonelinas, 1999) that are still being revised and elaborated. Despite the relatively advanced state of theory development in cognitive psychology, these comprehensive models are complex, difficult to analyze, often theoretically controversial, and frequently give way to new and improved versions. Thus, they do not provide stable, reliable, or practical measurement tools for analyzing cognitive deficits in special populations.

For assessment purposes, relatively simple models are needed that are paradigm specific and well validated under standard testing conditions. Such models are too simple and specific to constitute basic theory, but they often approximate well-established theory when they are confined to a particular paradigm. Because the statistical theory for these models is relatively straightforward, they can easily be used as measurement tools to measure underlying cognitive deficits and test hypotheses about specific deficits between different clinical populations. In addition, models of this type can be tested for validity and reliability by repeated applications in experiments, in which independent variables are chosen to affect certain parameters (and not others) in psychologically meaningful ways (Batchelder & Riefer, 1999). If the model passes these sorts of validity tests, then it is reasonable to interpret the estimates of a particular parameter as validly tapping a particular cognitive subskill. Because of their statistical simplicity and versatility, MPT models are especially suited for supplementing other more standard methodologies in clinical assessment. For example, Knight and Silverstein (2001) have advocated a process-oriented approach to clinical research and have acknowledged that mathematical modeling can be useful in complementing their approach. Neufeld (1998, p. 308) has argued that process models provide an "expanded methodological arsenal" that clinicians can draw on for their assessment of cognitive deficits in special populations.

The purpose of this article is to provide a detailed example of the cognitive psychometrics approach advocated by Batchelder (1998). The example involves a well-established MPT model of category clustering called the *pair-clustering model* (Batchelder & Riefer, 1980, 1986, 1999). The model is designed to separately measure two cognitive subskills in episodic memory: (a) memory organization and storage during study and (b) memory retrieval during test. The distinction between storage and retrieval is a very common and important one in human memory research. Many theories exist that attempt to explain a specific memory phenomenon in terms of either storage or retrieval factors. In fact, researchers often explain memory deficits, such as those due to aging or clinical dysfunction, as being caused by a deficit of storage or retrieval capacity (see Butters, Delis, & Lucas, 1995, for a review). Batchelder and Riefer (1999, pp. 65–68) reviewed a number of MPT models that have been designed to separately measure storage and retrieval processes in human memory, including the pair-clustering model discussed in this article. Models of this type have the capability to provide theoretically based interpretations of data patterns, and thus they supplement standard statistical analyses of data in an attempt to better understand the unobservable cognitive processes that underlie memory phenomena.

This article is organized as follows. First, the model and some of its statistical theory are described. Then, a new approach to applying the model to a multitrial study is developed. After this, we concentrate on showing that the model gives a valid interpretation of the subskills it purports to measure. This analysis involves a review of previous evidence for the validity of the model as well as a presentation of two new studies involving normal participants. Both of these studies demonstrate that a basic experimental manipulation has selective and predicted effects on one of the parameters of the model. The final section of this article presents two new experiments that examine storage–retrieval deficits in a pair of well-studied clinical populations: patients diagnosed with schizophrenia and patients with organic brain damage resulting from prolonged alcoholism. In the Conclusion section, we discuss the usefulness of this approach and its advantages over traditional methods for assessing clinical populations.

Pair-Clustering Model

Model Description

Although the pair-clustering model has been described elsewhere (e.g., Batchelder & Riefer, 1980, 1986, 1999), it is helpful to briefly describe it here. The basic experimental task is a variant of the free-recall paradigm, in which participants study a list of items and then attempt to recall the list in any order. In a pair-clustering task, the items consist of category pairs (e.g., *doctor*, *lawyer*) and possibly a number of singletons (words without a category mate in the list). Recall of a category pair is scored in one of four categories: E_1, both items recalled consecutively; E_2, both items recalled but not consecutively; E_3, one and only one item recalled; and E_4, neither item recalled. Recall of a singleton is classified into two categories: F_1, recalled, and F_2, not recalled. We denote N_i to be the total number of E_i events ($i = 1, 2, 3, 4$), with $N = N_1 + N_2 + N_3 + N_4$. We denote M_j to be the total number of F_j events ($j = 1, 2$), with $M = M_1 + M_2$. Usually these category frequencies are obtained by pooling data over participants and items within an experimental treatment group when there is reason to suppose that the aggregated data are reasonably homogeneous.

The model assumes that the probabilities of the recall event categories are a function of three hypothetical (latent) parameters: c, r, and u. Parameter c represents the probability of forming a cluster for a category pair during study and storing and maintaining that cluster in memory until the time of test. Parameter r is the conditional probability that a stored cluster is successfully retrieved during recall. Parameter u is the probability that an unclustered item is both stored and retrieved. In essence, parameters c and r constitute separate measures of cluster storage and cluster retrieval, respectively. The parameter u is a composite of storage and retrieval processes. From a statistical viewpoint, u is a *nuisance parameter*, in the sense that it is not of primary interest to investigators but is needed to complete the model. Figure 1 presents the processing-tree diagram for the model, which expresses the recall events as a function of the model's parameters. Separate trees are postulated for the category pairs and the singletons. In Figure 1, parameter u appears in both trees, which represents the way that items in unclustered pairs are recalled like singletons.

The model incorporates a number of simplifying assumptions that make it tractable. For example, it is assumed that consecutive recall of a category pair (event E_1) can occur only if the pair is stored and retrieved as a cluster. In addition, it is assumed that if

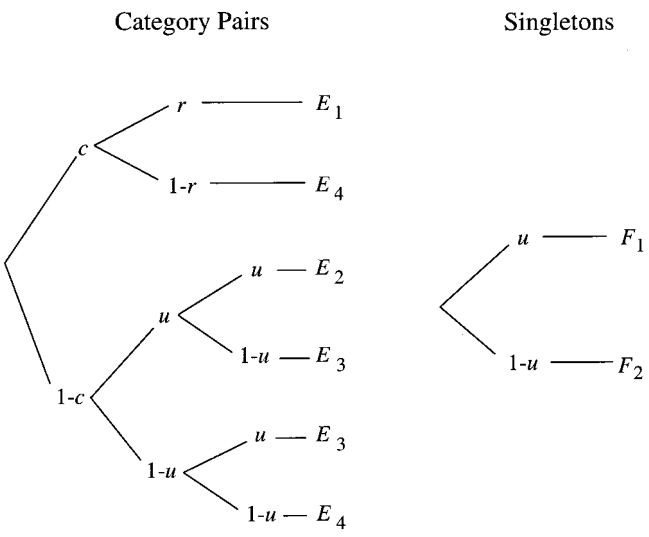

Figure 1. Batchelder and Riefer's (1980, 1986) pair-clustering model. c = probability of forming a cluster for a category pair during study; r = conditional probability that a stored cluster is successfully retrieved during recall; u = probability that an unclustered item is both stored and retrieved; E_1 = both items recalled consecutively; E_2 = both items recalled but not consecutively; E_3 = one and only one item recalled; E_4 = neither item recalled; F_1 = recalled; F_2 = not recalled.

a pair of related items is not clustered (with probability $1 - c$), then the two items in the pair are treated as singletons and are recalled independently (each with probability u). Thus, if $c = 1$, the contents of memory consist entirely of clustered pairs, whereas if $c = 0$, the contents are only single items. For intermediate values of c, the contents are a mixture, and retrieval is about bringing out the items in storage. Batchelder and Riefer (1986, 1999) discussed the logic behind these simplifying assumptions in more detail.

Model Analysis

With three parameters and 4 *df* in the data (3 *df* from the pair data plus 1 *df* from the singleton data), the goodness of fit for the model can be evaluated using the log-likelihood ratio statistic G^2 (Read & Cressie, 1988). In essence, G^2 measures the difference between the N_i data frequencies and the predicted values of those frequencies based on the "best fit" of the model, as obtained by maximizing the likelihood function. Larger values of G^2 mean a greater disparity between the model's predictions and the actual data and, hence, a poorer fit of the model. Interpretation of G^2 for large data sets is the same as for the chi-square statistic; in particular, the critical values for testing the significance of G^2 are the same as those used with chi-square (see Riefer & Batchelder, 1988, for details). If the model provides a poor fit to data for a particular study, the assumption that items in unclustered pairs are recalled at the same rate as singletons may be false. Under these circumstances, it is customary not to use the singleton data but to estimate c, r, and u using data only from the category pairs (see Batchelder & Riefer, 1986). In this case, unique estimates of c, r, and u can be obtained, but there are no degrees of freedom left for testing the goodness of fit of the model. One can check to see if the basic pattern of results for parameters c and r are the same under each of these two approaches, which is another way to evaluate the role of u in the model.

Once the N_i and M_j category frequencies have been obtained, a data analysis program called GPT.EXE (Hu, 1999; described in Hu & Phillips, 1999) can be used to provide maximum likelihood estimates for the model's parameters. Maximum likelihood estimates are the values of the model's parameters that maximize the likelihood function, an equation that expresses the probability of the data frequencies as a function of the parameter values (see Riefer & Batchelder, 1988, p. 321). The GPT.EXE program can also be used to conduct hypothesis tests for the model's parameters. These tests compare the null hypothesis, which assumes that parameter values are constant across experimental conditions, with an alternative hypothesis that assumes parameters have different values across conditions. Like goodness-of-fit tests, hypothesis tests are conducted using the log-likelihood ratio statistic G^2 (Riefer & Batchelder, 1988). In this case, G^2 is a measure of the difference between the expected data frequencies based on the parameter restrictions from the null hypotheses and predicted frequencies based on the version of the model without parameter restrictions.

Because of their simplicity, an advantage of MPT models is that their statistical analysis can be supplemented in an informative way with computer simulations of the model. Model simulations allow researchers to explore various measurement issues, such as bias in the parameter estimates, the robustness of the model when there are individual differences in the model's parameters, and the power to detect significant differences between different groups (Batchelder & Riefer, 1999). Riefer and Batchelder (1991b) have conducted extensive simulations of the pair-clustering model and have found that if the values of the parameters are not extreme, then the model is sufficiently robust and has negligible bias if the number of data observations (N) equals at least 150.

We conducted computer simulations of the pair-clustering model for each of the four experiments presented in this article. For each simulation, we took the maximum likelihood estimates of the parameters from the experimental data and used them as the values of the model's "true" parameters in the simulations. We then used the simulation option in GPT.EXE to conduct 1,000 Monte Carlo runs of the model. Each run used the same values of N and M as in the original experiment and set the parameters equal to their estimates from the experiment to generate values of the N_i and M_j statistics. The simulation then used those statistics to compute new parameter estimates. The values of these new estimates across 1,000 runs give a sampling distribution of the estimators for each of the parameters. If the means of these sampling distributions are close to the estimated values of the parameters from the experiment, then the evidence is strong that the estimators have negligible bias.

Another advantage of computer simulations is that they can be used to estimate the standard error of the estimators for each parameter. In this case, a standard error for an estimator is based on the variability of the estimator's theoretical sampling distribution, given the sample size used in the experiment. Riefer and Batchelder (1988) discussed traditional statistical methods for obtaining approximate standard errors based on so-called asymptotic (or large sample) methods; however, in this article, we use a more exacting method based on simulations for MPT models described by Riefer and Batchelder (1991b). Once the 1,000 simulated data

sets and the sampling distributions of the parameter estimates are created, one can compute the standard deviations of these sampling distributions. These standard deviations become estimates for the standard error of the maximum likelihood estimates that are not based on the asymptotic approximations. This approach is advantageous because it can be used to estimate standard errors for any sample size. In particular, Riefer and Batchelder (1991b) found that when the estimate of c is small, unrealistically large sample sizes often are needed to obtain accurate standard errors from the asymptotic theory, especially for the estimate of the retrieval parameter r.

Application to Multiple Trials

When a memory task is used in clinical assessment, it is often useful to conduct multiple study–test trials on the task. Patients can then be assessed in the early, middle, and late stages of learning; deficits may occur in some of these stages and not others. Multitrial data can yield measures of cognitive deficits that are more stable because they are measured over many trials, which offers more power in differentiating between clinical populations. It is, therefore, natural to consider running multiple trials when conducting a pair-clustering experiment. Most of the studies reported in this article involve memory performance across multiple presentation trials.

