Psychological Assessment
2016, Vol. 28, No. 3, 279–293

1040-3590/16/$12.00 http://dx.doi.org/10.1037/pas0000163

A Cognitive Psychometric Model for the Psychodiagnostic Assessment of Memory-Related Deficits

Gregory E. Alexander
University of California, Irvine

Timothy A. Satalich and W. Rodman Shankle
Medical Care Corporation, Newport Beach, California

William H. Batchelder
University of California, Irvine

Clinical tests used for psychodiagnostic purposes, such as the well-known Alzheimer's Disease Assessment Scale: Cognitive subscale (ADAS-Cog), include a free-recall task. The free-recall task taps into latent cognitive processes associated with learning and memory components of human cognition, any of which might be impaired with the progression of Alzheimer's disease (AD). A Hidden Markov model of free recall is developed to measure latent cognitive processes used during the free-recall task. In return, these *cognitive measurements* give us insight into the degree to which normal cognitive functions are differentially impaired by medical conditions, such as AD and related disorders. The model is used to analyze the free-recall data obtained from healthy elderly participants, participants diagnosed as having mild cognitive impairment, and participants diagnosed with early AD. The model is specified hierarchically to handle item differences because of the serial position curve in free recall, as well as within-group individual differences in participants' recall abilities. Bayesian hierarchical inference is used to estimate the model. The model analysis suggests that the impaired patients have the following: (1) long-term memory encoding deficits, (2) short-term memory (STM) retrieval deficits for all but very short time intervals, (3) poorer transfer into long-term memory for items successfully retrieved from STM, and (4) poorer retention of items encoded into long-term memory after longer delays. Yet, impaired patients appear to have no deficit in immediate recall of encoded words in long-term memory or for very short time intervals in STM.

Keywords: Alzheimer's disease, Bayesian Inference, free recall, Hidden Markov model, mild cognitive impairment, Hierarchical Bayesian model

Supplemental materials: http://dx.doi.org/10.1037/pas0000163.supp

This article was published Online First July 27, 2015.

Gregory E. Alexander, Department of Cognitive Science, University of California, Irvine; Timothy A. Satalich and W. Rodman Shankle, Medical Care Corporation, Newport Beach, California; William H. Batchelder, Department of Cognitive Science, University of California, Irvine.

Timothy A. Satalich is now at the Institute for Mathematical Behavioral Sciences, University of California, Irvine.

Support for this article was by an award from Oak Ridge Institute for Science and Education (ORISE) to William H. Batchelder for work with Dr. Mike Young at Air Force Research Labs at Wright Patterson Air Force Base. In addition, support for this article to Timothy A. Satalich came from the Institute for Mathematical Behavioral Sciences at the University of California, Irvine. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2–0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Health care; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro-Imaging at the University of Southern California. All authors own stock options in Medical Care Corporation, whose Chief Medical Officer is W. Rodman Shankle. There is a pending patent application for the model submitted through Medical Care Corporation that includes Gregory E. Alexander, Timothy A. Satalich, and W. Rodman Shankle. Timothy A. Satalich was a former employee of the Corporation, and Gregory E. Alexander was an occasional consultant for it. Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Correspondence concerning this article should be addressed to William H. Batchelder, Department of Cognitive Science, University of California, Irvine, Irvine, CA 92697. E-mail: whbatche@uci.edu

The most notable symptoms associated with Alzheimer's disease (AD) are the impairment of memory-related cognitive functions (Hodges, Salmon, & Butters, 1992; Nebes, 1992). Often these symptoms go unreported until those suffering from AD are either pressured by family into getting tested or their level of impairment causes disruptions in their daily lives. Unfortunately, by the time the impairment has affected their daily lives, there is little chance of improvement, making early detection of AD much more crucial. To assess memory-related cognitive functions, clinicians have adopted the use of cognitive tests developed by memory researchers. In return, these cognitive tests have given clinicians the opportunity of diagnosing earlier stages of AD, which allow for early interventions that afford patients more control over the progression of AD.

A prominent cognitive test used by medical doctors to measure memory-related deficits is a free-recall task. The design of the task involves a study trial, where words are sequentially presented to the participant, followed by a test trial where the participant is asked to recall as many of the presented words as they can. Despite its simplicity, the free-recall task provides a way to test the strength of a participant's episodic memory for familiar words presented on a study trial. A U-shaped serial position curve is often observed from the recall behavior in a free-recall paradigm, where words presented at the beginning and end of the study list have a higher probability of being recalled than words in the middle of the list (Murdock, 1962). The two peeks of the U-shape are commonly referred to as the *primacy* and *recency* effects, respectively.

Generally, two different cognitive processes are assumed to underlie these effects. For the primacy effect, the increased recall probability is assumed to be because of an additional amount of rehearsal time allotted for encoding into a long-term episodic memory (LTM) system (Rundus, 1971). During this additional time, words presented at the beginning of the study list have fewer competitors to rehearse and encode than words presented later in the study list. In addition, a shared characteristic of words that are successfully encoded into LTM is a recall probability that decays slowly over time. As for the recency effect, words toward the end of the study list do not have the extra time available for rehearsal before the test trial is administered, so the augmentation in their recall probability is thought to be a function of a different system. Namely, words toward the end of the study list may be in a temporary memory system that affords direct access for recall. This system is often referred to as the short-term memory (STM) storage, and words in the STM have recall probabilities that decay rapidly with time. Thus, in the STM, words whose proximity is closest to the test trial are more likely to be recalled than items further away from the test trial. Therefore, the recency effect is thought to be a function of STM (Howard & Kahana, 2001). This interpretation is supported by studies where the period between study and test is occupied with a distracting task (Bjork & Whitten, 1974). In this case, words at the end of the study list do not show a recency effect and may result in lower recall probabilities than words presented earlier in the list.

A widely used test to assess AD-related deficits is the Alzheimer's Disease Assessment Scale: Cognitive subscale (ADAS-Cog; Chu et al., 2000; Graham, Cully, Snow, Massman, & Doody, 2004). The ADAS-Cog includes a free-recall subtest administered to patients as part of their assessment. The goal of this article is to provide a new, model-based assessment method to analyze data in the ADAS-Cog free-recall task. For this purpose, the article develops a cognitively grounded Hidden Markov model (HMM). The next section will review a few operationally defined methods that clinicians have used for analyzing specific latent memory processes in free recall with the progression of AD. This review prompts the need to establish a formal cognitive psychometric model that combines known memory theory to assess the latent processes associated with the free-recall task. To do so, the next section will provide specifications and predictions of our model. Next, a Method section will provide a description of the research design and data gathered from three groups of participants by the Alzheimer's Disease Neuroimaging Initiative (ADNI) using the ADAS-Cog free-recall task. In the same section, estimation theory of our model will be presented. Following these sections, preliminary results will provide evidence showing the need for a modification of the model to further facilitate its use in clinical assessment. Finally, there is a discussion of the results and conclusion.

Clinical Assessment Using Free-Recall Data

An important aspect of using the free-recall paradigm in populations showing memory deficits is the finding that serial position effects are sensitive in differentiating healthy participants and those suffering from dementia (Egli et al., 2014; Howieson et al., 2011). For example, testing patients with AD-related deficits has revealed significant decline in the primacy effect (Capitani, Della Sala, Logie, & Spinnler, 1992; Gibson, 1981). This decline is supported by known LTM deficits associated with the progression of AD and is thought to be because of an impairment of encoding items into LTM. A standard operational method used to measure LTM related processes from the primacy effect is simply to calculate the proportion correct on the first few items in a study list. Researchers are then able to test whether there is a significant difference between healthy and AD participants using conventional statistics, such as the analysis of variance (ANOVA).

While the proportion correct for some items at the beginning of the list has been used as a proxy for LTM strength, other methods have been proposed. A systematic approach using a Selective Reminding Test involves measuring LTM by counting the number of words continually retrieved without further presentation (Buschke, 1973). While this method involves an experimental manipulation different from that of the ADAS-Cog, there have been cognitive models for the Selective Reminding Test (Kraemer, Peabody, Tinklenberg, & Yesavage 1983; Wenger, Negash, Petersen, & Petersen, 2010). Another method of measuring LTM abilities stems from studies of STM on the recency effect. Waugh and Norman (1965) proposed a method that uses performance on the middle words of the list as a proxy for LTM ability, with the assumption that STM processes do not influence the words in the middle serial positions. Regardless of the methodology used, overwhelming evidence for deficits in LTM is reported for patients showing symptomatology of AD (see Carlesimo & Oscar-Berman, 1992, for a review).

The second latent memory process associated with the serial position curve is retrieval from a STM system. Similar to the primacy effect, the recency effect is measured by calculating the proportion correct for a prespecified number of words at the end of the study list (Tulving & Patterson, 1968). Unlike the primacy effect, clear evidence of recency impairment in AD is not always

demonstrated. For example, Martin, Brouwers, Cox, and Fedio (1985) and Miller (1971) reported finding a significant reduction in the recency effect for participants with AD etiology. However, a study by Spinnler, Sala, Bandera, and Baddeley (1988) found normal levels of the recency effect for patients showing signs of AD progression when restricting the analysis of the recency effect to only the last five words. Similarly, Bayley et al. (2000) reported normal recency effects in patients with AD when the analysis only included the last two words. On balance, no definitive conclusion can be made about the decline in the primacy effect in AD.

Other, more sensitive, methods have been proposed to measure the latent memory processes associated with the recency effect. For example, Tulving and Colotla (1970) proposed measuring STM using the performance scores of items with a relatively small distance between the presentation at study and a recall during the test phase. Results using this procedure show comparable STM ability for patients with AD and healthy participants for the last 2–3 items in the study list, with a significant reduction in STM ability for patients with AD on items further away from the test (Carlesimo, Fadda, Sabbadini, & Caltagirone, 1996; Wilson, Bacon, Fox, & Kaszniak, 1983). Methodological differences and severity differences may be the reasons behind the variability in results. However, without a standardized procedure, measurement of these latent memory processes is dependent on the number of words a researcher deems to be part of the primacy or recency effects.