Because the structure of a pair-clustering experiment offers the learner repeated opportunities to store and retrieve items, it is psychologically reasonable to assume that all three parameters of the pair-clustering model are nondecreasing over successive trials. When decreases in parameter estimates over trials occur, they could be due to random variability in small samples, due to extraneous factors such as boredom or fatigue, or possibly due to the fact that the model assumptions are not valid for the data. In general, however, any decrease across trials is difficult to rationalize in the context of the psychological interpretation of the parameters.

Fortunately, there is an approach to analyzing multitrial MPT models that can incorporate the assumption of nondecreasing parameters over trials. The idea is to apply the model to the category frequencies from each trial separately but to impose parametric order constraints across trials. This approach is developed in detail by Knapp and Batchelder (2001). To apply this approach here, we assume that none of the parameters of the pair-clustering model decrease over trials. For an experiment with n trials, this creates the following constraints:

$$0 \leq c_1 \leq c_2 \leq \ldots \leq c_n \leq 1,$$
$$0 \leq r_1 \leq r_2 \leq \ldots \leq r_n \leq 1,$$
$$0 \leq u_1 \leq u_2 \leq \ldots \leq u_n \leq 1.$$

For an n-trial experiment without order constraints, the parameter space for the model is $[0,1]^{3n}$, because each of the three parameters is free to vary independently in [0,1] for each of the n trials. However, the parameter space that satisfies the imposed order constraints is much smaller and, in fact, reduces the size of the unconstrained space by the factor of $\left(\frac{1}{n!}\right)^3$. For example, in a six-trial experiment, the fraction of the unconstrained parameter space consistent with the order constraints is just $\frac{1}{373{,}248{,}000}$ of the original space. Because of this extreme reduction, it is not surprising that parameter values for the unconstrained model may sometimes violate the constraints simply due to random variation. However, the constrained model will impose these constraints over the estimates and, thus, may be viewed as a psychologically reasonable way to smooth the pattern of estimates over trials.

Unfortunately, the computer program for analyzing MPT models described by Hu and Phillips (1999) cannot be used directly under parametric order constraints. For the search algorithm to work, each parameter must be independent and free to vary in the full interval [0,1] (see Hu & Batchelder, 1994). One possible solution would be the development of a special optimization program to search only in the fraction of the parameter space that satisfies the order constraints for the parameter estimates maximizing the likelihood function. A second approach, first described by Batchelder (1999) and developed for multitrial MPT models by Knapp and Batchelder (2001), is construction of a new MPT model that is "statistically equivalent" to the pair-clustering model, with the inclusion of the order constraints described earlier. Two models are said to be statistically equivalent if they are indistinguishable on data, such as models that are reparameterizations of each other. Knapp and Batchelder (2001) have shown that construction of such a model is always possible for any MPT model that is subject to nondecreasing, multitrial parametric order constraints. The basic idea is to reparameterize the model to reflect the order constraints and then to show that this reparameterized version can be represented as an MPT model.

To illustrate this approach, consider a simple case of two trials and a single order constraint involving the storage parameter, $c_1 \leq c_2$ (this example is also applicable to the other two parameters, r and u). One approach is to introduce a new parameter β, $0 \leq \beta \leq 1$, and to define $(1 - c_2) = \beta(1 - c_1)$. If this relation holds, it follows that

$$c_2 = (1 - \beta) + \beta c_1, \tag{1}$$

and thus $c_1 \leq c_2$. The new model has parameters β and c_1, instead of c_1 and c_2; once estimates of β and c_1 are obtained from the data, an estimate of c_2 can be obtained through Equation 1. In fact if $\hat{\beta}$ and $\hat{c}_1$ are maximum likelihood estimates in the new model, then once they are inserted into Equation 1, $\hat{c}_2$ becomes a maximum likelihood estimate for the original model subject to the order constraint. The key to this approach is that the reparameterized model is also an MPT model, where all the parameters are free to vary in the [0,1] interval (Batchelder, 1999). Thus, one can obtain maximum likelihood estimates for these new parameters using GPT.EXE described by Hu and Phillips (1999). Then, it is straightforward to calculate the maximum likelihood estimates of the original parameters from the maximum likelihood estimates of the parameters from the new model. The specific procedure is described in detail in Knapp and Batchelder (2001), so we omit the methods of creating the new MPT here. However, it should be noted that this approach sometimes considerably expands the size of the trees representing the model. For example, for a six-trial experiment involving just pairs (and no singleton items), the model without imposing order constraints has 36 end branches. In contrast, the model reflecting the order constraints on all three parameters over six trials yields a tree with 256 end branches.

In general, if there are n trials in an experiment, then the reparameterized model will contain $(n - 1)$ β_i parameters ($i = 2$ to n) and one model parameter c_1. These arise from the constraints

$\beta_i(1 - c_{i-1}) = 1 - c_i$; given these constraints, it is easy to express all the c_i in terms of c_1 and the β_i. The result is

$$c_i = 1 - (1 - c_1)\prod_{j=2}^{i}\beta_j. \quad (2)$$

Separate but similar equations, with different values of β_i, can also be generated for parameters r and u. We refer to the β_i as the *rate parameters* of the new model, because they give an indication as to the rate of learning over trials. In addition to imposing order constraints, these rate parameters themselves have a psychologically useful and interesting interpretation. Specifically, the β_i parameters can be interpreted as the error reduction rate from trial $i - 1$ to trial i; that is, they represent the proportion of reduction in the failure rate of a cognitive process from one trial to the next. In general, lower values of β_i signify faster rates of learning (greater error reduction), and a value of $\beta_i = 1$ indicates that the cognitive subskill has not increased over trials ($c_i = c_{i-1}$).

Parameters with a rate interpretation allow for many hypothesis tests that are not available when multinomial modeling in a multitrial situation without order restrictions is used. For example, this methodology allows researchers to compare clinical groups not only on their overall level of performance on a cognitive subskill but also on their rate of improvement of that subskill over trials. Another possibility is that two groups may learn a subskill at the same rate but have different starting points. This is achieved in the model by allowing the parameter estimate for that subskill to vary on Trial 1 for each group while having the same values of the rate parameters β_i for each group across the remaining trials. Furthermore, if we assume that the rate parameters do not vary over trials (i.e., $\beta_i = \beta$ for all i), then Knapp and Batchelder (2001) have shown that this version of the model is a testable submodel that is equivalent in form to the well-known Bush–Mosteller (Bush & Mosteller, 1955) learning model. The resulting equation for this model is easy to derive from Equation 2 and is given by

$$\theta_i = 1 - \beta^{i-1}(1 - \theta_1) \quad (3)$$

for all i, where $0 \leq \beta \leq 1$ and θ_i is either c_i, r_i, or u_i in the pair-clustering model. Equation 3 represents a geometric reduction in the error rate leading to perfect performance as the number of trials increases. In the experiments to follow that involve multiple learning trials, we impose similar order constraints on the parameters for the model's analysis. In these cases, we analyze the data with different values of the rate parameters as well as the special case of constant error rate reduction in the parameters over trials.

Validation Studies

When any measurement model for clinical assessment is utilized, it is important to determine if the model's parameters are in fact valid measures of their corresponding cognitive capacities. This type of validity is not assessed merely by showing that the model fits the data, because a good statistical fit does not in itself imply that the parameters are playing the role that the theorist thinks they are. In the pair-clustering model, the parameter c should ideally measure storage capacity and should not be affected by experimental manipulations that impact other cognitive functions. Similarly, the parameter r should in theory measure only the retrieval capacity of successfully stored clusters and thus should be unaffected by variables that do not impact retrieval. An essential step, then, in the development of any MPT model is the performance of validation studies for the model. Validity testing of a model typically involves a series of experiments in which selected experimental variables are manipulated. The goal is to show that certain variables have a selective influence on the key parameters of the model that is interpretable on theoretical or logical grounds (see Batchelder & Riefer, 1999, p. 76, for more on validity testing of MPT models).

A good example of a validity test for the pair-clustering model can be found in a study by Batchelder and Riefer (1980). They conducted a between-groups pair-clustering experiment in which the pairs consisted of high-associate members of categories for one group (Experiment 1A) or low associates for the other group (Experiment 1B). It is reasonable to hypothesize that the level of association is a variable that should have its primary effect on the clustering and storage of category pairs. It is presumably easier to see the relationship between category members when their association level is high, which should aid their encoding and storage. Although Batchelder and Riefer (1980) did not conduct any formal analysis of association level, they did present the estimates of c and r for both high and low associates (see Batchelder & Riefer, 1980, Table 4). In general, level of association had strong and predictable effects on parameter c but no systematic effect on parameter r. We reanalyzed the original Batchelder and Riefer (1980) data by using GPT.EXE and found that high-associate pairs had significantly higher values of c, $G^2(5) = 300.18$, $p < .001$. In contrast, there was no significant effect (at the .05 level) of association level on parameter r, $G^2(5) = 1.64$. This is a theoretically reasonable result that is consistent with the psychological assumptions behind the pair-clustering model and provides some face validity for the model.

Another test of the model's validity can be found in a study by Riefer and Batchelder (1987) that examined memory performance for free versus cued recall. In cued recall, the participant is given cues during the memory test that are designed to aid retrieval. The assumption made by many theorists is that cued recall tends to facilitate retrieval factors in memory without affecting the original storage factors. If so, then cued recall should have its primary effect on parameter r of the pair-clustering model, with little or no effect on the storage parameter c. In the control condition of the Riefer and Batchelder (1987) study, participants recalled without any cues, whereas in the cued condition, participants were given the names of the categories for the pair clusters during their recall. The model's analysis revealed that there was no difference in storage capacity between free and cued recall but that cued recall significantly improved retrieval capacity.

In this section, we present two previously unpublished experiments designed to further test the validity of the pair-clustering model as a measurement tool. The first experiment manipulates the variable of presentation rate during list study, which should have its main influence on the storage parameter c. The second experiment examines a phenomenon called part-list cuing: Participants memorize a list of words, and then their recall is cued during the memory test with a subset of the items from the original list. Most theories of part-list cuing assume that this manipulation disrupts the participants' retrieval strategies for recalling the list; the manipulation should therefore affect the retrieval parameter r.

Experiment 1: Presentation Rate

When stimuli are presented serially, the rate of presentation for those stimuli can have an influence on how they are processed. Researchers in the field of human memory have manipulated presentation rate to study a number of different phenomena, including the bizarreness effect (Kline & Groninger, 1991), word comprehension (Warrington & Cipolotti, 1996), and priming effects (Weldon & Jackson-Barrett, 1993). More recently, presentation rate has been used in a number of studies to examine memory deficits in various clinical populations (e.g., Fazio, 1998; Turkstra, 1998). In essence, variation in the rate of stimulus presentation is a manipulation that occurs during the study phase of a memory experiment. As such, it seems reasonable that such a manipulation should have its primary effect on the clustering and storage of related items (although it is logically possible that presentation rate could effect their retrievability as well). Thus, a natural prediction for the pair-clustering model would be a significant effect of presentation rate on parameter *c*.

Method. The participants were 42 undergraduates from the University of California, Irvine. Each participant was presented with a list of 40 words for memorization, consisting of a primacy buffer (first 7 words in the list), recency buffer (last 8 words in the list), pairs of categorically related nouns (10 pairs), and unpaired nouns (5 words). To control for serial position effects, we did not include the primacy and recency items in the reported analyses. All words were taken from the Battig and Montague (1969) category norms and consisted of high associates (i.e., exemplars that are the most frequently associated with each category). For each participant, items for the category pairs, singletons, and buffer regions were randomly selected from a larger pool of 24 category pairs. Specifically, the category pairs used in the list were randomly selected from 10 of these categories, the singletons were selected from 5 additional categories, and words in the primacy and recency buffer were taken from the remaining categories. Presentation of the items from the list was random under the constraint that items from the same category were not presented consecutively.