By employing formal cognitive models, cognitive psychologists have focused on modeling latent memory structures and processes to improve clinical measures of free recall (e.g., Batchelder, Chosak-Reiter, Shankle, & Dick, 1997; Brainerd et al., 2014). For instance, Batchelder et al. (1997) developed a cognitive model that identified differing cognitive processes underlying the free-recall task of the Consortium to Establish a Registry for Alzheimer's disease (CERAD; Fillenbaum et al., 2008). Batchelder et al. (1997) demonstrated that it was possible with the model to measure differences between AD and cerebrovascular etiologies using the immediate free-recall portion of the task. Successful applications of other cognitive models for psychodiagnostic purposes are evident with the many publications of articles in special issues of *Psychological Assessment* (Neufeld, 1998), *Journal of Mathematical Psychology* (Neufeld & Townsend, 2010), and chapters in special books on clinical modeling (e.g., Neufeld, 2007).

Although improvement has been made using memory-based measurement tools for clinical assessment, often clinical tests used for psychodiagnostic assessments use different task designs that are at variance from those normally used to study memory processes by experimental psychologists. As discussed in Batchelder (1998), the design of assessment tests, like the ADAS-Cog free-recall task, is structured so that all participants receive exactly the same test and testing procedure. In contrast, psychological experiments generally control for possible confounding variables known to cause spurious results since episodic memory is sensitive to experimental design. For example, to avoid item effects, words are chosen to be unrelated to each other, and their presentation order on any given study trial is randomized over participants. On the other hand, the ADAS-Cog free-recall subtest is similar to many other clinical tests in that every participant receives the same set of 10 words over the same three fixed shuffled-order study trials. The added complexity can be problematic for formal cognitive models attempting to quantify the latent memory processes in the ADAS-Cog free-recall test. Any model attempting to analyze such data would have to distinguish between underlying signals and noise created by experimental conditions that do not control for confounding variables.

One solution often used by clinicians is to assume that the added noise occurring from methodology is constant across all participants and thus the true cognitive ability of a person can be approximated by a statistic of their observed responses (e.g., normally the number of correct recalls). In fact, the manual for ADAS-Cog provides scoring rules that create an aggregate summary score for the free-recall subtest to diagnose patients showing early signs of AD. With the summary score, researchers can then take advantage of statistical models such as ANOVA to analyze the differences between participant groups. In the Results section, we provide a between-group repeated measures ANOVA analysis of ADAS-Cog data (Table 2) on the observed recall behavior to show results using a standard statistical method. Our inclusion of this analysis is designed to point out that summary scores used to quantify behavior on the ADAS-Cog free-recall task not only fail to tap into much of the signal in the data, but also do not measure the latent cognitive processes underlying the behavior of those tested on the ADAS-Cog. Instead of analyzing aggregate performance scores, this article develops and applies a formal modeling-based approach that combines known memory theory for a more complete assessment of latent memory processes associated with the free-recall task.

A HMM for Free Recall

The framework for the model in this article can be traced back to established cognitive models designed for list memory experiments. These memory models stem from the class of models called HMMs, whose structure involves latent (unobservable) cognitive memory states and observable response sequences. Starting in the 1960s, HMMs became a popular approach to cognitive modeling that led to a number of models that successfully fit data in simple memory paradigms, such as paired-associate learning and free recall, (e.g., Greeno & Bjork, 1973; Wickens, 1982). In these memory models, learning is represented as a function of storage and retrieval processes from latent memory encoding states.

In the case of a multitrial memory task such as the free-recall task, an HMM model postulates that on any trial a to-be-remembered item occupies one of a small set of memory states. Associated with each memory state is a retrieval parameter representing the probability of a correct recall for any item occupying that state on a test trial. The role of the study trials is to prompt transitions among the memory states through a network of state-to-state transition probabilities specified in terms of the model parameters. Such a model is called a HMM because the observable recall/not-recall response sequence for an item over test trials does not uniquely identify (hides) the sequence of underlying latent memory states behind the observed response sequence. For example, an error on a test trial could come from any of several memory states. The term Markov comes from a class of stochastic processes where transition probabilities between the states depend only on the current state and not on previous state transitions.

Basic Model Assumptions

The proposed HMM for the ADAS-Cog recall task postulates three memory states corresponding to different levels of episodic memory storage. The first cognitive state is the Unlearned state (U-State) that represents a state where the participant has not yet encoded a word into episodic memory. The second cognitive state is the Intermediate state (I-State). The I-State is analogous to STM. It is a state where a word is encoded at a shallow level, and the probability of retrieval from that state is expected to decrease rapidly since the occurrence of encoding. The third and final state is the Learned state (L-State). The L-State can be thought of as LTM because it represents a state where an item is fully encoded into episodic memory and it is expected that the recall probability of words in the L-State is subject to slow decay.

It is common practice to display a HMM as a graphical representation of nodes and connections between nodes. The nodes represent model states and the directed connections between nodes represent transition probabilities. The model is represented pictorially in Figure 1. The parameter r is the probability that a word in the U-State is encoded into the L-State on any study trial. If some encoding occurred but did not result in a transition into the L-State then, with probability $(1 - r)a$, the word transitions into the I-State. If no transition from the U-State is made into either of these states then, with probability $(1 - r)(1 - a)$, the word remains in the U-State. Now if a word is in the I-State at the start of a study trial it has probability v of making a transition to the L-State, and with probability $(1 - v)$ it remains in the I-State. Finally, once a word is in the L-State, it does not make any further transitions.

The observation recall sequence for the model is the compilation of the recall performance on each item across the four test trials of the free-recall task. These recall events are generated probabilistically as a function of the state an item is in on a test trial. In Figure 1, the recall probabilities are written inside the nodes for each state. If an item is in the U-State on a test trial, there is a zero probability of recall, if in the I-State the recall probability is t for immediate test trials, and if in the L-State it is l. The model as currently specified has five parameters, r, a, v, t, and l_1, that represent various transition and recall probabilities.

Adapting the HMM to the ADAS-Cog Task

To adapt the model in Figure 1 to the ADAS-Cog free-recall task, three important additional specifications of the model are needed. First, if an item is in the L-State on any of the three immediate test trials, the recall probability is l_1; and if an item in the L-State is recalled on a delayed trial (the fourth test trial), it has a probability l_2 of doing so. Having two recall probabilities for the L-State stems from memory research showing that a memory trace in the LTM decays after a delay (e.g., Burgess & Hitch, 2006). Naturally, one would expect that $l_2 < l_1$ because of memory decay during the delay before the fourth test trial. Unlike the L-State, the I-State is a STM system with rapidly decreasing recall probability, so it is assumed that there is a zero probability of recall on the delayed test trial. Thus, the model assumes that only items in the L-State have a chance to be recalled on the delayed test trial.

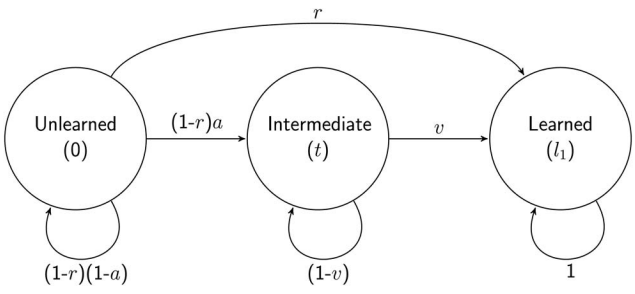

Figure 1. Hidden Markov model with the state-to-state transition and state recall over the three latent memory states. State-to-state transition probabilities are written next to the arrows and recall probabilities are written in the circles that represent states.

Second, in addition to the transitions made possible during the study trial, depicted in Figure 1, there is one other way that a state transition can occur in the model. In particular, if an item is in the I-State on any of the first three study trials and if, with probability t, it is successfully recalled on the following test trial, then a transition to the L-State during the test trial is possible with probability b, and with probability $(1-b)$ the item remains in the I-State. This additional transition parameter represents the possibility of learning during the test trial, which is related to the Testing Effect (Goldstein, 2010), and memory research in paired associate learning has shown that learning can occur during a test trial for both healthy and memory-impaired participants (e.g., Bozoki, Grossman, & Smith, 2006).

The third additional specification of the model concerns how the parameters are tied to the presentation order on a study trial. When models like the one in Figure 1 are applied to many list memory experiments, a convenient assumption has been to apply the cognitive processes envisioned by the model to each item independently and with identical values of the model parameters. The assumption of identical model parameters regardless of the location of an item in the study list is directly inconsistent with known results provided by the serial position curve. As a consequence, our third modification of the model is to adapt the model to the variable study list orders by associating the state-to-state transition parameters and state recall probabilities in Figure 1 with each possible word order position on a study list. Thus, the transition and recall probabilities that apply to any word on a particular study trial or test trial depend on the location of that word in the study order for that trial. Because the serial positions for each word change across the study trials, it follows that any given word will have different state-to-state transition probabilities on each study trial and different recall probabilities on each test trial, depending on its study list position.

This modification means that for each parameter type in the model, with two exceptions, there is an associated set of 10 different parameters, each corresponding to one of the 10 study order positions in the 10-word study list. The two exceptions are on the b parameter (learning on a test trial) and l_2 (delayed recall from the L-State) because neither parameter is tied to the study list order. However, because the parameter b is a measure of possible learning effects of each item on a test trial, the parameter is indexed by item rather than list order. As a consequence, there are also 10 possible values for the parameter b. The purpose of these generalizations of the model in Figure 1 is to allow the storage probabilities and recall probabilities to reflect the cognitive processes behind the serial position curve discussed earlier. For example, we would expect the r parameter for storage in the L-State

in Figure 1 to be higher for a word presented in a position at the beginning of the study list than at the end because of the extra amount of encoding time. In addition, we would expect the recall probability t from the I-State to be higher for words studied at the end of the list because of their proximity to the test phase.