Participants were run individually; the words in each list were presented for study one at a time on a computer screen. Half of the participants saw each list at a rate of 2 s per word; the other half saw each list at a 5-s presentation rate. Each participant was presented with four study–test trials for each list, with a different random order of the list for each presentation. After each list presentation, participants were given 3 min to freely recall in writing whatever words they remembered.

Results and discussion. Many researchers in the field of human memory have traditionally used empirical statistics to measure memory processes, such as the proportion of words correctly recalled, $P(C)$. Such an analysis for the current experiment reveals that recall was significantly better with a 5-s presentation rate than a 2-s rate, $F(1, 40) = 16.27$, $p < .001$, $\omega^2 = .06$, and that recall improved over the four study–test trials, $F(3, 120) = 175.71$, $p < .001$, $\omega^2 = .58$. However, an important question is whether these effects are due to storage processes, retrieval processes, or possibly both. This is the kind of question that MPT models, as opposed to empirical summary statistics, are capable of answering.

Table 1 shows the N_i and M_j statistics pooled over participants and items and presented for the 2-s and 5-s presentation rates for each of the four study–test trials. Table 1 also includes the G^2 goodness-of-fit statistics for each condition. As it turns out, the fit of the model is very good across all conditions. (This is based on a critical value of 3.84 for each goodness-of-fit test, using 1 *df* and the .05 significance level. For all remaining statistical tests, we report *p* values only if they are less than .05.) We applied the GPT.EXE computer program to the data in Table 1 to estimate the model's parameters for each condition. Because this is a multitrial experiment, the parameter estimates can be computed under the order constraints described earlier. First, the maximum likelihood estimates of the reparameterized model, namely β_i and c_1, are obtained in the usual way from GPT.EXE. These estimates are then entered into Equation 2 to obtain the maximum likelihood estimates c_i. The same approach yields maximum likelihood estimates for the parameters r_i and u_i, with different sets of β_i for each parameter. These estimates are presented in Table 1 along with the standard error for these estimates derived from computer simulations of the model.

Figure 2 presents the parameter estimates from Experiment 1; storage and retrieval improve across the four study–test trials. This increase across trials is significant for both *c*, $G^2(6) = 46.84$, $p <$

Table 1
Recall Statistics, Goodness-of-Fit Measures, and Parameter Estimates for Experiment 1

Trial	N_1	N_2	N_3	N_4	M_1	M_2	$G^2(1)$	*c*	*r*	*u*
						2 s				
1	41	13	55	101	23	82	1.74	.26 (.08)	.75 (.20)	.25 (.03)
2	68	20	61	61	41	64	0.01	.39 (.06)	.82 (.11)	.39 (.04)
3	107	19	44	40	63	42	2.74	.62 (.04)	.82 (.05)	.55 (.04)
4	126	31	33	20	80	25	2.51	.67 (.04)	.90 (.03)	.72 (.03)
						5 s				
1	59	17	40	94	35	70	2.16	.55 (.07)	.51 (.08)	.37 (.04)
2	99	21	38	52	66	39	1.61	.66 (.04)	.71 (.04)	.59 (.04)
3	138	30	16	26	89	16	0.92	.77 (.03)	.85 (.03)	.83 (.03)
4	154	33	13	10	97	8	3.23	.78 (.03)	.94 (.02)	.89 (.02)

Note. Numbers in parentheses are the standard errors for each parameter. $G^2(1)$ is the log-likelihood ratio statistic based on 1 *df*. All values of G^2 are below the critical value of 3.84, indicating a good fit of the model. N_1 = both items recalled consecutively; N_2 = both items recalled not consecutively; N_3 = one and only one item recalled; N_4 = neither item recalled; M_1 = singleton recalled; M_2 = singleton not recalled; *c* = probability of forming and storing a cluster; *r* = probability of retrieving a cluster, if stored; *u* = probability of recalling a nonclustered item as a singleton.

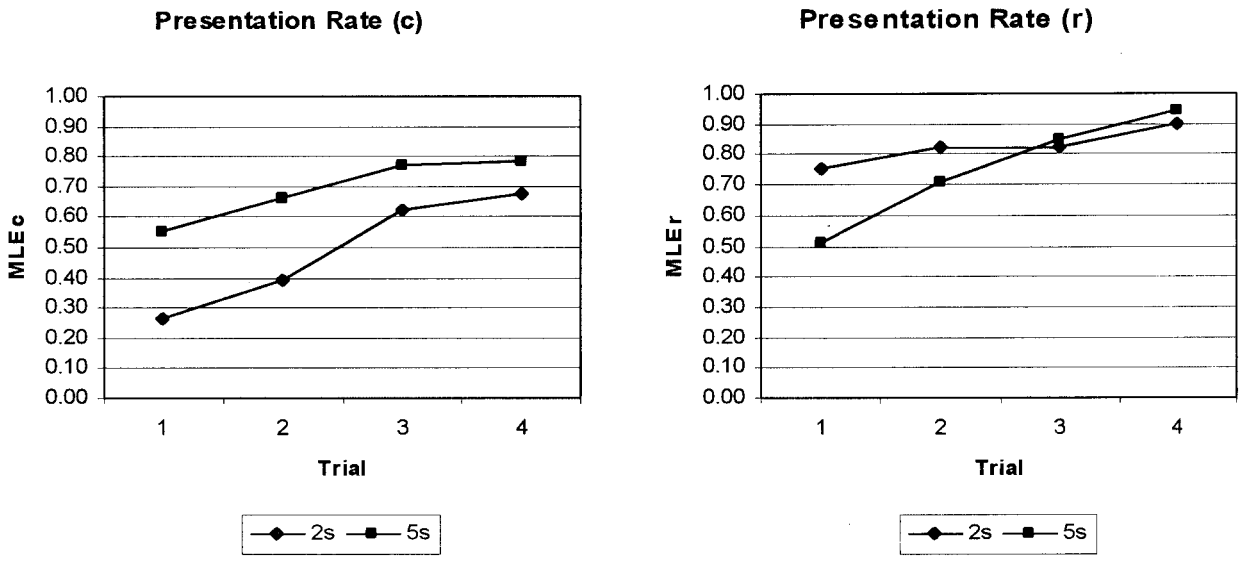

Figure 2. Estimates of *c* (left panel) and *r* (right panel) for the 2-s and 5-s presentation rates. MLE = maximum likelihood estimate.

.001, and *r*, $G^2(6) = 32.05$, $p < .001$. There is also a clear and consistent storage advantage, as measured by parameter *c*, for lists presented at a 5-s rate, $G^2(4) = 34.41$, $p < .001$. In contrast, there is no significant effect of presentation rate on retrieval, as measured by parameter *r*, $G^2(4) = 3.87$. This result is consistent with our earlier hypothesis that presentation rate should have its primary effect on storage and not retrieval capacity.

Because this is a multitrial experiment, it is also possible to compare the rate of increase for parameters *c* and *r* over trials. This is done directly by using the MPT model that incorporates the order constraints and comparing the values of the rate parameters (β_i) for storage and retrieval. This analysis reveals that the increase in parameter *c* across trials is not significantly different for 2-s versus 5-s presentations, $G^2(3) = 0.68$. The rate of increase for parameter *r* also shows no significant difference between the two presentation rates, $G^2(3) = 3.83$. Thus, the significant differences in *c* between presentation rates can be attributed to different starting values for c_1 for the first trial but equivalent rates of learning across additional trials.

Experiment 2: Part-List Cuing

Part-list cuing is an interesting memory phenomenon first demonstrated by Slamecka (1968). In a typical experiment, participants memorize a list of words and attempt either an uncued recall of the words or a cued recall in which some of the list words are given as cues. The usual finding is that the cuing of memory with words from the list does not improve memory, and, in fact, it generally has a harmful effect. Because list items can be used as cues to the recall of other list items, this negative effect of recall cues has intrigued memory theorists for many decades and is still of interest to researchers in this field (e.g., Raaijmakers & Phaf, 1999). Nickerson (1984) reviewed a number of different theories that have been proposed to explain part-list cuing. All the theories propose that cued items somehow disrupt or inhibit the retrieval of the uncued items. For example, the strategy disruption hypothesis (Basden & Basden, 1995) states that when participants are forced to organize their recall around arbitrary list items, they engage in a less efficient retrieval strategy than when they use their own subjective organization. If part-list cuing is essentially a retrieval phenomenon, then an analysis of this phenomenon by the pair-clustering model should primarily focus on changes in parameter *r*. Specifically, estimates of *r* should be significantly lower under part-list cuing as compared with free recall. Experiment 2 tests this prediction.

Method. Participants were 191 undergraduates from the University of California, Irvine. Each participant was shown a list of 40 words, which were presented by slide projector at a rate of 5 s per word. The list consisted of a primacy buffer (first 5 words), recency buffer (last 5 words), pairs of categorically related nouns (12 pairs), and unpaired nouns (6 words) from Battig and Montague (1969). As in Experiment 1, the primacy and recency items were not included in the final data analysis. Short lags were used in Experiment 2; specifically, the number of intervening items between words from the same category ranged from 1 to 3 items. After presentation of the list, participants were given 3 min for recall. There were 96 participants in the free-recall condition, and these participants were allowed to recall the words without any constraints or cues. There were 95 participants in the part-list cuing condition, in which participants were presented with a recall sheet containing 18 words from the list. These 18 words consisted of 6 of the 10 buffer items, 3 of the 6 singletons, and 9 of the 24 category items. Of the 12 category pairs, 2 pairs had both words as cues, 5 pairs had one of the words in the pair as a cue, and 5 pairs had neither category pair as a cue. The 18 cued words were chosen randomly for each participant contingent on the aforementioned constraints, and the words appeared in alphabetical order on the participants' recall sheets. Participants who were presented with the part-list were instructed to recall as many additional words from the list as they could remember.

Results and discussion. The proportion of available words recalled was significantly lower in the part-list condition than in the free-recall condition, $P(C) = .36$ versus .48, respectively; $F(1,$

189) = 24.01, $p < .001$, $\omega^2 = .11$; this result is consistent with previous studies on part-list cuing. The question, however, is whether this deficit is due to storage or retrieval factors, which can now be explored with the pair-clustering model. Table 2 presents the N_i and M_j statistics for the model, along with the model's goodness of fit and parameter estimates for each condition. Only category pairs and singletons that were not cued during recall were included in this analysis. Thus, all items from the free-recall lists were analyzed, but only the five uncued category pairs and the three uncued singletons from the part-list cuing condition were analyzed. As measured by G^2, the fit of the model is fairly poor (critical value = 3.84). One reason for this is a high level of power to reject the model created by the large number of data observations for the experiment (191 participants, creating 1,627 total N_i events for the model; see Riefer & Batchelder, 1991b, for a discussion of sample size and model fit tests). We decided, therefore, to analyze the data from Experiment 2 by using the version of the model that does not estimate parameter u from the singletons. As it turns out, both versions of the model produced the same pattern of results.

Because Experiment 2 did not involve multiple trials, no order constraints were imposed on the parameter values, so there are no rate parameters to analyze. Figure 3 presents the values of parameters c and r for both the free and part-list recall conditions. Analysis with the GPT.EXE program revealed that part-list cuing had virtually no effect on storage (parameter c), $G^2(1) = 0.01$. However, the retrievability of items (parameter r) was significantly lower as a function of part-list cuing, $G^2(1) = 7.61$, $p < .01$. This result is consistent with our hypothesis based on previous theories of part-list cuing.

Experiments 1 and 2 provide converging evidence for the validity of the pair-clustering model as a method for separately measuring storage and retrieval. In Experiment 1, presentation rate is shown to affect the storage parameter c but has no effect on the retrieval parameter r. In contrast, Experiment 2 shows that part-list cuing reduces the value of r but has no comparable effect on c. Thus, both manipulations demonstrate a selective influence on the parameters of the model that is consistent with logical and theoretical expectations. Moreover, these conclusions about storage and retrieval would have been impossible to make using only an examination of empirical measures, such as $P(C)$.

The aforementioned results, as well as the results of prior validity studies, speak well for the usefulness of the pair-clustering model as a tool for measuring storage and retrieval. However, as we stated earlier, the current model makes some strong assumptions about the storage and recall of category pairs. It is possible that other models, making different assumptions, might also prove to be viable models for measuring storage and retrieval capacities. Development of models and validation testing is always an ongoing issue. Fortunately, because Tables 1 and 2 contain the N_i and M_j frequency counts, theorists who are interested in testing alternative versions of the model have the necessary information to do so.

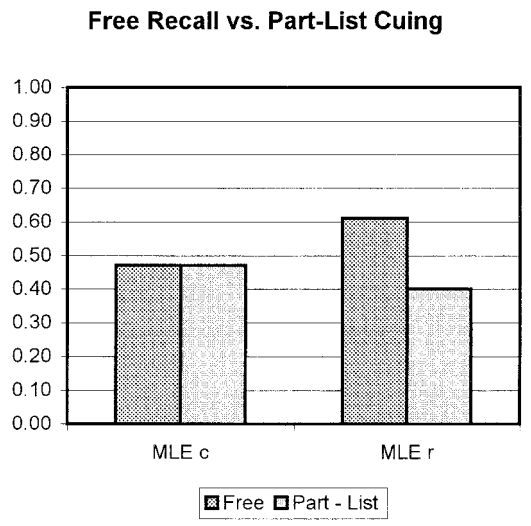

Figure 3. Estimates of c and r for free recall and part-list cuing conditions. MLE = maximum likelihood estimate.

Assessing Memory Deficits

Once a model has been supported through a series of validation studies, it can then be used as a measurement tool to explore a variety of theoretical and applied issues. In the case of the pair-clustering model, the new results from Experiments 1 and 2, along with previously published applications of the model, have established it as a valid tool for the measurement of storage and retrieval in the context of free recall of clusterable pairs.

As stated in the introduction, the purpose of this article is to show how MPT models can be used as diagnostic tools for the assessment of cognitive deficits in special populations, such as older adults or people with cognitive dysfunction. This strategy, cognitive psychometrics, has been discussed at the programmatic

Table 2
Recall Statistics, Goodness-of-Fit Measures, and Parameter Estimates for Experiment 2

Condition	N_1	N_2	N_3	N_4	M_1	M_2	$G^2(1)$	c	r	u
Free recall	333	111	276	432	215	361	4.44	.47 (.03)	.61 (.04)	.40 (.02)
Part-list cuing	90	42	102	241	86	199	7.96	.47 (.04)	.40 (.06)	.34 (.02)

Note. Numbers in parentheses are the standard errors for each parameter. $G^2(1)$ is the log-likelihood ratio statistic based on 1 *df*. Both values of G^2 exceed the critical value of 3.84, indicating a significantly bad fit of the model. N_1 = both items recalled consecutively; N_2 = both items recalled not consecutively; N_3 = one and only one item recalled; N_4 = neither item recalled; M_1 = singleton recalled; M_2 = singleton not recalled; c = probability of forming and storing a cluster; r = probability of retrieving a cluster, if stored; u = probability of recalling a nonclustered item as a singleton.

level in an article by Batchelder (1998). That article outlined procedures for using MPT models to make inferences about cognitive deficits at either the individual or group level. Such information, assessed at either level, has the potential to aid clinicians in the early detection and treatment of cognitive dysfunction. In fact, there are several applications of MPT models for the assessment of cognitive deficits that have appeared before and after Batchelder's (1998) programmatic article (e.g., Batchelder, Chosak-Reiter, Shankle, & Dick, 1997; Batchelder & Riefer, 1990; Bayen & Murnane, 1996; Bender, Wallsten, & Ornstein, 1996; Chechile & Roder, 1998; Chosak-Reiter, 2000; Gerrein & Chechile, 1977; Gutowski & Chechile, 1987; Howe & Brainerd, 1989; Keefe, Arnold, Bayen, & Harvey, 1999; Kraemer, Peabody, Tinkleberg, & Yesavagé, 1983). Our goal in this section is to illustrate how the validated pair-clustering model can be used to make group-level inferences about deficits in storage and retrieval processes.

An example of how the pair-clustering model can be used for this purpose comes from a study by Riefer and Batchelder (1991a) that examined memory deficits in older adults. A number of researchers (e.g., Schonfield & Robertson, 1966; Shaps & Nilsson, 1980) have theorized that poorer recall performance in older adults is due to problems with retrieval processes. To explore this, Riefer and Batchelder compared the memory performance of young versus older participants and analyzed the results using the pair-clustering model. The data revealed poorer recall in older participants, and the model showed that this recall deficit was in fact due primarily to retrieval differences, as measured by *r*, and not to differences in storage capacity, as measured by *c* (see Bayen, 1990, for a more extensive study on the effects of aging on storage and retrieval).

In the two experiments to follow, we explore storage–retrieval deficits in two clinical groups that are known to exhibit memory impairment: people diagnosed with schizophrenia and individuals with organic brain damage resulting from alcoholism.

Experiment 3: Memory Deficits in Schizophrenia

It is well documented that schizophrenics exhibit memory deficits relative to comparable controls (Heinrichs & Zakzanis, 1998; Stip, 1996). However, it is still an open question as to which cognitive functions are responsible for this impairment. Much research has concentrated on the specific mechanisms that underlie memory dysfunction in schizophrenics, and a portion of this research has focused on potential problems with either storage or retrieval capacity. Sengel and Lovallo (1983), for example, have called for studies "to determine whether the schizophrenic recall deficit is due primarily to encoding or retrieval dysfunction, or both" (p. 426).

However, no consistent pattern has emerged to suggest conclusively that schizophrenics suffer primarily from either storage or retrieval deficits. For example, early research has shown that although recall memory is impaired in schizophrenics, their recognition memory is often at the same level as controls (e.g., Bauman & Kolisnyk, 1976; Koh, 1978; Traupmann, 1975). Paulsen et al. (1995) have also noted that schizophrenics exhibit a disproportionate improvement in memory relative to controls when tested on recognition as compared with recall. From these results, some theorists have concluded that memory impairment in schizophrenics is basically due to problems with retrieval. But some researchers (Calev, 1984a; Gold, Randolph, Carpenter, Goldberg, & Weinberger, 1992) have found that schizophrenics can have poorer recognition memory. For example, Bamber (1979) observed that although schizophrenics had both recognition and recall deficits as compared with normals, their recall deficit was no worse than their recognition deficit. Such findings have led other theorists to conclude that memory deficits in schizophrenics are primarily caused by problems with storage.

A fair amount of research has also examined schizophrenics' memory in terms of its semantic organization. Studies have shown that schizophrenics are poor at category clustering; that is, they exhibit poorer recall and organization of items that come from semantic categories (Calev, 1984b; Koh, Kayton, & Berry, 1973; Traupmann, 1980). This has led some theorists (Brebion, Amador, Smith, & Gorman, 1997; Lyons et al. 1995) to conclude that schizophrenics exhibit an encoding deficit in their ability to organize semantic material. However, as illustrated in the pair-clustering model, poor clustering can be due to either a storage or retrieval deficit; so, it is possible that the problem lies in the retrieval process and not in the clustering process itself. In fact, other studies have shown that under certain conditions the provision of category cues during recall can result in greater improvement in the memory performance for schizophrenics as compared with controls (McClain, 1983; Traupmann, 1980), which implies that schizophrenics have a problem with retrieving information. For example, Barker (1977) used a modification of a paradigm developed by Tulving and Pearlstone (1966) in which recall of categorized lists is broken down into the number of categories recalled and the number of items recalled per category. He found that schizophrenics improved their memory of a categorized list when they received cues during recall and that the number of categories recalled, but not the number of items recalled per category, differed between schizophrenics and normals. Barker concluded from this that information in memory is stored and is thus potentially available to schizophrenics but is inaccessible (i.e., a problem with retrieval). In contrast, Brebion et al. (1997) found that the rate of memory improvement in schizophrenics for cued recall was equal to the benefit exhibited by normal controls. From this, they concluded that the original free-recall deficit was due to problems of encoding and not retrieval.

Other empirical results have been used to examine storage and retrieval factors in schizophrenics' memory, with equally divergent results. Theorists who favor an encoding deficit hypothesis have argued that experimental procedures that ensure sufficient encoding should eliminate this deficit. One way to accomplish this is with an incidental learning paradigm (Hyde & Jenkins, 1969) in which participants are given an orienting task that leads to deep processing of information. Many researchers have observed that schizophrenics' memory is equal to that of normal controls under these procedures (e.g., Koh, Kayton, & Peterson, 1976; Larsen & Fromholt, 1976), which supports the encoding deficit hypothesis. However, Traupmann (1980) has pointed out that the findings from incidental learning tasks have not been consistent and that the evidence for an encoding deficit is "less than compelling" (p. 705). Other researchers have noted that schizophrenics show faster forgetting rates than normal controls (e.g., Beatty, Jocic, Monson, & Staton, 1993). Calev, Venables, and Monk (1983) saw this as evidence for a postencoding deficit. Sengel and Lovallo (1983)

have speculated that such a finding is most likely due to problems with retrieval, either because of output interference or poor organization of information. However, Bamber (1979) observed that schizophrenics did not forget faster than normal controls over a 24-hr retention interval.

The aforementioned research illustrates the potential difficulty of using ad hoc procedures, rather than mathematical models, to examine separate cognitive processes such as storage and retrieval. Empirical measures like free recall, cued recall, or recognition are often influenced by many cognitive processes and, thus, are not a pure measure of any single process. For this reason, conclusions drawn from the changes in these measures from one clinical group to another can be difficult to interpret. Mathematical models, designed specifically to measure underlying cognitive processes, can overcome some of these difficulties by providing more direct, theoretically motivated measures of these processes.

To explore this issue, Experiment 3 applies the storage–retrieval model to a study comparing memory performance in schizophrenics versus normal controls. A standard memory test was administered to a group of diagnosed schizophrenics and to a group of nonpsychotic controls. Each group memorized and recalled a word list consisting entirely of clusterable pairs over a series of six study–test trials. The pair-clustering data from this study enable us to use the storage–retrieval model to determine if memory impairment in schizophrenia is due to deficits in storage or retrieval capacity.

Method. All participants were male patients at Veterans Administration Medical Center in St. Cloud, Minnesota, who volunteered to be in the experiment. The 29 schizophrenic patients met the criteria established by Feighner et al. (1972) for probable or definite schizophrenia, and all were psychotic at the time of testing. The 25 controls were nonpsychotic patients who were neither diagnosed with nor suspected of having schizophrenia, major affective disorder, or organic brain syndrome. The schizophrenics and controls were selected so that they were matched on both their mean age and level of education.

Participants memorized a 40-word list that consisted of two relatively high-associate exemplars from each of 20 categories in Battig and Montague (1969). The words were presented on a memory drum at a rate of 4 s per word. After viewing the list, participants worked on a series of addition problems for 30 s, followed by a 5-min free recall. This procedure was repeated for a total of six study–test trials. Presentation of the words was random on each trial subject to the constraint that words from the same category were widely separated. Specifically, category pairs appeared in either the first and third quarter of the list or the second and fourth quarter.

Results and discussion. An analysis of variance (ANOVA) conducted on the proportion of words correctly recalled revealed that the patients diagnosed with schizophrenia recalled significantly fewer words than the controls, $F(1, 52) = 15.39$, $p < .001$, $\omega^2 = .12$. The groups also improved their recall performance over the six study–test trials, $F(5, 260) = 215.68$, $p < .001$, $\omega^2 = .32$. It is of greater interest to apportion these effects over storage or retrieval mechanisms. Table 3 presents the N_i data statistics across trials for both the schizophrenics and controls. Because there are no singletons, we applied GPT.EXE to these data by using the same version of the pair-clustering model as we used in Experiment 2. However, like Experiment 1, this is a multitrial study, so we were able to use the reparameterized version of the model incorporating order constraints. The maximum likelihood estimates under the order constraints from this version of the model are also presented in Table 3, along with the standard error for each estimate derived from the computer simulations.