Method

Data used in the preparation of this article come from the ADNI database (adni.loni.usc.edu). ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and nonprofit organizations, as a $60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial MRI, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. For current information, see www.adni-info.org.

Participants

ADNI's first longitudinal study (ADNI-1) recruited 744 participants from 50 sites in the United States and Canada. Participants enrolled in ADNI-1 were between 55 and 90 years of age. Normal control participants ($N = 205$) had no memory complaints, a Mini-Mental State Exam (MMSE; Folstein, Folstein, & McHugh, 1975) of 24–30, a Clinical Dementia Rating (CDR) of zero, nondepressed, non-MCI, and nondemented. MCI participants ($N = 362$) have had a memory complaint by the participant or their partner, MMSE 24–30, objective memory loss based on education adjusted scores on the Wechsler Memory Scale Logical Memory II, a CDR of 0.5, absence of significant levels of impairment in other cognitive domains, preserved activities of daily living, and nondemented as determined by the site physician at the time of screening. Mild AD participants ($N = 177$) had MMSE scores between 20 and 26, CDR of 0.5 or 1.0, and met the criteria established by the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) and the Alzheimer's Disease and Related Disorders Association (ADRDA) known as the NINCDS-ADRDA criteria for probable AD.

Materials and Procedure

ADAS-Cog was developed in 1983 (Mohs, Rosen, & Davis, 1983) and revised in 1997 (Mohs et al., 1997) for trained personnel to assess cognitive functions affected during the dementia stage of AD using a single, aggregate summary score. The 11 subsections used by ADNI-1 (in no particular order) are as follows: (1) orientation to date and time, (2) constructional praxis, (3) following commands, (4) following multistep instructions, (5) object naming, (6) ideational praxis, (7) spoken language, (8) word finding, (9) word recognition, (10) immediate and delayed free recall, and (11) delayed recognition memory. The data obtained from a 10-word list immediate and delayed free-recall tasks is used for the current analysis.

The immediate free-recall task has three study trials where 10 words are presented to the participant one at a time on a white index card. During the first study trial, participants were instructed to read each individual word and to repeat it aloud. After the 10 words were studied, the participants were asked to recall the words just presented to them, in any order, within a 2-min window. The administrator of the ADAS-Cog would then record each word correctly recalled by the participant. This procedure was repeated two more times with the same 10 words but with different presentation orders. Each participant received the same instructions and the same material. See Table 1 for the list of 10 words and their presentation order on the three study trials. After the three study-test trials, two intervening tasks were administered followed by the Delayed Recall task. The two intervening tasks tested the participant's ability to follow commands and constructional praxis. Neither intervening task had any common elements with the word recall task. The Delayed Word Recall task tested the participant's ability to recall words after intervening tasks. The task required that each participant recall the 10 words studied during the three study trials after a delay of approximately 5 min.

HMM Equations

Discrete trial HMMs are traditionally specified with matrices representing state-to-state transition probabilities, an initial starting state probability vector, and a vector representing state-to-observable response probabilities (e.g., Wickens, 1982). For our current model, two transition operators are required for the proposed design. The first transition matrix, $\mathbf{T}$, in Equation 1, designates the possible transitions on a study trial from one of the row states into one of the column states.

$$\mathbf{T} = \begin{array}{c|ccc|cc} & \mathbf{L_{n+1}} & \mathrm{I}_{n+1} & \mathbf{U_{n+1}} & \text{Immediate} & \text{Delay} \\ \hline \mathbf{L_n} & 1 & 0 & 0 & l_{1,i} & l_2 \\ \mathbf{I_n} & v_i & (1-v_i) & 0 & t_i & 0 \\ \mathbf{U_n} & r_i & (1-r_i)a_i & (1-r_i)(1-a_i) & 0 & 0 \end{array} \quad (1)$$

The subscript, i, on the parameters refers to the ith study list position on a study trial, because there are different transition probabilities for each study-list order position. Without the study-list order subscripts, the transition matrix in Equation 1 is just another way of representing the transition probabilities in Figure 1. Note that we have added a column vector of recall probabilities, given the row state, to the right of the transition matrix, $\mathbf{T}$. The two

Table 1
Studied Words in the Order Presented to the Participants on Each of the Three Study Trials

Order	Trial 1	Trial 2	Trial 3
1	Butter	Pole	Shore
2	Arm	Letter	Letter
3	Shore	Butter	Arm
4	Letter	Queen	Cabin
5	Queen	Arm	Pole
6	Cabin	Shore	Ticket
7	Pole	Grass	Engine
8	Ticket	Cabin	Grass
9	Grass	Ticket	Butter
10	Engine	Engine	Queen

recall probabilities in the L-State correspond to the immediate test and the delayed test.

The second state-to-state transition matrix, Γ, found in Equation 2, is used to account for the possibility of learning on a test trial.

$$\Gamma = \begin{array}{c|ccc} & \mathbf{L_{n+1}} & \mathbf{I_{n+1}} & \mathbf{U_{n+1}} \\ \hline \mathbf{L_n} & 1 & 0 & 0 \\ \mathbf{I_n} & t_i b_k & (1 - t_i b_k) & 0 \\ \mathbf{U_n} & 0 & 0 & 1 \end{array} \quad (2)$$

Unlike the first transition matrix, $\mathbf{T}$, which covers transitions on any one of the three study trials, $\mathbf{\Gamma}$ covers possible transitions during any one of the three immediate test trials. Such test-trial transitions are only possible when an item is in the I-State at the beginning of a test trial, and with probability $t_i\ b_k$, the word k at position i transitions into the L-State after the test phase. This probability represents the joint probability of recalling an item located in the I-State and transitioning to the L-State on the test trial. In the proposed HMM, this is the only process that results in learning during the test trial.

To complete the HMM and obtain the equations needed for statistical inference using the likelihood function of the model we designate the initial start vector of state probabilities before the first study trial as $\Lambda_0 = [0, 0, 1]$, which indicates that every item is assumed to be in the U-State before the first study trial. Now, it is a property of the HMM structure that the probabilities of every sequence of observable responses (recall success or failure) can be obtained by suitable matrix operations on the start vector, transition matrices, and recall vector; for example, Wickens (1982). There are 16 possible sequences of observable responses for each of the 10 words. For example, it is possible for a participant to fail to correctly recall a word on all four test trials (0000) or recall that word for all four trials (1111), or anything in between. The observed recall performance sequence for each participant and each word constitutes the basic data that will be used to estimate the parameters of the model.

Hierarchical Bayesian Inference

A central task of psychological assessment focuses on the evaluation of the individual patient's performance or lack thereof. While many cognitive-based models are in the service of studying memory-related functions at the group level (see Batchelder, 1998, for a review), a resolution toward analyzing individual performances is needed for psychodiagnostic assessments. One standard method to augment a statistical model to handle individual differences is to make it hierarchical. This approach makes the assumption that every participant's parameters are a sample from a hierarchical population distribution with its own hierarchical parameters. Given a group of participants classified with similar symptomology, estimation of population level latent variables can shed some light on the disorder. For example, given the three groups of participants in the current study, clinicians may be interested in knowing the average latent ability associated with encoding into a long-term storage system for each stage of memory impairment. In addition, the approach of making the model hierarchical allows the analysis for individual participants within a group.

The use of Bayesian based inference for parameter estimation in hierarchical models is an established practice in statistics, for example, Bernardo and Smith (2009). In the last decade, it has become an increasingly popular approach to estimating hierarchical cognitive models; for example, Lee and Wagenmakers (2013). The advantages over classical likelihood based analysis are both practical and conceptual (Gelman, Carlin, Stern, & Rubin, 2014). From a pragmatic point of view, easily available software such as WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000) and JAGS (Plummer, 2003) allows users to estimate complex cognitive models with relative ease. At the conceptual level, Bayesian statistical inference facilitates the augmentation of formal mathematical models to include hierarchical assumptions to estimate participant parameters. By augmenting the model to include hierarchical assumptions and adopting the Bayesian statistical inference framework, we sidestep many problems posed by classical likelihood methods such as the assumption that the data constitute a large sample of independent and identically distributed observations.

To apply hierarchical Bayesian estimation to the HMM, we associate each parameter with a hierarchical distribution, where individual participant parameters are drawn from these distributions. A popular hierarchical population distribution used by statisticians is the Gaussian distribution with mean μ and precision $1/\sigma^2$ hyperparameters. Of course, values sampled from the Gaussian distribution are on the real line, so for our application, values drawn from a Gaussian distribution will require a transformation to the probability space of (0,1). A common transformation that takes values on the real line to values in probability space is the inverse-probit transformation (Gowans, Fraser, & Hyltoft Petersen, 1989).

For the current application, draws from a distribution for the mean and precision of the Gaussian hierarchical distributions are on the real line and on the positive half, respectively. We selected the hyperprior for the mean, μ, to be normally distributed with the mean and precision set at 0 and 1, respectively. This hyperprior is exactly the distribution of the probit of an uninformative uniform distribution on the probability space (0,1). For the precision hyperparameter, a Gamma distribution is used with scale and shape hyperpriors set at 5 and 5, respectively. The use of uninformative hyperparameter distributions is selected so that before evidence of the data, no parameter value is expected to be more likely than any other value, thus allowing the data to drive the posterior parameter distributions rather than our prior beliefs.

The supplementary material, which is available online, provides the equations for the 16 response patterns for each of the 10 words along with the model likelihood function written in terms of JAG's code. Furthermore, a graphical model representation of the hierarchical HMM is presented in the supplementary section. The analysis of the hierarchical model in JAGS used four chains of 1,000 samples each with a burn-in of 500 samples. A collective total of 2,000 samples were retained for the current analysis. For a detailed explanation of Markov Chain Monte Carlo (MCMC) sampling, see Gelman et al. (2014). In the Results, the reported means over participants of each parameter will be presented in the figures on the natural probability scale rather than on the real line. To obtain the mean of a particular parameter in (0,1), first an inverse probit is taken of each draw for that parameter from the hierarchical Gaussian, then the posterior mean and *SD* of these transformed draws are presented in the figures.

Results

Before analyzing the recall data with the HMM, it is useful to inspect several aspects of the data. Figure 2 provides the group average recall probabilities for each of the 10 words and four test trials in each of the three groups. The most obvious fact about these plots is a decreased performance across the three study groups. In addition, the bar plots for Trial 1 in Figure 2 tend to reveal the expected form of the serial position curve. For all three participant groups, the U-shaped serial position curve is not evident after the first test trial. By changing the presentation order of words in the study list for the second and third trials as shown in Table 1, words may no longer be governed by the same STM and LTM processes affecting them on the first trial. As a consequence, there is no reason to expect the U-shaped serial position curve on those trials.

Although the bar plots for Trial 1 tend to show the expected serial position curve, there are noticeable exceptions, with the largest being in position seven. The word "Pole" in Position 7 on the first study trial has consistently higher recall probabilities than its neighbors for all three study groups. For some reason, "Pole" is more memorial than its neighbors in the context of the particular words and word orders that are fixed for all participants in the ADAS-Cog task. Violations of the expected serial position curve such as for Item 7 are a likely consequence of the fact that the ADAS-Cog task does not counterbalance the assignment of words to study positions. This is an example of one of the difficulties in applying cognitive models to data from clinical assessment batteries. Then, if every participant receives the same set of words in the same order, like in the ADAS-Cog test, specific item effects may become evident such as the word "Pole" in the ADNI data.

Standard ADAS-Cog Analysis

Next, we examine the recall data by employing standard statistical tests on the aspects of the data that are suggested by the ADAS-Cog manual. Table 2 shows the results of a split-plot repeated measures ANOVA (Kirk, 1968, pp. 248–251). The repeated measures factor was the recall trials for the healthy, MCI, and AD groups, and the dependent variable was the number of correctly recalled words. Group and trial main effects and Group × Trial interaction effect significantly influenced word recall performance ($p < .01$), such that increasing severity results in lower numbers of correctly recalled words, especially on the delayed recall trial. The effect size for group was calculated using eta-squared, $\eta^2 = .31$, which by conventional standards is considered large. One difficulty with this type of analysis is with the aggregation of word recall into a single score. Aggregate scores do not offer much insight into the psychological differences underlying the three groups. Although the ANOVA model provides useful information, it does not indicate whether normal aging, MCI or AD differentially influences particular cognitive processes that underlie the performance on different words in the aggregate score.

Participant Heterogeneity

To test participant heterogeneity in light of possible item heterogeneity, a nonparametric Monte Carlo permutation test in Smith and Batchelder (2008) was employed. In that article, their test was applied to free-recall data obtained from a study similar to the ADAS-Cog free-recall task. The test calculates the variance of the participants' performance scores on each of the four test trials. The permutation test obtains a distribution of these variances under permutations of the performance data across participants, and this distribution represents the variability of this statistic under the null

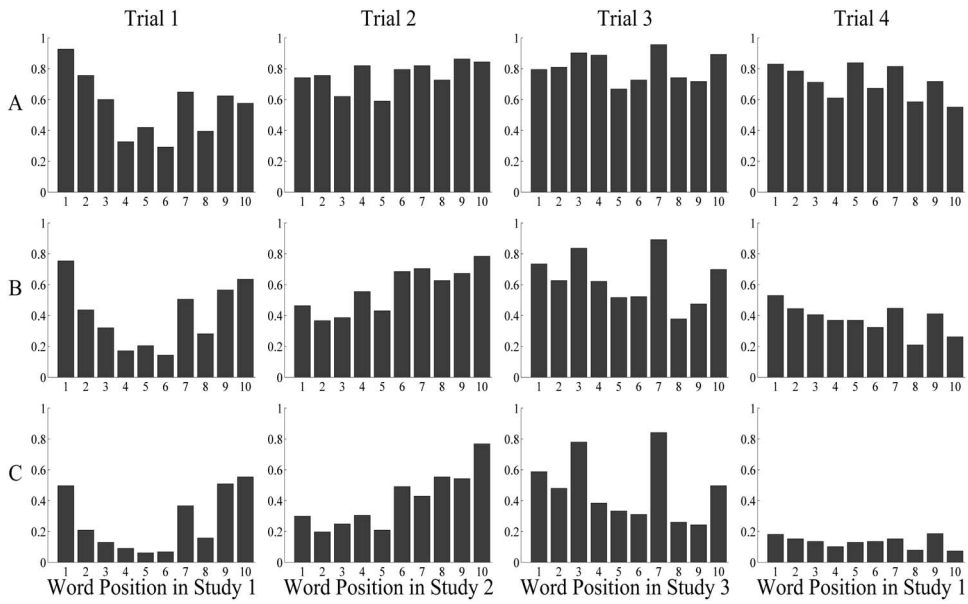

Figure 2. The aggregate recall probability for (A) healthy, (B) mild cognitive impairment (MCI), and (C) Alzheimer's disease (AD) participants for each word over the four trials. The words are positioned to reflect their assignment during the three study trials, and the fourth trial matches the first trial order.

Table 2
Split-Plot Repeated Measures Analysis of Variance (ANOVA) of the Number of Words Recalled in Each of the Four Trials by Impairment Groups

Source	SS	*df*	MS	*F*	Prob > *F*
Between subjects	12,718.29	743	—	—	—
Group	5,878.33	2	2,939.16	318.41	<.01
Subjects w/in groups (error)	6,839.96	741	9.23	—	—
Within subjects	6,266.26	2,232	—	—	—
Trial	3,115.31	3	1,038.44	869.13	<.01
Group × Trial	495.00	6	82.50	69.05	<.01
Trial × Subjects w/in groups (error)	2,655.95	2,223	1.19	—	—
Total	18,984.54	2,975	6.38	—	—

Note. SS = sum of squares; MS = mean squares.

hypothesis of participant homogeneity within a study group. The current application tests the null hypothesis that subject variability for the ADAS-Cog data on each trial is what would be expected from random error. A sample of 100,000 permutations provided a distribution of possible variances of participant's performance scores on each trial under the null hypothesis of participant homogeneity. The observed variance for each group across the 4 trials and the 95-percentile distribution of possible variances under the null hypothesis is provided in Table 3. The null hypothesis of participant homogeneity was rejected because the *p* values were outside the .05-level (2-tailed) for all four trials for all three-study groups, with the exception of the first trial of the healthy participant group. The results indicate that it is important to utilize an estimation method for the HMM that handles random effects on the parameters because of participant heterogeneity within a study group. The hierarchical Bayesian inference discussed earlier is ideal for accomplishing this purpose.

Preliminary Results of the HMM

The presence of several defining characteristics from memory theory was discovered in an initial application of the model to the ADAS-Cog data. Two psychological phenomena in free recall described previously were the primacy and recency effects. As mentioned before, memory theory dictates that the primacy effect is reflected by a system responsible for encoding words into long-term episodic memory storage. In the current HMM, encoding into a LTM storage system corresponds to the *r* parameter, which indicates the transition probability into the L-State from the U-State. Figure 3 shows the mean parameter estimates on the probability scale for each serial position averaged over all participants, along with a 1-*SD* bar from the Bayesian analysis for the healthy, MCI, and AD participants. Figure 3A reveals the pattern suggested by memory theory; namely, parameter *r* estimates belonging to the beginning of the list tend to be larger than those toward the end of the list for all three participant groups. In addition, there are large drops in the *r* parameter with increasing levels of dementia, especially for the early study list positions. It is important to emphasize that these results, suggesting a primacy effect in parameter *r*, were in no way forced by the methodology used to estimate the parameters.

Figure 3A does have exceptions to the strictly decreasing pattern one would expect to see for the primacy effect. The most prominent exception is in Position 7 of parameter *r* for the three participant groups. As noted earlier, there is some inherent unsystematic noise in the data shown in Figure 2 that can be attributed to experimental procedures. Unfortunately, this inherent noise now seems to be carried over into the parameter estimates of the model showing that the words in position seven have a higher probability of entering LTM than expected by the amount of rehearsal time allotted to that position. Although no model can be assumed to be completely correct, for the purpose of understanding the latent memory processes, an evidenced-based revision of our model will be proposed to control for effects that may arise from the use of testing procedures that do not counterbalance items and their order.

The recency effect is thought to be based on a systematic retrieval process from STM. Figure 3B provides the mean value from the population posterior distribution for the *t* parameters on the probability scale along with 1-*SD* bars for each study list position corresponding to the healthy, MCI, and AD participants. The parameter *t* represents a retrieval probability from the I-State, which our model assumes is a short-term storage system with a rapidly decaying memory trace similar to the theoretical STM. This suggests that retrieval from this memory storage should reflect the memory theoretic predictions of the recency effect.

Table 3
Permutation Test (and 95% Confidence Intervals) for the Three Groups

Group	Trial 1	Trial 2	Trial 3	Trial 4
Healthy	2.355 (1.7466, 2.5212)	2.500* (1.4610, 2.1081)	2.353* (1.2059, 1.7452)	3.457* (1.6136, 2.3390)
MCI	2.67* (1.7419, 2.2959)	3.326* (1.9578, 2.5838)	3.497* (1.8186, 2.3892)	5.399* (1.9745, 2.6006)
AD	2.253* (1.3100, 1.9349)	2.805* (1.7137, 2.5432)	3.636* (1.7036, 2.5447)	2.358* (0.9260, 1.3805)

Note. MCI = mild cognitive impairment; AD = Alzheimer's disease. Asterisks indicate significance at the .05 level (2-tailed).