Table 3
Recall Statistics and Parameter Estimates for the Schizophrenics and Their Controls in Experiment 3

Trial	N_1	N_2	N_3	N_4	c	r	u
				Schizophrenics			
1	31	15	154	380	.15 (.11)	.36 (.35)	.19 (.03)
2	79	45	163	293	.38 (.07)	.36 (.10)	.35 (.04)
3	127	63	160	230	.44 (.05)	.50 (.06)	.44 (.04)
4	148	74	149	209	.47 (.04)	.53 (.05)	.49 (.04)
5	176	73	152	179	.47 (.04)	.63 (.06)	.49 (.04)
6	198	67	138	177	.52 (.04)	.65 (.05)	.49 (.04)
				Controls			
1	49	31	148	272	.29 (.09)	.34 (.18)	.30 (.04)
2	116	76	141	167	.42 (.04)	.54 (.06)	.51 (.04)
3	190	66	136	108	.47 (.04)	.82 (.06)	.51 (.04)
4	243	68	108	81	.57 (.03)	.84 (.04)	.57 (.04)
5	269	77	79	75	.64 (.03)	.84 (.03)	.65 (.03)
6	301	76	67	56	.68 (.02)	.88 (.02)	.70 (.04)

Note. Numbers in parentheses are the standard errors for each parameter. N_1 = both items recalled consecutively; N_2 = both items recalled not consecutively; N_3 = one and only one item recalled; N_4 = neither item recalled; c = probability of forming and storing a cluster; r = probability of retrieving a cluster, if stored; u = probability of recalling a nonclustered item as a singleton.

The parameter estimates from Experiment 3 are presented in Figure 4; there is a general increase in $\hat{c}$ (the estimate of c) and $\hat{r}$ (the estimate of r) across the six study–test trials for both the schizophrenics and controls. This increase across trials is significant for both c, $G^2(10) = 89.74$, $p < .001$, and r, $G^2(10) = 29.22$, $p < .01$. In addition, a direct comparison of the two groups shows that the schizophrenics' performance is poorer than the controls on both storage and retrieval. These differences, especially for $\hat{c}$, tend to be most pronounced in the later trials. In fact, the difference between schizophrenics and controls on parameter c is not significant across the first three trials, $G^2(4) = 4.55$, but it does reach statistical significance on the last three trials, $G^2(4) = 34.36$, $p < .001$. In contrast, the differences in parameter r are statistically reliable for the first three trials, $G^2(4) = 15.46$, $p < .01$, as well as the last three, $G^2(4) = 35.04$, $p < .001$. In general, even though the estimates of c and r are lower for the schizophrenics, the effect is larger and occurs at earlier trials for the retrieval parameter r.

It is also possible to compare the schizophrenics and their controls on the rate of error reduction across trials. This is done by examining the rate parameters (β_i) for both storage and retrieval and by testing to see if those parameters significantly differ between the two groups. For this analysis, we allowed the rate parameters to differ across trials (as opposed to using a constant value of β from the Bush–Mosteller model in Equation 3). The analysis reveals that, in addition to the general deficits in storage and retrieval, schizophrenics differ from the controls in their rate of error reduction. For parameter c, schizophrenics tend to increase their storage at a higher rate than the controls for the early trials (with higher values of $\hat{\beta}_i$) but then increase their storage at a slower rate for the later trials (lower values of $\hat{\beta}_i$). These differ-

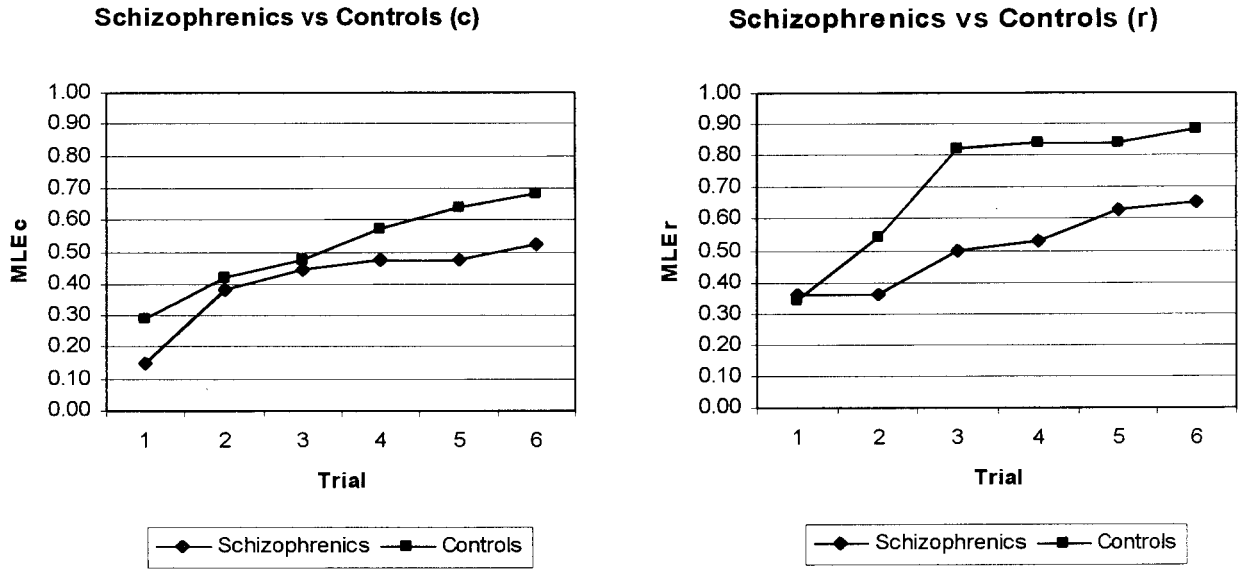

Figure 4. Estimates of *c* (left panel) and *r* (right panel) for the schizophrenics and their controls. MLE = maximum likelihood estimate.

ences between the β_i for the two groups proved to be statistically reliable, $G^2(5) = 11.53$, $p < .05$. For parameter *r*, the increase in retrieval over trials occurs at a significantly slower rate for the schizophrenics than the controls, $G^2(5) = 13.32$, $p < .05$.

Experiment 4: Memory Deficits in Alcoholics With Organic Brain Damage

Severe and prolonged alcoholism can result in impairment of memory and other cognitive functioning. In its worse form, it can even result in organic brain damage, as is the case with alcoholic Korsakoff (AK) syndrome. Alcoholics with organic brain damage exhibit a wide range of memory deficits, depending on the area of the brain that is affected. The most notable deficit is anterograde amnesia, which is characterized by a severe impairment in the capacity to commit new information to memory. Repeated exposure to stimuli often results in a relatively flat learning curve, and delayed testing typically results in rapid forgetting of information (see Butters et al., 1995; Carlen & Menzano, 1995; and Kopelman, 1998, for more detailed reviews).

Much research has been conducted to determine what specific cognitive processing is damaged in alcoholics with organic brain damage, and a portion of this effort has focused on storage and retrieval processes (e.g., Warrington & Weiskrantz, 1970). Unfortunately, the empirical data on this issue are not always clear-cut or consistent. Similar to the research with schizophrenics, a few studies have examined the improvement in memory for AK patients when recognition is compared with recall. Some researchers (e.g., Butters et al., 1988; Warrington & Weiskrantz, 1970) have found that recognition memory for AK patients is not significantly better than recall memory. This would be an indication of a storage deficit that is evident even when the original stimuli are presented as cues. However, other studies (Cermak, Butters, & Goodglass, 1971; Kopelman, 1989; Kopelman, Stanhope, & Kingsley, 1999) have shown that a recognition test can improve the memory of AK patients and that this improvement is greater than that shown by normal controls. Kopelman et al. (1999) have used this result to argue for an underlying retrieval deficit in AK patients.

Another line of research has explored whether alcoholics with organic brain damage can benefit from cues during learning or recall (Marslen-Wilson & Teuber, 1975). Cermak and Butters (1973) found in a series of studies that semantic cuing, such as category names, generally does not improve the memory of AK patients. From this, they concluded that AK patients suffer from an encoding deficit; specifically, these patients usually engage in low-level acoustic or nonsemantic processing of stimuli, although they are capable of semantic processing under certain conditions. This theory is supported by studies using the release from proactive inhibition (PI) paradigm (Wickens, 1970). This is a memory test in which PI is established for a certain class of stimuli over repeated presentations, followed by a release from PI initiated by an item from a different class on a final presentation. Cermak, Butters, and Moreines (1974) and Freedman and Cermak (1986) have shown that AK patients show little or no release from PI when words from one category are shifted to a different taxonomic category. This lack of a release with semantically related stimuli has been cited by some (e.g., Freedman & Cermak, 1986) as evidence that AK patients exhibit a deficit in semantic encoding.

A number of other studies (e.g., Kopelman et al., 1999; Warrington & Weiskrantz, 1970) have demonstrated that AK patients can reliably benefit from cued recall, which indicates an impairment of retrieval processes. For example, Jaffe and Katz (1975) examined a single clinical patient and found that category cues can benefit recall when they are provided at both study and recall. They also reported that category clustering was at chance levels for this patient and that the level of clustering did not improve over trials without cues but did improve when cues were provided. McDowall (1979) observed that AK patients who were instructed to encode semantically improved their memory performance as

compared with AK patients who received no instructions but that category cues that were provided during recall improved the memory of all participants, even those who did not receive initial instructions to encode semantically. From this result, McDowall (1979) concluded that AK patients are capable of encoding semantically but are impaired in their ability to generate appropriate retrieval cues during recall.

Another empirical measure, used by some researchers to differentiate between storage and retrieval factors in clinical populations, is rate of forgetting. As mentioned in the previous section, Calev et al. (1983) observed different forgetting curves for schizophrenics and controls and viewed this as evidence for a postencoding deficit. Huppert and Piercy (1977) applied this same technique to AK patients. In this paradigm, clinical patients receive longer stimulus exposures so that their recognition performance is the same as that for controls after a 10-min delay. Once memory levels have been equated, various delayed recall tests are given so the forgetting curves for each group can be compared. In a series of studies, Huppert and Piercy (1977, 1978, 1979) found that the rate of forgetting was the same for AK patients and normals. From this, they concluded that the memory impairment for AK patients was exclusively at the encoding stage; this conclusion also was reached by Squire (1981). However, Martone, Butters, and Trauner (1986) observed this same pattern of results for AK patients and patients with Huntington's disease and, in a different interpretation, concluded that their results were consistent with a deficit in retrieval not storage.

The previously discussed review demonstrates that a number of model-free measures have been used in an effort to separate storage from retrieval factors in the memory deficits exhibited by alcoholics with organic brain damage. It is clear from this review, however, that the results from these empirical analyses have not been entirely consistent nor have theoreticians always agreed on the interpretation of the results. This brings us back to our central point: Without a model-based method for measuring underlying cognitive processes, the use of ad hoc empirical statistics alone can result in findings that are inconsistent or difficult to interpret.

Experiment 4 uses the pair-clustering model to separately measure storage and retrieval factors in the memory of alcoholics with organic brain damage. Experiment 4 examines the memory of patients suffering from brain damage due to prolonged alcoholism as compared with a control group of alcoholic patients whose diagnosis does not indicate brain damage. As with Experiment 3, the participants attempted to memorize a list of clusterable pairs for a series of six study–test trials. Through the application of the pair-clustering model, it should be possible to examine the storage and retrieval differences between these two groups and to determine how storage and retrieval capacity develops across repeated trials.

Method. The word list and the procedure were the same as those used in Experiment 3. Participants were male patients at Veterans Administration Medical Center in St. Cloud, Minnesota, who volunteered to be in the experiment. The 21 alcoholics with organic brain damage had a staff diagnosis of organic brain syndrome associated with alcoholism and met the corresponding criteria established by Feighner et al. (1972). The 21 alcoholic controls had a staff diagnosis of some form of alcoholism, but none of the controls were suspected of organic brain syndrome. Both the alcoholics with organic brain damage and their controls were selected so that they were matched on mean age and level of education. All participants were sober at the time of testing.