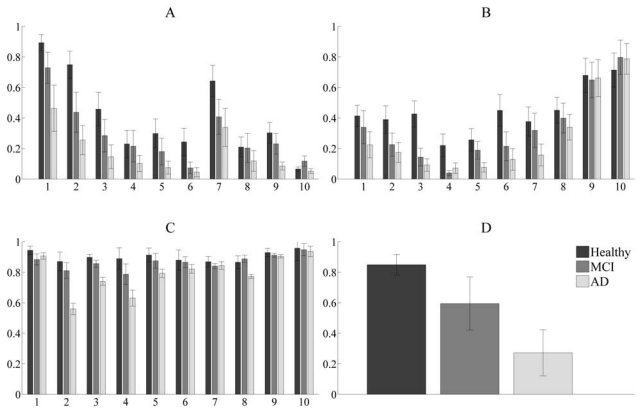

Figure 3. (A) The mean *r* storage parameters (error bar: ± 1 *SD*) for the 10 serial positions for healthy, mild cognitive impairment (MCI), and Alzheimer's disease (AD) groups, obtained from the posterior distributions. (B) The mean *t* retrieval parameters (error bars: ± 1 *SD*) of the 10 word list positions, for healthy, MCI, and AD groups, obtained from the posterior distributions. (C) The mean l_1 retrieval parameters (error bars: ± 1 *SD*) of the 10 word list positions for healthy, MCI, and AD groups, obtained from the posterior distributions. (D) The mean l_2 retrieval parameter (error bars: ± 1 *SD*) for the healthy, MCI, and AD groups.

Figure 3B shows that the estimates of *t* tend to increase with proximity to the test phase for all three participant groups. Furthermore, there is evidence that the overall recall probability from the I-State decreases with levels of dementia further away from the end of the list. It can be noted that there exists some deviations from an expected recency effect; in particular, it appears that the parameters for serial positions at the beginning of the list have slightly higher recall probabilities relative to positions in the middle of the list than might be expected from memory theory. Again, a revision of the model will be introduced to standardize the signal from the unsystematic noise.

The final psychological phenomena discovered in the analysis of the HMM is found in the L-State retrieval parameter. Figure 3C presents the mean l_1 values for each serial position with 1-*SD* bar on the probability scale. As a reminder, parameter l_1 is the probability of recalling a word from the L-State on a test trial presented immediately after a study trial. Memory theory suggests that when words have been encoded into a long-term episodic memory system, the recall probabilities would not have a rapid decay. As a consequence, it is consistent with memory theory to expect that the recall probability l_1 would not show a presentation order effect because the time between study and test is relatively short for each study list position. This property is found in the estimates of the recall parameter for the L-State, shown in Figure 3C. In particular, values of l_1 are similar across different word positions for the three groups of participants.

An additional characteristic of the estimates of l_1 is the lack of a sharp decline in the MCI and AD groups as found for the other parameters. This is interesting because it shows that not all latent cognitive processes specified in the model have equivalent amounts of deficit across various levels of cognitive impairment. However, the level of retrieval from the L-State after a delay, measured by l_2, shows the difference a time delay can have on different levels of deficits. Comparing Figures 4C and 4 D shows that there is a relatively small decline in retrieval from the L-State for the healthy participants as expected from memory theory, while also revealing a large drop in retrieval ability from the L-State for the patients with AD. As for the remaining parameters, no discernable trend was observed across the 10 serial positions. Table 4 displays the means of each of these parameters averaged over the10 serial positions along with 1 *SD* obtained from the average posterior distributions of each participant parameter. Differences between groups are seen in the parameters, *a*, and *v* but not in *b.*

In summary, the analysis of the ADAS-Cog data with our model has revealed patterns in the parameters *r*, *t*, l_1, and l_2 that are consistent with memory theory, which provides some construct validity for the model and indicates that it may be useful for interpreting the differences in the groups as shown by the simple ANOVA in Table 2. Of course the patterns of these estimates were not in perfect accord with expectations from memory theory, and we attribute this to a combination of ordinary random variability, as well as a result of the experimental design, where all participants had the same list of words in the same set of trial-to-trial orders. Modifications to the model in the next section will focus on strengthening the signal and eliminating noise produced by the experimental procedures.

Evidence-Based Revision of HMM

The goal of the preliminary analysis in the previous subsection was twofold. The first goal of the analysis was to accentuate the similarities between memory theory and the results obtained by analyzing the data with the model. In this way, evidence for construct validity of interpreting the parameters as tapping latent cognitive process was obtained. The second goal, assuming success of the first, was to discover which latent cognitive processes are affected by increasing levels of impairment. To complete the second goal, it would be beneficial if the cognitive measurements were less affected by the fixed structure of the ADAS-Cog experimental design. To do this, the relationship between memory theory and our model parameters is explored and further strengthened by adding constraints on the parameters to match known psychological phenomena. By adjusting our model's parameters to match memory theory assumptions, we create a cognitive psychometric model whose application gives more interpretable measurements of the latent processes that are affected by increasing levels of impairment.

For the current data, we will modify three parameters to get more interpretable results. Based on the patterns discernable in Figure 3, we add parameter specifications to *r* and *t* by requiring the underlying parameters to satisfy weak order constraints. The

Table 4
Average Parameter Values for a, v, and b

Variable	Healthy	MCI	AD
a	0.6967 (.1630)	0.6240 (.2043)	0.4934 (.1790)
v	0.3457 (0.1785)	0.1641 (0.0844)	0.1465 (0.0803)
b	0.3986 (0.1439)	0.3377 (0.1818)	0.3552 (0.2223)

Note. MCI = mild cognitive impairment; AD = Alzheimer's disease. Means of all participant posterior distributions are averaged across the 10 serial positions and presented here along with the 1 *SD* of the estimates pertaining to the 10 serial positions.

order constraints on both r and t are as follows: if $j < k$ then $r_j \geq r_k$ and $t_j \leq t_k$. Imposing these order constraints on r and t does not reduce the number of parameters just their relationships to each other. This approach has been used in another cognitive psychometric models applied to special clinical populations, (e.g., Riefer, Knapp, Batchelder, Bamber, & Manifold, 2002). The third modification will be on parameter l_1. The modification of the parameter l_1 is based on the findings in Figure 3C that show little difference across the 10 l_1 parameters. As a consequence, we equate the 10 l_1 parameters within each study group over the study list positions. The remaining parameters that reflect study list positions, a and v, showed no discernible patterns so we imposed no constraints on them, and in addition the parameters b and l_2 were not constrained.

The hyperparameter distributions of the unconstrained parameters (a_i, v_i, b_k, l_2) in the model are set as before to be drawn from independent Gaussian distributions with mean 0 and precision 1, as is the single l_1 parameter in the constrained model. These draws for each participant are transformed via an inverse probit as before, and then for the figures they are averaged to create values on the probability scale. In the case of the order constrained parameters r and t, one addition to the sampling scheme for the unconstrained HMM is needed to impose the order constrains. Order statistics (David & Nagaraja, 2003) on parameters r and t are applied after the inverse probit transforms. In particular, an inverse probit is applied to each participant's set of draws for the 10 t parameters, and then they are ordered from smallest to largest. The means in the figures are based on the participants' means of these ordered draws. This approach in essence assumes a uniform distribution on all ordered sequences $0 \leq t_1 \leq \ldots \leq t_{10} \leq 1$ making them equiprobable. The same approach generated the distribution of the 10 r parameters, except the study position subscripts are reversed. The likelihood function for the modified model has the same functional form as the original model, but its domain is restricted to parameters that satisfy the constraints of the modified model. In other words, it is the restriction on the hierarchical population distribution samples that assures that the posterior distribution of the modified parameters will satisfy the model constraints. The modified model is analyzed with JAGS and the model code can be found in the supplementary section of this article. For the following figures, the parameter value reported is obtained from the average posterior distribution means of each participant's individual model parameters. One-*SD* bar will be presented to indicate the dispersion of parameter values across participants for each group in the current study.

Results of Modified HMM

Figure 4 provides the results of applying the modified model to the data. The first result concerns the parameter r. Figure 4A reveals that words located at the beginning of the study list have a higher probability of being stored into the L-State than words presented at the end of the list, which of course reflects the effect of imposing order constraints. What is of interest is the difference in the r parameters between the three participant groups. Figure 4A shows that the Healthy group has the highest value of the r parameters followed by the MCI group and then the AD group. It appears that the decline in ability to transition a word into the L-State for the MCI participants is closer to the healthy group for the beginning two positions and after the third position it drops to similar levels as for the participants with AD.

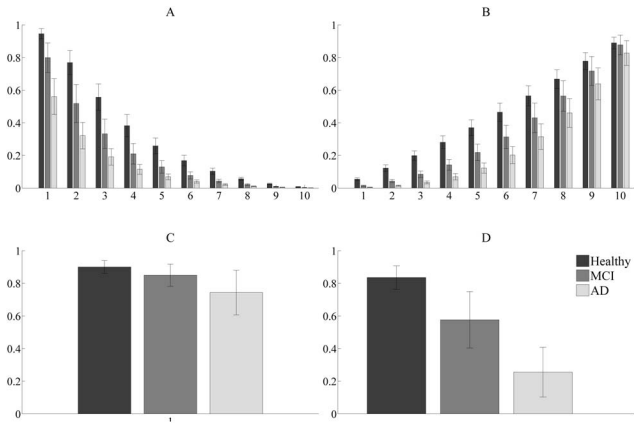

Figure 4. (A) The mean r storage parameters (error bars: ± 1 *SD*) of the 10 serial positions for healthy, mild cognitive impairment (MCI), and Alzheimer's disease (AD) groups, obtained from the posterior distributions of the order constraint model. (B) The mean t retrieval parameters (error bars: ± 1 *SD*) of the 10 word list positions for healthy, MCI, and AD groups, obtained from the posterior distributions of the order constraint model. The mean l_1 (C) and l_2 (D) retrieval parameters (error bars: ± 1 *SD*) for healthy, MCI, and AD groups, obtained from the posterior distributions of the order constraint model.