Results and discussion. An ANOVA, with the proportion of words correctly recalled as the dependent measure, showed that the alcoholics with organic brain damage recalled significantly fewer words than the alcoholic controls, $F(1, 40) = 40.44$, $p < .001$, $\omega^2 = .34$. There was a significant increase in performance for both groups over trials, $F(5, 200) = 163.52$, $p < .001$, $\omega^2 = .19$, as well as a significant Group × Trial interaction, $F(5, 200) = 40.32$, $p < .001$, $\omega^2 = .05$. The interaction indicated that, although the alcoholics with organic brain damage improved their performance over trials, they did so at a slower rate than the controls.

Table 4 presents the N_i statistics across the six study–test trials for both groups. We input these data into GPT.EXE to estimate the values of c and r for each condition and used the version of the model without singletons and under order constraints across trials. Figure 5 presents the maximum likelihood estimates for c and r, and Table 4 presents the standard errors for these estimates derived from the computer simulations. As can be seen from Figure 5, the alcoholics with organic brain damage exhibit deficits of both storage and retrieval as compared with the controls. Similar to the results of Experiment 3, these differences tend to be stronger for the later trials. In particular, the differences in storage are not significantly different over the first three trials, $G^2(4) = 3.02$, but are significant for the last three trials, $G^2(4) = 27.83$, $p < .001$. However, the differences between the groups on retrieval are significant for the first three trials, $G^2(4) = 19.52$, $p < .001$, as well as the last three, $G^2(4) = 27.08$, $p < .001$. Also similar to Experiment 3, the size of the differences is larger for retrieval than for storage.

What is particularly interesting about the results in Figure 5 is the performance for the two groups across the six study–test

Table 4
Recall Statistics and Parameter Estimates for the Alcoholics With Organic Brain Damage and Their Controls in Experiment 4

Trial	N_1	N_2	N_3	N_4	c	r	u
				Organic alcoholics			
1	20	9	91	300	.22 (.15)	.21 (.37)	.17 (.04)
2	34	18	102	266	.37 (.12)	.22 (.18)	.26 (.05)
3	43	30	102	245	.44 (.08)	.23 (.08)	.35 (.05)
4	57	25	114	224	.44 (.08)	.30 (.08)	.35 (.05)
5	58	33	100	229	.47 (.07)	.30 (.06)	.38 (.05)
6	65	29	100	226	.49 (.07)	.31 (.06)	.38 (.05)
				Controls			
1	45	24	97	254	.48 (.08)	.22 (.07)	.33 (.05)
2	106	41	107	166	.49 (.05)	.52 (.07)	.44 (.05)
3	171	40	110	99	.51 (.05)	.80 (.07)	.46 (.05)
4	202	50	79	89	.61 (.04)	.80 (.04)	.55 (.05)
5	217	64	69	70	.64 (.03)	.81 (.03)	.65 (.04)
6	243	64	58	55	.68 (.03)	.85 (.03)	.69 (.04)

Note. Numbers in parentheses are the standard errors for each parameter. N_1 = both items recalled consecutively; N_2 = both items recalled not consecutively; N_3 = one and only one item recalled; N_4 = neither item recalled; c = probability of forming and storing a cluster; r = probability of retrieving a cluster, if stored; u = probability of recalling a nonclustered item as a singleton.

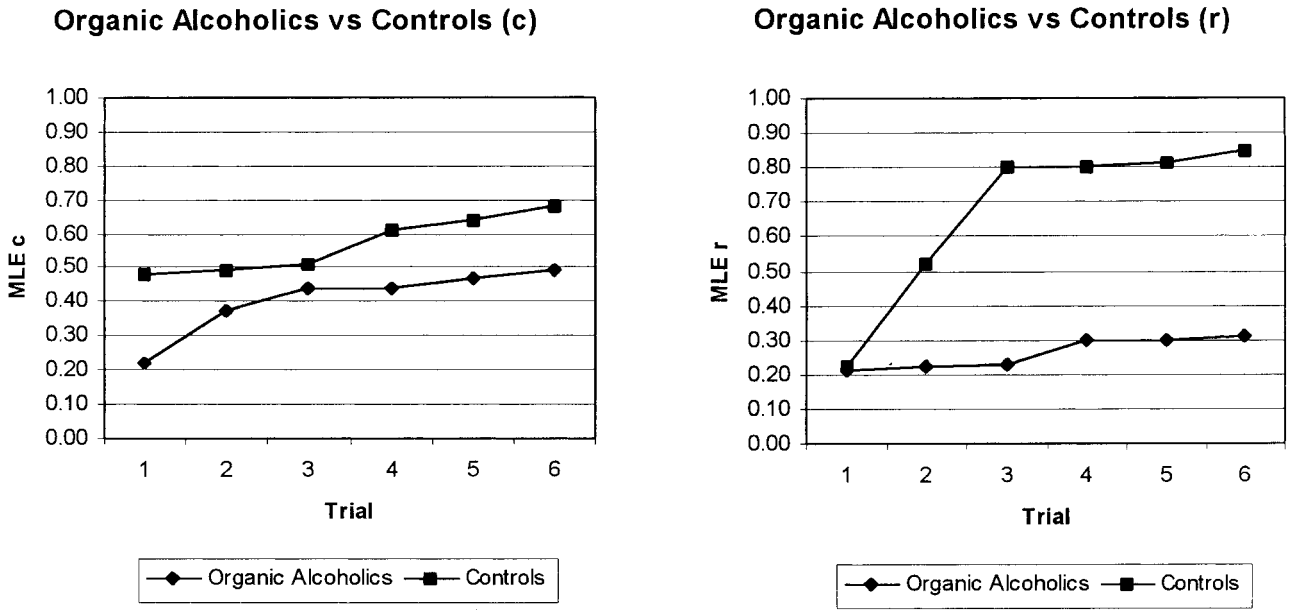

Figure 5. Estimates of *c* (left panel) and *r* (right panel) for the alcoholics with organic brain damage and their controls. MLE = maximum likelihood estimate.

trials. There is a significant increase across trials for both *c*, $G^2(10) = 35.43$, $p < .001$, and *r*, $G^2(10) = 24.06$, $p < .01$, but a more revealing pattern of results emerges when these changes across trials are examined separately for the alcoholics with organic brain damage and their controls. For parameter *c*, there is no significant difference in the rate parameters (β_i) between the alcoholics with organic brain damage and their controls, $G^2(5) = 7.83$, which indicates that, even though the alcoholics with organic brain damage exhibit an absolute deficit in storage, their rate of improvement over trials is equivalent to that of the controls. In contrast, the alcoholics with organic brain damage and their controls show a significant difference in their rate parameters for the retrieval parameter *r*, $G^2(5) = 20.47$, $p < .01$. In fact, even though the control participants exhibit substantial increases in $\hat{r}$ across trials, the curve for the alcoholics with organic brain damage appears to be approximately flat. It is possible to verify this by statistical testing to see whether the retrieval rate parameters for alcoholics with organic brain damage are significantly different from 1.00. This hypothesis was not confirmed, $G^2(5) = 4.58$, which indicates no significant improvement in retrieval capacity across trials for the alcoholics with organic brain damage. Thus, despite repeated presentation and recall attempts for the list of words, the alcoholics with organic brain damage exhibit no substantial improvement in their ability to retrieve information from memory.

General Discussion

Effects of Multiple Trials

Experiments 3 and 4 involved clinical patients who were tested on their memory over six study–test trials. It is unusual in clinical test batteries for a particular memory task to undergo as many as six trials on the same list of words. For example, the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) free-recall task includes only three study–test cycles, although these are followed by two additional test trials. In addition, most studies on schizophrenic or AK patients typically entail one presentation and test for any specific list of items. Consequently, it is reasonable to ask what conclusions might have been drawn if the experiments had involved fewer trials. To assess this, we focused on the first three trials in Experiments 3 and 4. For both experiments, there was no significant difference between the clinical groups and their controls for the storage parameter *c* across the first three trials. Instead, these differences were more pronounced across the last three trials. On the other hand, the retrieval parameter *r* differed for the clinical groups and their controls across all six study–test trials. Indeed, if these experiments had only tested memory across three study–test trials, they would have provided converging evidence that the cognitive deficits were produced by pronounced differences in retrieval and not storage.

It is interesting to speculate why significant differences in cluster storage occurred strongly in the latter three trials but not in the first three. The most likely explanation may be that (in a six-trial experiment) a short attention span, perhaps coupled with fatigue or boredom, could become a factor in the later trials. If this possibility is correct, then the fatigue factor would have to affect the clinical groups more strongly than their controls. This explanation is only conjecture, and further work is needed to validate it and, if true, determine its specific characteristics. In any event, Experiments 3 and 4 strongly support retrieval deficits in the clinical populations relative to their controls.

Individual Differences

An important issue in the application of any MPT model is the question of how robust are the results of the model when there are individual differences in its parameters. To explore this issue, we used computer simulations to analyze the robustness of the pair-clustering model under the assumption of individual differences in parameters *c*, *r*, and *u*. We examined the effect of parameter variation in the two experiments involving the clinical groups (Experiments 3 and 4), because it is commonly acknowledged that individual differences are an important concern when clinical populations are being assessed. This analysis utilized one of the options in the Monte Carlo component of GPT.EXE to simulate data from the model in which observations are determined by parameter values that are drawn from distributions designed to model individual differences. One distribution that is commonly used for this purpose is the beta distribution; Batchelder (1998) described how to incorporate beta distributions into MPT models (see also Batchelder & Riefer, 1999, and Riefer & Batchelder, 1991b, for more detail). The first step in this process is the creation of independent beta distributions for the parameters *c*, *r*, and *u*. Next, instead of generating each category observation with constant parameter values, the parameters underlying each observation are drawn randomly from their corresponding beta distributions. Finally, the category observations are generated from the MPT model represented in Figure 1. This process is repeated for every observation in each of the simulated data sets.

The Monte Carlo runs with parameter variability were conducted in exactly the same way as the computer simulations described in the introduction with one exception. The maximum likelihood estimates of the parameters from each experiment were used as starting values in the simulations, but they were set to the mean of the corresponding beta distribution for each parameter. The simulations added variability to the betas, which creates the assumption that there is variation in the sampled parameters underlying each of the series of observations in a Monte Carlo run. In Experiments 3 and 4, we set the standard deviation of the sampling distribution to a value of 0.10. Riefer and Batchelder (1991b) argued that this is a moderate amount of variability, as might be expected to occur with natural individual differences in participants and items.

For Experiments 3 and 4, we conducted 1,000 Monte Carlo runs with individual differences in the parameters to compare the results with the 1,000 runs without individual differences. This allowed us to see if the parameter estimates and standard errors obtained with each method were substantially different, and it enabled us to see if the conclusions drawn from an analysis of the data with the model were affected by individual differences. We conducted the simulations with independent observations and statistically independent parameter distributions, although it may be of interest in future work to explore what effect dependencies and correlations between the parameters might have on the parameter estimates and their standard errors.

For the simulations without individual differences, the average parameter values from the simulations were generally equal to or close to the parameter values input to the simulations from each experiment (i.e., within .02). When we conducted additional simulations with individual differences in the parameter values, these results did not substantially differ. In other words, estimates and standard errors for the parameters were generally stable even under conditions of moderate parameter variability. These results speak well for the stability and accuracy of the pair-clustering model as applied to the experiments in this article. The only exceptions to this trend were three occasions when the mean of the estimates of *r* from the simulations was substantially larger than the original estimate (with deviations ranging from .17 to .23). These typically occurred during early trials in the clinical groups, when the values for *c* and *r* were very low. This finding matches the results from a series of computer simulations conducted by Riefer and Batchelder (1991b). They observed that the variance of the estimators for *r* depends on the value of *c*. In particular, low values of *c* tend to lead to less stable estimators of *r* (i.e., estimates with higher variance).