The first retrieval probability of interest is the probability of recalling a word from the I-State, namely the parameter t. The probability of recalling a word from this temporary storage state for each position for the order-constrained model is in Figure 4B. Again, we see the imposed order effect on the parameter t. Figure 4B shows that words in the last two serial positions are about equally likely to be recalled by a participant in any of the three groups. The groups begin to diverge in their recall ability at Position 7. The recollection by participants with AD was the first to drop at the 7th position, followed by MCI participants at the 6th position. The healthy participant's probabilities of recall from the I-State remained above the MCI and AD group for the preceding positions.

The two retrieval parameters governing the recall probability from the L-State are presented in Figure 4C and 4D. The immediate recall from the L-State during the study-test portion of the task, l_1, shows that the three groups have similar rates of retrieval in the L-State, as was also seen in the unmodified model. The difference between the three groups shows that not all the cognitive processes are affected by the progression of dementia. The largest difference between the three groups on any parameter is shown in Figure 4D for the parameter l_2, which represents the probability of recalling a word from the L-State on the delayed test. The recall proportion after the delay is much lower for the participants in the AD group than in the MCI group.

Now, when a word is not encoded and stored into the L-State from the U-State, the system allows encoding to occur by three other processes. The first process is conditional on the word not being encoded into L-State; in this case, with probability a the word can be encoded into the I-State. No discernible patterns were noticed across the 10 positions for any particular group; however,

the estimates showed a difference between the groups. Table 5 presents the average value of the 10 *a* parameters for each of the three groups. This shows a decreasing pattern across the three groups in ability of encoding into the I-State. A second method of encoding a word into the L-State is through the parameter *v* that encodes a word from the I-State into the L-State during a study trial. Once again, no detectable patterns were observed in each group across the 10 serial list positions; however, participants in the MCI and AD groups showed a decreased probability compared with those in the healthy group. Finally, the *b* parameter in Table 5 represents the third way a word can be encoded into the L-State from the I-State. The transition is only possible during a correct recall of a word from the I-State on the test trial with probability *b*. The result in Table 5 for parameter *b* shows that learning during the test trial is not reduced for those in the MCI and AD groups. As a generalization, it is noteworthy that the results for these parameters in Table 4 match those of the modified model in Table 5.

Discussion of Model Results

Analysis of the ADAS-Cog free-recall data with the HMM revealed several interesting explanations behind the significant differences between the three participants groups found in the ANOVA test. The measurement model revealed that, compared with healthy participants, MCI participant's show an impairment in some but not all of the latent memory processes. Furthermore, the latent memory processes found to be impaired in the MCI group, showed a greater degree of impairment for the AD group. The remainder of this section will be devoted at discussing the differences between the MCI group and the AD group.

By operationally defining latent variables, previous research has shown that encoding into a long-term episodic memory system is impaired with the progression of AD. In our current analysis using a cognitive model, we observe the same decline with the progression of AD. This is important because the model, unlike the operational approach of calculating recall proportions for various items, combines interacting memory processes that are simultaneously at play during the free-recall experiment. A comparison between the three groups reveals a progressive decline in a patient's ability to encode information into a long-term storage system. Although the decline is most notable for those in the AD group, individuals in the MCI group do not show an equivalent ability of encoding as do the healthy participants. It seems that the encoding process is affected at an early stage of AD such that the system responsible for encoding items into LTM shows a marked decline for words in later serial positions compared with words at the beginning of the study list. In other words, MCI participants appear to be able to encode information into their episodic LTM provided that few items are competing for encoding.

As for words not encoded into LTM by the MCI group, their conditional transition into the STM as measured by *a* is nearly has high as the healthy participants. This finding suggests that a person diagnosed with MCI can still store information into a STM state as well as healthy participants. However, with a fast decay rate in the STM, it seems necessary that the memory trace be encoded into a longer, more permanent system to have a greater chance at remembering at a later time. To see whether this is indeed the case, we can look at the *v* parameter in the model, which corresponds to a transition from STM to LTM. It appears that for the MCI group that words encoded into STM are no more likely to transition into the LTM as in the AD group. This result suggests that the cognitive process used to encode words into the LTM from the STM is hindered during early stages of AD. As outlined by Gauthier et al. (2006), MCI is defined as the prodromal stage of AD where participants are classified by an inability to recall conversations or recent events. This matches the current finding of a large drop in ability to encode words into a more permanent memory structure from a temporary one. Mainly, if the memory trace of a conversation or recent event cannot be encoded into the LTM, the chances of retrieving that conversation or recent event is quite low after some time has lapsed.

The final encoding parameter described by the model that allows participants to encode a word into LTM is the parameter *b*. The parameter *b* is an item specific parameter that shows that storage for salient items may occur if they were recalled during the test phase. Similarly, patterns across the first trial for the three groups in Figure 2 showed that certain items were more likely to be recalled across the three groups, such as the word Pole. This suggests that memorable words are as likely to be recalled for healthy participants as for patients with AD. The parameter *b* shows similar values across the three groups, suggesting that encoding during the test phase may be item dependent, and not completely impaired in the progression of AD.

Next, we focus our discussion on the recall probabilities of the latent memory states. While overall there is a decline in *t* with impairment, it is not evident either for MCI or AD participants in *t* for words at the end of the study list whose proximity to the test phase is closest. One possibility for this result is that even with increasing levels of cognitive impairment, words in STM presented right before recall are still resonating in the STM and compete equally well for recall with other memory traces for all three groups. This interpretation is consistent with the findings of Bayley et al. (2000) and Carlesimo et al. (1996) showing that the last 2–3 items are equally recallable. STM impairment in the progression of AD can be viewed as an inability to retain words further away from the end of the list. In other words, for participants with MCI and AD, memory is subject to a faster rate of forgetting in STM as shown by a large drop in retrieval ability for words further away from the list.

Despite the deficits in recall outlined so far, not every cognitive processes is immediately diminished in participants with MCI and AD. The result for the parameter l_1, reflecting immediate recall from the L-State, is quite interesting. Basically, using both the

Table 5
Average Parameter Values for a, v, and B for the Modified Model

Variable	Healthy	MCI	AD
a	0.6645 (0.1617)	0.5925 (0.2194)	0.5452 (0.2369)
v	0.3921 (0.1939)	0.1905 (0.1218)	0.1940 (0.1165)
b	0.4581 (0.2176)	0.4535 (0.1885)	0.4419 (0.2842)

Note. MCI = mild cognitive impairment; AD = Alzheimer's disease. Means of all participant posterior distributions are averaged across the 10 serial positions and presented here along with the 1 *SD* of the estimates pertaining to the 10 serial positions. The results are obtained using the modified model.

original and revised HMM, it was shown that recall probabilities from the LTM on immediate test trials were independent of serial position. In addition, there did not seem to be any noticeable effect of impairment level on l_1, and if this result holds up in other applications of the model, it represents a new finding about episodic memory deficits. In particular, even though there is considerable impairment in achieving a long-term episodic memory trace as measured by *r* and *v,* there is no deficit in the ability of that trace to support recall when the recall test occurs soon after the encoding.

In contrast, the parameter l_2 that measures the delayed recall probability of the L-State, shows a large decline with increasing levels of impairment. This finding shows that after approximately five Minutes AD participant's ability to recall from LTM diminishes very fast. It seems that as a patient's impairment level increases their ability to retain information is no longer aided by what should be a long-term and slow decaying memory state. This interpretation of the L-state comes from the results of the healthy group, showing a small decline in their retrieval ability. While the relatively short time span between an initial measure of LTM retrieval ability and the delayed retrieval measure should not warrant such a decrease, it has been noted that forgetting occurs fastest after short time lag for AD participants (Hart, Kwentus, Taylor, & Harkins, 1987).

Assessing Model Adequacy

The new constrained HMM reflects knowledge of the latent cognitive processes based on psychological theory. To test whether the modifications do not limit the ability of the model to fit the observed data, we test the fit of the unconstrained model and the constrained model. To test each hierarchical model with the data, we use a Bayesian *p* value (Gelman et al., 2014) on the posterior predictive distributions. First, one selects a statistic of the data that is deemed important. Then, a distribution of this statistic is generated from various parameter sets obtained from samples during the Markov Chain Monte Carlo runs. Each such sampled parameter set is used to simulate a data set from the model, and from each such data set the value of the chosen statistic is obtained. A distribution made of these samples constitutes the posterior predictive distribution of the statistic, and it can be thought of as the distribution of future data conditioned on the model posterior parameters. Then a *p* value for the statistic is obtained by referring the observed value of the statistic to this distribution.

A testable statistic that is often used for free recall is the number of correctly recalled words over the 4 trials. For example, as mentioned the ADAS-Cog manual recommends analyzing the number of correctly recalled words for each participant on each trial. For the current test, the fit of each model will be evaluated on these statistics for each participant. Thus, distributions of replicated values predicted by each model for the number of correctly recalled words on each trial and the total number correct across all trials were computed for each participant. Then, the *p* value for each participant and trial is the location of the observed data in the distribution of posterior predictive replicated values (Gelman, Meng, & Stern, 1996). Table 6 shows the proportion of participants whose *p* values lie within the 95% probability interval of the distribution of replicated values, which is known as the predictive concordance of the model. Support for a model is indicated when this value is near to 95% (Gelfand, 1996). Both models performed fairly well using this test since the total sum values are not far from the desired value of .95, which shows that the modifications did not impair the model's ability to fit the data. In the current study, the goal of the comparison is not to deem one model version to be better than other; rather, in any particular application one can decide whether the original model or the order-constrained version is the better way to analyze the data. The modification of the model is motivated by psychological theory and the check of fit to the observed data shows that the changes do not create a worse fit. By focusing on model fit, the Bayesian *p* value can be used to measure the discrepancy as a measure of model adequacy (Meng, 1994).

Table 6
Proportion of Bayesian p-Value's Within the Corresponding 95% Credible Interval

Groups	Trial 1	Trial 2	Trial 3	Trial 4	Total sum
Healthy					
M_1	0.9610	0.9024	0.8244	0.9220	0.9463
M_2	0.9707	0.9024	0.8293	0.9220	0.9659
MCI					
M_1	0.9641	0.9392	0.9088	0.9807	0.9448
M_2	0.9696	0.9696	0.9503	0.9807	0.9614
AD					
M_1	0.9548	0.9887	0.9266	1.000	0.9379
M_2	0.9718	0.9887	0.9887	1.000	0.9774

Note. MCI = mild cognitive impairment; AD = Alzheimer's disease. For each trial, model, and group, the entries report the proportion of participants whose observed response score was within the 95% credible interval of the posterior predictive distribution.

Now that the hierarchical model has shown the ability to account for the observed aggregate recall scores, it is of interest to see whether certain individual model parameters can perform well in explaining variations across participants within a study group in the observed scores. Of course, there are many model parameters that the model combines to achieve the fits reported in Table 6, so we selected two central parameters as candidates to study, namely l_1 and l_2. In addition, we selected the proportion of correct recalls on the first three test trials (score between 0 and 30) and the proportion correct on the delayed trial (score between 0 and 10). Table 7 presents the Pearson product–moment correlation coefficients of three comparisons.

It is noted in Table 7 that the correlations between l_2 and Trial 4 performance scores are highest in each group. The model assumes that correct recall is possible on the delayed trial only if the item has reached the L-State, so this is a nice predictive result for the model. Note, performance on a delayed recall test has been shown to be sensitive in differentiating AD and Healthy participants (Welsh, Butters, Hughes, Mohs, & Heyman, 1992), so l_2 may help explain the differences between the groups. Now, the correlations between l_1 and the first three test trials are lower. Because performance on the first three test trials can come about both from the I-State and the L-State, one would not expect that l_1 would be able to explain as much variance in the first three trials as does l_2 for the delayed trial. The correlation between l_1 and l_2 is consistent with the model assumption that l_1 has decayed during the delayed test. Consequently participants with a larger l_1 are more likely to have a larger value of l_2 after the delay regardless of their group assignment.

Table 7
Pearson Product-Moment Correlation Coefficients Between Long-Term Memory (LTM) Retrieval Parameters, L_1 and L_2, and Performance Scores on the First Three Trials

Group	$\rho(l_1, l_2)$	$\rho(l_1, T_{1-3})$	$\rho(l_2, T_4)$
Healthy	.3970*	.6328*	.8281*
MCI	.2395*	.4811*	.8970*
AD	.3004*	.7245*	.9530*

Note. MCI = mild cognitive impairment; AD = Alzheimer's disease. The asterisks indicate significance at the .01-level (two-tailed).

Conclusion

Early detection of AD is quite important to clinicians and families of those with the disease. An advantage that clinicians have is their use of cognitive tests to classify the likelihood a person is impaired. These cognitive tests are similar to the experimental procedures used to develop cognitive models to measure latent memory processes. For this reason, it makes sense to attempt to apply cognitive models rather than simple statistical analyses to data obtained from the cognitive tests used by clinicians. By combining a formal cognitive model with established psychological theory, we are able to measure a few latent cognitive processes associated with learning and memory from the behavioral measures collected by clinicians using the ADAS-Cog free-recall task. Doing so allows a more complete picture of which cognitive processes are affected by the progression of dementia.

The class of HMMs is adopted to accomplish our purposes. Our HMM demonstrates that with a simple two memory storage state system, the primacy and recency effects are a byproduct of underlying latent cognitive processes extrapolated from behavioral measures gathered using a variable order study list. Support for the two memory systems has been provided by both memory theory and neuropsychological studies showing different memory disorders associated with deficits in performance for early and later parts of the study list (Baddeley & Warrington, 1970; Basso, Spinnler, Vallar, & Zanobio, 1982).

The result of the analysis demonstrates that by using a mathematical model, we are able to circumvent potential problems caused by having to estimate where the primacy effect ends and where the recency effect begins. An immediate consequence of this can be seen with the measurement of the recency effect. The result of our model supports the findings that, although participants with AD show a decline in recall from STM, their problem arises from an inability to recall earlier items but not the last items in the study list. In effect, reducing the STM capacity of storage, possibly causing a diminished ability found in the recency effect. Another finding, after application of the model, was seen in the retrieval process in the long-term storage state, showing similarity to the hypothesized LTM. Furthermore, the retrieval characteristics of this state, as measured by the healthy group, indicated a level of forgetting consistent with a slow decaying process. The rapid forgetting measured for the AD group indicates that AD patient's ability to hold on to memory rapidly declines after an initial storage (Hart et al., 1987). The comparison between healthy and AD was possible because of similar initial immediate recall abilities from LTM, as measured by the parameter l_1. The effects of this decline is possibly the reason behind the low performance on delayed trials and may be the reason why the delayed recall task is sometimes used as a proxy for measuring LTM (Welsh et al., 1992).

One motivation for the current application of the model to the three groups was to establish the utility of standardizing measures of cognitive function in the progression of AD, any of which might be impaired with the progression of AD (Nebes, 1992). Other applications of our model can prove useful, for example, by quantifying the latent variables, clinicians may be able to use the model for assessing changes, if any, in psychological processes with certain drug interventions. Other applications of the model can include the study of memory disorders, such as vascular dementia; to understand what latent processes are affected by their disorder. By broadening the scope of the model's application to other memory disorders, a comparison across memory disorders can then be attainable. Doing so may reveal additional benefits such as assessing which psychological processes are helped by certain prophylactics.

In addition, by making the model hierarchical it becomes possible to use it to detect participants that might be misclassified by physicians because the Bayesian hierarchical inference of the model returns posterior distributions of each parameter for each participant. Nevertheless, the estimation theory for the model may have to be augmented to allow it to better classify individual participants. One idea for future work would be to construct an empirical Bayesian prior based on a large sample of participants who enter a clinic and are tested. Such a prior would be highly informed unlike the priors that were used to analyze the HMM. Batchelder (1998) describes how a data bank of this sort could be constructed based on a cognitive model and used to classify individuals.

References

Baddeley, A. D., & Warrington, E. K. (1970). Amnesia and the distinction between long-and short-term memory. *Journal of Verbal Learning & Verbal Behavior, 9,* 176–189. http://dx.doi.org/10.1016/S0022-5371(70)80048-2

Basso, A., Spinnler, H., Vallar, G., & Zanobio, M. E. (1982). Left hemisphere damage and selective impairment of auditory verbal short-term memory. A case study. *Neuropsychologia, 20,* 263–274. http://dx.doi.org/10.1016/0028-3932(82)90101-4

Batchelder, W. H. (1998). Multinomial processing tree models and psychological assessment. *Psychological Assessment, 10,* 331–344. http://dx.doi.org/10.1037/1040-3590.10.4.331

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, 52*B, 206–215. http://dx.doi.org/10.1093/geronb/52B.5.P206

Bayley, P. J., Salmon, D. P., Bondi, M. W., Bui, B. K., Olichney, J., Delis, D. C., . . . Thal, L. J. (2000). Comparison of the serial position effect in very mild Alzheimer's disease, mild Alzheimer's disease, and amnesia associated with electroconvulsive therapy. *Journal of the International Neuropsychological Society, 6,* 290–298. http://dx.doi.org/10.1017/S1355617700633040

Bernardo, J. M., & Smith, A. F. (2009). *Bayesian theory* (Vol. 405). New York, NY: John Wiley & Sons.

Bjork, R. A., & Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology, 6,* 173–189. http://dx.doi.org/10.1016/0010-0285(74)90009-7

Bozoki, A., Grossman, M., & Smith, E. E. (2006). Can patients with Alzheimer's disease learn a category implicitly? *Neuropsychologia, 44,* 816–827. http://dx.doi.org/10.1016/j.neuropsychologia.2005.08.001

Brainerd, C. J., Reyna, V. F., Gomes, C. F. A., Kenney, A. E., Gross, C. J., Taub, E. S., . . . the Alzheimer's Disease Neuroimaging Initiative. (2014). Dual-retrieval models and neurocognitive impairment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 41–65. http://dx.doi.org/10.1037/a0034057

Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language, 55,* 627–652. http://dx.doi.org/10.1016/j.jml.2006.08.005

Buschke, H. (1973). Selective reminding for analysis of memory and learning. *Journal of Verbal Learning & Verbal Behavior, 12,* 543–550. http://dx.doi.org/10.1016/S0022-5371(73)80034-9

Capitani, E., Della Sala, S., Logie, R. H., & Spinnler, H. (1992). Recency, primacy, and memory: Reappraising and standardising the serial position curve. *Cortex, 28,* 315–342. http://dx.doi.org/10.1016/S0010-9452(13)80143-8

Carlesimo, G. A., Fadda, L., Sabbadini, M., & Caltagirone, C. (1996). Recency effect in Alzheimer's disease: A reappraisal. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 49,* 315–325. http://dx.doi.org/10.1080/713755622

Carlesimo, G. A., & Oscar-Berman, M. (1992). Memory deficits in Alzheimer's patients: A comprehensive review. *Neuropsychology Review, 3,* 119–169. http://dx.doi.org/10.1007/BF01108841

Chu, L. W., Chiu, K. C., Hui, S. L., Yu, G. K., Tsui, W. J., & Lee, P. W. (2000). The reliability and validity of the Alzheimer's Disease Assessment Scale Cognitive Subscale (ADAS-Cog) among the elderly Chinese in Hong Kong. *Annals of the Academy of Medicine, Singapore, 29,* 474–485.