This illustrates an important point that practitioners need to keep in mind when using MPT models. Many of the parameters in these models represent conditional processes in the sense that they are relevant only after other processes occur. In the case of the pair-clustering model, *r* is such a parameter because it represents the conditional probability of successful retrieval given successful storage. In general, the standard errors of the estimators for such parameters depend on the overall values of the parameters for the processes on which they are conditioned. This can be seen in the measures of standard error for *r*, some of which were quite large. An examination of Tables 1, 3, and 4 reveals that this tended to occur for early trials, where the estimated value of *c* was small. A low value of *c*, in essence, reduces the effective number of observations that enter into the estimation of parameter *r*. Riefer and Batchelder (1991b) have generally raised this point elsewhere, but it is a concern that is especially relevant when the pair-clustering model is applied to clinical populations. Such populations will often be poor at storage (as well as other processes), which will tend to negatively impact the stability of the maximum likelihood estimates for parameter *r*.

One solution is to increase the number of data observations, *N*. The most straightforward way of doing this is by collecting data from more participants. Typically, this will involve looking at the aggregated data combined across participants, which was the procedure we used in Experiments 1–4. It should be noted, however, that information about individual differences will be unavailable when data are pooled across participants. Especially in the case of clinical populations, there may be a cost to psychological assessment when such information is lost. Carter, Neufeld, and Benn (1998, p. 392) have discussed the potential problems of aggregating data in clinical settings when there are individual differences in patient populations, and they describe how this issue can be addressed through the use of probability mixture models.

Another option for addressing small sample sizes is the collection of enough data from each participant so that each individual can be analyzed separately. Batchelder (1998) described some of the procedures for using MPT models for individual participant assessment, and we have also attempted to meet some of the concerns of individual differences with the computer simulations involving the independent beta distributions. One straightforward way to collect additional data from participants is to measure performance on repeated study–test trials. As seen in Experiments 3 and 4, parameter estimates tended to be more stable for later trials, which resulted in more statistically reliable differences between the clinical groups and their controls.

Conclusion

The goal of this article has been to demonstrate how cognitive psychometrics can be used to assess underlying cognitive deficits in clinical populations. The basic approach takes the form of a substantive mathematical model that represents hypothesized latent cognitive subprocesses as parameters of the model. If the model provides an adequate approximation to how these subprocesses combine to yield manifest performance data, then the estimates of the parameters become measures of the capacity of each subprocess. In this article, we applied this technique to a pair-clustering model that is capable of taking free-recall data from clinical and control participants and using that data to separately measure storage and retrieval subcomponents. The processes of storage and retrieval are each represented by different parameters of the model, and statistically significant differences in the parameter estimates between clinical groups and their controls provide an indication of where cognitive deficits occur.

The pair-clustering model was used to explore storage and retrieval deficits in both schizophrenics and alcoholics with organic brain damage. In general, the results help to explain the diverse findings from prior studies on recall deficits in both of these clinical groups. According to the model's analysis, the schizophrenics and alcoholics with organic brain damage suffer from some impairment of both storage and retrieval capacity. In addition, these deficits tend to be larger for retrieval than for storage and are stronger over later study–test trials. In Experiments 3 and 4, significant storage differences only emerge in the later trials. Given these results, it is not surprising that different studies reach different conclusions regarding the storage–retrieval mechanisms behind the cognitive deficits in these populations. Because some type of deficit seems to occur for both storage and retrieval, final conclusions from empirical studies could depend on the specific memory task that is used or the particular empirical statistics that are used to measure memory. In addition, different conclusions can be reached depending on whether performance is measured on only one trial or over repeated study–test trials.

The differences between storage and retrieval are especially striking for the alcoholics with organic brain damage. Although they showed a deficit in storage capacity, the model revealed that their improvement in storage over trials, measured in terms of error reduction in the rate parameters, did not significantly differ from that of the controls. A very different story occurred for retrieval; alcoholics with organic brain damage exhibited no significant improvement in retrieval over trials. This is an important result, because it parallels previous findings that the learning curve for alcoholics with organic brain damage is flat despite repeated presentations of the same material (e.g., Butters et al., 1995; Jaffe & Katz, 1975). However, these prior results were based on the use of standard empirical measures, which give no indication as to whether the deficit is due to storage or retrieval failure. The results of the pair-clustering model suggest that this learning deficit is due mainly to an inability of patients to develop retrieval strategies and not necessarily to an impairment in storage capacity.

This last result provides a good illustration of one of our main points: Standard, ad hoc empirical measures are limited in their ability to separately measure underlying mental processes. Not only do these measures often confound different processes, but inferences drawn from global patterns of empirical results can potentially result in the wrong conclusions. A good example of this problem can be found in the separate studies by Barker (1977) and Brebion et al. (1997) described earlier. Both examined the benefits of cued recall for improving the memory of schizophrenics and normal controls. Barker observed in one condition that cued recall was significantly greater than free recall for schizophrenics but not for normal controls and concluded from this interaction that recall deficits were due to retrieval and not storage. Brebion et al. (1997) found no interaction in their study between type of patient (schizophrenic vs. normal) and type of recall (free vs. cued) and concluded that the recall deficits were due to storage and not retrieval. Riefer and Rouder (1992) have specifically warned about the potential risk of drawing conclusions about storage and retrieval based on the contrast between free and cued recall. Many theorists have pointed out the risks of relying on interactions to examine cognitive deficits in clinical populations (see Strauss, 2001, for a review).

The specific problem with using interactions between experimental variables to make conclusions about storage and retrieval can be illustrated with the following example. Suppose we assume a simple model in which recall is a function of storage and retrieval processes, with s = the probability of adequate storage, r = the conditional probability of retrieval given storage, and Pr(correct recall) = sr. Suppose also that we conduct a study comparing free and cued recall of schizophrenics versus normal controls, in which it is assumed that schizophrenics exhibit both storage and retrieval deficits. Specifically, let $s = .6$ and $r = .7$ for the normal controls and $s = .3$ and $r = .4$ for the schizophrenics. It is easy to see that the probability of free recall will be .42 and .12, respectively, for the two groups. Suppose we also give each group a cued-recall test and assume that when cues are given during recall retrieval of an adequately stored item occurs with certainty (i.e., $r = 1$). It is straightforward to see that the probability of correct cued recall will equal .6 and .3 for the normal controls and schizophrenics, respectively. This represents an increase in recall rate of .18 for both groups; in other words, there would be no interaction between clinical group and type of recall. Based on the standard interpretation used by Brebion et al. (1997), this would indicate that the initial memory deficit for schizophrenics was due to problems with storage and not retrieval, even though this example was created with the explicit specification that impairment exists in both processes.

The aforementioned demonstration illustrates the advantage that valid mathematical models have over standard, empirically based approaches for the measurement of underlying cognitive deficits. MPT models in particular have an added advantage as measurement tools, due to their relative mathematical simplicity and the availability of computer programs that can perform statistical inference. Because of this, a number of basic statistical techniques are easy to implement using these models, including hypothesis testing and the computation of standard errors for the parameters. The structure of MPT models allows one to conduct computer simulations. We were able to conduct simulations of the model for each of the four experiments reported here by using a special subroutine within the Hu (1999) computer program. Among other things, we were able to use these simulations to derive measures of the variability for all of the parameter estimates, which allowed us to compute the standard error for each estimate.

We think that the clinical studies in this article shed some light on the deficits found in alcoholics with organic brain damage and patients diagnosed with schizophrenia. However, our larger goal has been to illustrate in concrete settings the dynamics of cognitive psychometrics. We demonstrated how a simple but substantive memory model can be validated, how it can be extended for a multitrial situation with minimal theoretical additions, how failure rates for specific cognitive subprocesses can be measured using statistical theory, and how the model can be extended to include parameter variability designed to reflect individual differences in a group of participants. Mathematical modeling has been an active area in cognitive psychology since the 1950s; however, with a few exceptions, little of this work has been directed toward the assessment of cognitive deficits in clinical populations. It is our belief that mathematical modeling of the sort we describe can have a lasting role in psychological assessment.

References

Bamber, D. (1979). State-trace analysis: A method of testing simple theories of causation. *Journal of Mathematical Psychology, 19,* 137–181.

Barker, W. J. (1977). The role of retrieval in schizophrenic memory deficit. *Canadian Journal of Behavioral Science, 9,* 175–186.

Basden, D. R., & Basden, B. H. (1995). Some tests on the strategy disruption interpretation of part-list cuing inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1656–1669.

Batchelder, W. H. (1998). Multinomial processing tree models and psychological assessment. *Psychological Assessment, 10,* 331–344.

Batchelder, W. H. (1999). Multinomial processing tree models of cognition. In W. Hacker & M. Rinck (Eds.), *Bericht über den 41. Kongress der Deutschen Gessellschaft für Psychologie in Dresden* (pp. 65–73). Lengerich, Germany: Pabst Science.

Batchelder, W. H., Chosak-Reiter, J., Shankle, W. R., & Dick, M. B. (1997). A multinomial modeling analysis of memory deficits in Alzheimer's disease and vascular dementia. *Journal of Gerontology: Psychological Sciences, 52,* 206–215.

Batchelder, W. H., & Riefer, D. M. (1980). Separation of storage and retrieval factors in free recall of clusterable pairs. *Psychological Review, 87,* 375–397.

Batchelder, W. H., & Riefer, D. M. (1986). The statistical analysis of a model for storage and retrieval processes in human memory. *British Journal of Mathematical and Statistical Psychology, 39,* 120–149.

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review, 97,* 548–564.

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review, 6,* 57–86.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monographs, 80*(3, Pt. 2).

Bauman, E., & Kolisnyk, E. (1976). Interference effects in schizophrenic short-term memory. *Journal of Abnormal Psychology, 85,* 303–308.

Bayen, U. J. (1990). *Zur lokalisation von altersdifferenzen im episodischen gedaechtnis erwachsener* [The localization of adults' age differences in episodic memory] (Report No. 16-2). Bonn, Germany: University of Bonn, Psychological Institute.

Bayen, U. J., & Murnane, K. (1996). Aging and the use of perceptual and temporal information in source monitoring tasks. *Psychology and Aging, 11,* 293–303.

Beatty, W. W., Jocic, Z., Monson, N., & Staton, R. D. (1993). Memory and frontal lobe dysfunction in schizophrenia and schizo-affective disorder. *Journal of Nervous and Mental Disease, 181,* 448–453.

Bender, R. H., Wallsten, T. S., & Ornstein, P. A. (1996). Age differences in encoding and retrieval details of a pediatric examination. *Psychonomic Bulletin & Review, 3,* 188–198.

Brebion, G., Amador, X., Smith, M. J., & Gorman, J. M. (1997). Mechanisms underlying memory impairment in schizophrenia. *Psychological Medicine, 27,* 383–393.

Bush, R. R., & Mosteller, F. (1955). *Stochastic models of learning.* New York: Wiley.

Butters, N., Delis, D. C., & Lucas, J. A. (1995). Clinical assessment of memory disorders in amnesia and dementia. *Annual Review of Psychology, 46,* 493–523.

Butters, N., Salmon, D. P., Cullum, C. M., Cairns, P., Troster, A. I., & Jacobs, D. (1988). Differentiation of amnesic and demented patients with the Wechsler Memory Scale—Revised. *Clinical Neuropsychology, 2,* 133–148.

Calev, A. (1984a). Recall and recognition in chronic non-demented schizophrenics: Use of matched tasks. *Journal of Abnormal Psychology, 93,* 172–177.

Calev, A. (1984b). Recall and recognition in mildly disturbed schizophrenics: The use of matched tasks. *Psychological Medicine, 14,* 425–429.

Calev, A., Venables, P. H., & Monk, A. F. (1983). Evidence for distinct verbal memory pathologies in severely and mildly disturbed schizophrenics. *Schizophrenia Bulletin, 9,* 247–264.

Carlen, P. L., & Menzano, E. (1995). Neurologic pathology of alcohol abuse. In B. Tabakoff & P. L. Hoffman (Eds.), *Biological aspects of alcoholism* (pp. 245–267). Seattle, WA: Hogrefe & Huber.

Carter, J. R., Neufeld, R. W. J., & Benn, K. (1998). Application of process models in assessment psychology: Potential assets and challenges. *Psychological Assessment, 10,* 379–395.

Cermak, L. S., & Butters, N. (1973). Information processing deficits of alcoholic Korsakoff patients. *Quarterly Journal of Studies on Alcohol, 34,* 1110–1132.

Cermak, L. S., Butters, N., & Goodglass, H. (1971). The extent of memory loss in Korsakoff patients. *Neuropsychologia, 9,* 307–315.

Cermak, L. S., Butters, N., & Moreines, J. (1974). Some analyses of the verbal encoding deficit of alcoholic Korsakoff patients. *Brain and Language, 1,* 141–150.

Chechile, R. A., & Roder, B. (1998). Model-based measurement of group differences: An application directed toward understanding the information-processing mechanisms of developmental dyslexia. In S. A. Soraci & W. J. McIlvane (Eds.), *Perspectives on fundamental processes in intellectual functioning: Vol. 1. A survey or research approaches* (pp. 91–112). Norwich, CT: JAI Press.

Chosak-Reiter, J. (2000). Measuring cognitive processes underlying picture naming in Alzheimer's and cerebrovascular dementia: A general processing tree approach. *Journal of Clinical and Experimental Neuropsychology, 22,* 351–369.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review, 3,* 37–60.

Fazio, B. B. (1998). The effect of presentation rate on serial memory in young children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 41,* 1375–1383.

Feighner, J. P., Robins, E., Guze, S. B., Woodruff, R. A., Jr., Winokur, G., & Munoz, R. (1972). Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry, 26,* 57–63.

Freedman, M., & Cermak, L. S. (1986). Semantic encoding deficits in frontal lobe disease and amnesia. *Brain and Cognition, 5,* 108–114.

Gerrein, J. R., & Chechile, R. A. (1977). Storage and retrieval processes of alcohol-induced amnesia. *Journal of Abnormal Psychology, 86,* 285–294.

Gold, J. M., Randolph, C., Carpenter, C. J., Goldberg, T. E., & Weinberger, D. R. (1992). Forms of memory failure in schizophrenia. *Journal of Abnormal Psychology, 101,* 487–494.

Greeno, J. G., & Bjork, R. A. (1973). Mathematical learning theory and the new "mental forestry." *Annual Review of Psychology, 24,* 81–116.

Gutowski, W. E., & Chechile, R. A. (1987). Encoding, storage, and retrieval components of associative memory deficits of mildly retarded adults. *American Journal of Mental Deficiency, 92,* 85–93.

Heinrichs, R. W., & Zakzanis, K. K. (1998). Neurocognitive deficit in schizophrenia: A quantitative review of the evidence. *Neuropsychology, 12,* 426–445.

Howe, M. L., & Brainerd, C. J. (1989). Development of children's long-term retention. *Developmental Review, 9,* 301–340.

Hu, X. (1999). GPT (Version 1.0) [Computer software]. Memphis, TN: PsyWare.

Hu, X., & Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika, 59,* 21–47.

Hu, X., & Phillips, G. A. (1999). GPT.EXE: A powerful tool for the visualization and analysis of general processing tree models. *Behavior Research Methods, Instruments, and Computers, 31,* 220–234.

Huppert, F. A., & Piercy, M. (1977). Recognition memory in amnesic patients: A defect of acquisition? *Neuropsychologia, 15,* 643–652.

Huppert, F. A., & Piercy, M. (1978). Dissociation between learning and remembering in organic amnesia. *Nature, 275,* 317–318.

Huppert, F. A., & Piercy, M. (1979). Normal and abnormal forgetting in organic amnesia: Effect of locus of lesion. *Cortex, 15,* 385–390.

Hyde, T. S., & Jenkins, J. J. (1969). Differential effects of incidental tasks on the organization of recall of a list of highly associated words. *Journal of Experimental Psychology, 82,* 472–481.

Jacoby, L. L. (1998). Invariance in automatic influences of memory: Towards a user's guide for the process-dissociation procedure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 3–26.

Jaffe, P. G., & Katz, A. N. (1975). Attenuating anterograde amnesia in Korsakoff's psychosis. *Journal of Abnormal Psychology, 84,* 559–562.

Keefe, R. S. E., Arnold, M. C., Bayen, U. J., & Harvey, P. D. (1999). Source monitoring deficits in patients with schizophrenia: A multinomial modeling analysis. *Psychological Medicine, 29,* 903–914.

Kline, S., & Groninger, L. D. (1991). The imagery bizarreness effect as a function of sentence complexity and presentation time. *Bulletin of the Psychonomic Society, 29,* 25–27.

Knapp, B. R., & Batchelder, W. H. (2001). *Representing parametric order constraints in multi-trial applications of multinomial processing tree models* (Mathematical Behavioral Sciences Tech. Rep. No. 01–14). Irvine: Institute for Mathematical Behavioral Sciences, University of California, Irvine.

Knight, R. A., & Silverstein, S. M. (2001). A process-oriented approach for averting confounds resulting from general performance deficiencies in schizophrenia. *Journal of Abnormal Psychology, 110,* 15–30.

Koh, S. D. (1978). Remembering of verbal materials by schizophrenic young adults. In S. Schwartz (Ed.), *Language and cognition in schizophrenia* (pp. 59–69). Hillsdale, NJ: Erlbaum.

Koh, S. D., Kayton, L., & Berry, R. (1973). Mnemonic organization in young non-psychotic schizophrenics. *Journal of Abnormal Psychology, 81,* 299–310.

Koh, S. D., Kayton, L., & Peterson, R. A. (1976). Affective encoding and subsequent remembering in schizophrenic young adults. *Journal of Abnormal Psychology, 85,* 156–166.

Kopelman, M. D. (1989). Remote and autobiographical memory, temporal context memory, and frontal atrophy in Korsakoff and Alzheimer patients. *Neuropsychologia, 27,* 437–460.

Kopelman, M. D. (1998). Clinical and neuropsychological studies of patients with amnesic disorders. In M. A. Ron & A. S. David (Eds.), *Disorders of brain and mind* (pp. 147–174). New York: Cambridge University Press.

Kopelman, M. D., Stanhope, N., & Kingsley, D. (1999). Retrograde amnesia in patients with diencephalic temporal lobe or frontal lesions. *Neuropsychologia, 37,* 939–958.

Kraemer, H. C., Peabody, C. A., Tinkleberg, J. R., & Yesavage, J. A. (1983). Mathematical and empirical development of a test of memory for clinical and research use. *Psychological Bulletin, 94,* 367–380.

Larsen, S. F., & Fromholt, P. (1976). Mnemonic organization and free recall in schizophrenia. *Journal of Abnormal Psychology, 85,* 61–65.

Lyons, M. J., Toomey, R., Seidman, L. J., Kremen, W. S., Farone, S. V., & Tsuang, M. T. (1995). Verbal learning and memory in relatives of schizophrenics: Preliminary findings. *Biological Psychiatry, 37,* 750–753.

MacMillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide.* New York: Cambridge University Press.

Marslen-Wilson, W. D., & Teuber, H. L. (1975). Memory for remote events in anterograde amnesia: Recognition of public figures from newsphotographs. *Neuropsychologia, 13,* 353–364.

Martone, M., Butters, N., & Trauner, D. (1986). Some analyses of forgetting of pictorial material in amnesic and demented patients. *Journal of Clinical and Experimental Neuropsychology, 8,* 161–178.

McClain, L. (1983). Encoding and retrieval in schizophrenics' free recall. *Journal of Nervous and Mental Disease, 171,* 471–479.

McDowall, J. (1979). Effects of encoding instructions and retrieval cuing on recall in Korsakoff patients. *Memory and Cognition, 7,* 232–239.

Neufeld, R. W. J. (1998). Introduction to the special section on process models in psychological assessment. *Psychological Assessment, 10,* 307–315.

Nickerson, R. S. (1984). Retrieval inhibition from part-list cuing: A persisting enigma in memory research. *Memory and Cognition, 12,* 531–552.

Paulsen, J. S., Heaton, R. K., Sadek, J. R., Perry, W., Delis, D. C., Braff, D., et al. (1995). The nature of learning and memory impairment in schizophrenia. *Journal of the International Neuropsychology Society, 1,* 88–99.

Raaijmakers, J. G. W., & Phaf, R. H. (1999). Part-list cuing revisited: Testing the sampling-bias hypothesis. In C. Izawa (Ed.), *On human memory: Evolution, progress, and reflections on the 30th anniversary of the Atkinson-Shiffrin model* (pp. 87–104). Mahwah, NJ: Erlbaum.

Read, T. R. C., & Cressie, N. A. C. (1988). *Goodness-of-fit statistics for discrete multivariate data.* New York: Springer-Verlag.

Riefer, D. M., & Batchelder, W. H. (1987, November). *Further tests of a model for measuring storage and retrieval.* Paper presented at the 28th Annual Meeting of the Psychonomic Society, Seattle, WA.

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review, 95,* 318–339.

Riefer, D. M., & Batchelder, W. H. (1991a). Age differences in storage and retrieval: A multinomial modeling analysis. *Bulletin of the Psychonomic Society, 29,* 415–418.

Riefer, D. M., & Batchelder, W. H. (1991b). Statistical inference for multinomial processing tree models. In J.-P. Doignon & J.-C. Falmagne (Eds.), *Mathematical psychology: Current developments* (pp. 313–336). Berlin, Germany: Springer-Verlag.

Riefer, D. M., & Rouder, J. N. (1992). A multinomial modeling analysis of the mnemonic benefits of bizarre imagery. *Memory and Cognition, 20,* 601–611.

Schonfield, D., & Robertson, B. A. (1966). Memory storage and aging. *Canadian Journal of Psychology, 20,* 228–236.

Sengel, R. A., & Lovallo, W. R. (1983). Effects of cueing on immediate and recent memory in schizophrenics. *Journal of Nervous and Mental Disease, 171,* 426–430.

Shaps, L. P., & Nilsson, L. G. (1980). Encoding and retrieval operations in relation to age. *Developmental Psychology, 16,* 636–643.

Slamecka, N. J. (1968). An examination of trace storage in free recall. *Journal of Experimental Psychology, 76,* 504–513.

Squire, L. R. (1981). Two forms of human amnesia: An analysis of forgetting. *Journal of Neuroscience, 6,* 635–640.

Stip, E. (1996). Memory impairment in schizophrenia: Perspectives from psychopathology and pharmacotherapy. *Canadian Journal of Psychiatry, 41*(8, Suppl. 2), S27–S34.

Strauss, M. E. (2001). Demonstrating specific cognitive deficits: A psychometric approach. *Journal of Abnormal Psychology, 110,* 6–14.

Traupmann, K. L. (1975). Effects of categorization and imagery on recognition and recall by process and reactive schizophrenics. *Journal of Abnormal Psychology, 84,* 307–314.

Traupmann, K. L. (1980). Encoding processes and memory for categorically related words by schizophrenic patients. *Journal of Abnormal Psychology, 89,* 704–716.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior, 5,* 381–391.

Turkstra, L. S. (1998). The effect of stimulus presentation rate on syntax performance in brain-injured adolescents. *Aphasiology, 12,* 421–433.

Warrington, E. K., & Cipolotti, L. (1996). Word comprehension: The distinction between refractory and storage impairments. *Brain, 119,* 611–625.

Warrington, E. K., & Weiskrantz, L. (1970). Amnesic syndrome: Consolidation or retrieval? *Nature, 228,* 628–630.

Weldon, M. S., & Jackson-Barrett, J. L. (1993). Why do pictures produce priming on the word-fragment completion test? A study of encoding and retrieval factors. *Memory and Cognition, 21,* 519–528.

Wickens, D. E. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review, 77,* 1–15.

Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source-memory judgments: A formal dual-process model and an analysis of receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1415–1434.

Received June 12, 2001
Revision received January 22, 2002
Accepted January 24, 2002 ■