David, H. A., & Nagaraja, H. N. (2003). *Order statistics*. Hoboken, NJ: John Wiley & Sons. http://dx.doi.org/10.1002/0471722162

Egli, S. C., Beck, I. R., Berres, M., Foldi, N. S., Monsch, A. U., & Sollberger, M. (2014). Serial position effects are sensitive predictors of conversion from MCI to Alzheimer's disease dementia. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association, 10*(Suppl.), S420–S424. http://dx.doi.org/10.1016/j.jalz.2013.09.012

Fillenbaum, G. G., van Belle, G., Morris, J. C., Mohs, R. C., Mirra, S. S., Davis, P. C., . . . Heyman, A. (2008). Consortium to Establish a Registry for Alzheimer's Disease (CERAD): The first twenty years. *Alzheimer's & Dementia: The Journal of the Alzheimer's Associaiton, 4,* 96–109. http://dx.doi.org/10.1016/j.jalz.2007.08.005

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12,* 189–198. http://dx.doi.org/10.1016/0022-3956(75)90026-6

Gauthier, S., Reisberg, B., Zaudig, M., Petersen, R. C., Ritchie, K., Broich, K., . . . the International Psychogeriatric Association Expert Conference on mild cognitive impairment. (2006). Mild cognitive impairment. *The Lancet, 367,* 1262–1270. http://dx.doi.org/10.1016/S0140-6736(06)68542-5

Gelfand, A. (1996). Model determination using sampling-based methods. In W. Gilks, S. Richardson, & D. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in practice* (pp. 145–161). Boca Raton, FL: Chapman & Hall.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). *Bayesian data analysis* (Vol. 2). Boca Raton, FL: Chapman; Hall/CRC.

Gelman, A., Meng, X. L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica, 6,* 733–760.

Gibson, A. J. (1981). A further analysis of memory loss in dementia and depression in the elderly. *The British Journal of Clinical Psychology, 20,* 179–185. http://dx.doi.org/10.1111/j.2044-8260.1981.tb00516.x

Goldstein, B. E. (2010). *Cognitive psychology: Connecting mind, research and everyday experience*. Belmont, CA: Wadsworth Publishing Company, Inc.

Gowans, E. M. S., Fraser, C. G., & Hyltoft Petersen, P. (1989). A guide to the use of probit transformation of Gaussian distributions. *Biochimica Clinica, 13,* 327–336.

Graham, D. P., Cully, J. A., Snow, A. L., Massman, P., & Doody, R. (2004). The Alzheimer's Disease Assessment Scale-Cognitive subscale: Normative data for older adult controls. *Alzheimer Disease and Associated Disorders, 18,* 236–240.

Greeno, J. G., & Bjork, R. A. (1973). Mathematical Learning Theory and the new "mental forestry." *Annual Review of Psychology, 24,* 81–116. http://dx.doi.org/10.1146/annurev.ps.24.020173.000501

Hart, R. P., Kwentus, J. A., Taylor, J. R., & Harkins, S. W. (1987). Rate of forgetting in dementia and depression. *Journal of Consulting and Clinical Psychology, 55,* 101–105. http://dx.doi.org/10.1037/0022-006X.55.1.101

Hodges, J. R., Salmon, D. P., & Butters, N. (1992). Semantic memory impairment in Alzheimer's disease: Failure of access or degraded knowledge? *Neuropsychologia, 30,* 301–314. http://dx.doi.org/10.1016/0028-3932(92)90104-T

Howard, M. W., & Kahana, M. J. (2001). A distributed representation of temporal context. *Journal of Mathematical Psychology*. Advance online publication.

Howieson, D. B., Mattek, N., Seeyle, A. M., Dodge, H. H., Wasserman, D., Zitzelberger, T., & Jeffrey, K. (2011). Serial position effects in mild cognitive impairment. *Journal of Clinical and Experimental Neuropsychology, 33,* 292–299. http://dx.doi.org/10.1080/13803395.2010.516742

Kirk, R. E. (1968). *Experimental design: Procedures for the behavioral sciences*. Belmont, CA: Wadsworth Publishing Company, Inc.

Kraemer, H. C., Peabody, C. A., Tinklenberg, 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. http://dx.doi.org/10.1037/0033-2909.94.2.367

Lee, M. D., & Wagenmakers, E. J. (2013). *Bayesian cognitive modeling: A practical course*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139087759

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS-A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing, 10,* 325–337.

Martin, A., Brouwers, P., Cox, C., & Fedio, P. (1985). On the nature of the verbal memory deficit in Alzheimer's disease. *Brain and Language, 25,* 323–341. http://dx.doi.org/10.1016/0093-934X(85)90088-4

Meng, X. L. (1994). Posterior predictive p-values. *Annals of Statistics, 22,* 1142–1160. http://dx.doi.org/10.1214/aos/1176325622

Miller, E. (1971). On the nature of the memory disorder in presenile dementia. *Neuropsychologia, 9,* 75–81. http://dx.doi.org/10.1016/0028-3932(71)90064-9

Mohs, R. C., Knopman, D., Petersen, R. C., Ferris, S. H., Ernesto, C., Grundman, M., . . . Thal, L. J. (1997). Development of cognitive instruments for use in clinical trials of antidementia drugs: Additions to the Alzheimer's Disease Assessment Scale that broaden its scope. The Alzheimer's Disease Cooperative Study. *Alzheimer Disease and Associated Disorders, 11*(Suppl. 2), S13–S21. http://dx.doi.org/10.1097/00002093-199700112-00003

Mohs, R. C., Rosen, W. G., & Davis, K. L. (1983). The Alzheimer's disease Assessment Scale: An instrument for assessing treatment efficacy. *Psychopharmacology Bulletin, 19,* 448–450.

Murdock, B. B., Jr. (1962). The serial position effect of free recall. *Journal of Experimental Psychology, 64,* 482–488. http://dx.doi.org/10.1037/h0045106

Nebes, R. D. (1992). Cognitive dysfunction in Alzheimer's disease. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 373–446). Hillsdale, NJ: Erlbaum.

Neufeld, R. (Ed.). (1988). Process models in psychological assessment [Special issue]. *Psychological Assessment, 10,* 307–443.

Neufeld, R. (Ed.), (2007). *Advances in clinical cognitive science: Formal modeling of processes and symptoms* (pp. 19–50). Washington, DC: American Psychological Association Books. http://dx.doi.org/10.1037/11556-000

Neufeld, R., & Townsend, J. (Eds.). (2010). Contributions of mathematical psychology to clinical sciences and assessment. *Journal of Mathematical Psychology, 54,* 1–214.

Plummer, M. (2003, March). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing* (Vol. 124, p. 125). Vienna.

Riefer, D. M., Knapp, B. R., Batchelder, W. H., Bamber, D., & Manifold, V. (2002). Cognitive psychometrics: Assessing storage and retrieval deficits in special populations with multinomial processing tree models. *Psychological Assessment, 14,* 184–201. http://dx.doi.org/10.1037/1040-3590.14.2.184

Rundus, D. (1971). Analysis of rehearsal procedures in free recall. *Journal of Experimental Psychology, 89,* 63–77. http://dx.doi.org/10.1037/h0031185

Smith, J. B., & Batchelder, W. H. (2008). Assessing individual differences in categorical data. [Chicago.]. *Psychonomic Bulletin & Review, 15,* 713–731. http://dx.doi.org/10.3758/PBR.15.4.713

Spinnler, H., Sala, S. D., Bandera, R., & Baddeley, A. (1988). Dementia, ageing, and the structure of human memory. *Cognitive Neuropsychology, 5,* 193–211. http://dx.doi.org/10.1080/02643298808252933

Tulving, E., & Colotla, V. (1970). Free recall of trilingual lists. *Cognitive Psychology, 1,* 86–98. http://dx.doi.org/10.1016/0010-0285(70)90006-X

Tulving, E., & Patterson, R. D. (1968). Functional units and retrieval processes in free recall. *Journal of Experimental Psychology, 77,* 239–248. http://dx.doi.org/10.1037/h0025788

Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review, 72,* 89–104. http://dx.doi.org/10.1037/h0021797

Welsh, K. A., Butters, N., Hughes, J. P., Mohs, R. C., & Heyman, A. (1992). Detection and staging of dementia in Alzheimer's disease. Use of the neuropsychological measures developed for the Consortium to Establish a Registry for Alzheimer's Disease. *Archives of Neurology, 49,* 448–452. http://dx.doi.org/10.1001/archneur.1992.00530290030008

Wenger, M. K., Negash, S., Petersen, R. C., & Petersen, L. (2010). Modeling and estimating recall processing capacity: Sensitivity and diagnostic utility in application to mild cognitive impairment. *Journal of Mathematical Psychology, 54,* 73–89. http://dx.doi.org/10.1016/j.jmp.2009.04.012

Wickens, T. D. (1982). *Models for behavior: Stochastic processes in psychology*. San Francisco, CA: Freeman and Company.

Wilson, R. S., Bacon, L. D., Fox, J. H., & Kaszniak, A. W. (1983). Primary memory and secondary memory in dementia of the Alzheimer type. *Journal of Clinical Neuropsychology, 5,* 337–344. http://dx.doi.org/10.1080/01688638308401181

Received August 25, 2014
Revision received April 9, 2015
Accepted April 15, 2015 ■