Modeling Change in Recognition Bias with the Progression of Alzheimer's

James P. Pooley (jpooley@uci.edu)
Michael D. Lee (mdlee@uci.edu)
Department of Cognitive Sciences, 3151 Social Science Plaza
University of California, Irvine, CA, 92697-5100

William R. Shankle (rshankle@mccare.com)
Medical Care Corporation, 19782 MacArthur Boulevard
Irvine, CA 92612

Abstract

One of the key memory tests in the clinical assessment and diagnosis of Alzheimer's Disease (AD) is the recognition memory task. Models developed in cognitive psychology have previously been applied to help understand clinical data. In particular, Signal Detection Theory (SDT) models have been used, to separate people's memory capabilities from their decision-making strategies. An important finding in this literature is that people with AD change their decision strategy in response to memory impairment, applying a more liberal criterion than people without AD. In this paper, we analyze clinical data that measures the progression of AD in a detailed way, using a theoretically motivated version of SDT, and applying hierarchical Bayesian methods to model individual differences. Our results corroborate many of the previous findings, but provide a more detailed focus on recognition performance with AD progression.

Keywords: Alzheimer's disease; Cognitive psychometrics; Hierarchical Bayesian modeling; Human recognition memory; Signal detection theory

Introduction

The clinical assessment and diagnosis of Alzhiemer's disease (AD) routinely involves the administration of memory tests that are familiar to cognitive scientists who study human memory. In particular, recognition, immediate free recall, and delayed free recall are large subcomponents of assessment tools such as the MCIS and the ADAS-Cog (e.g., Morris, Heyman, & Mohs, 1989). This link means there is an important role for theories and models of memory, as developed in the cognitive sciences (for an overview, see Norman Detre, & Polyn, 2008), in helping understand AD. In particular, memory models can provide quantitative measurement tools that allow for patient behavior to be interpreted in terms of psychologically meaningful latent parameters (e.g., Riefer, Knapp, Batchelder, Bamber, & Manifold, 2002).

A good example of the potential for applying memory models to clinical data is provided by a literature that uses equal-variance Signal Detection Theory (SDT) models (e.g., MacMillan & Creelman, 2004). SDT is widely-used as a basic model of the recognition memory task, and has the theoretical attraction of separating memory capabilities from decision processes when explaining people's behavior (e.g., Budson Wolk, Chong, & Waring, 2006; Snodgrass & Corwin, 1988). This is a very important capability, because there is considerable evidence that AD patients do have different decision-making strategies in tasks like recognition memory.

The recent review by Budson et al. (2006) notes that the application of SDT models to clinical data has repeatedly shown that patients with AD use a more liberal criterion in identifying previously studied words. This strategy is usually interpreted as a response to awareness of diminishing memory capabilities. Additionally, Budson et al. (2006) report the results of an experiment which addressed several potential confounds in the existing experiments, including unequal numbers of old and new words and semantic and/or perceptual relatedness of the old and new words. Again, AD patients were found to have abnormally liberal response biases compared to non-AD patients.

In this paper, we extend the application of SDT models to clinical recognition memory data. We do this in a number of ways. First, we use a large new clinical database, which has the advantage of measuring the progression of AD in some detail. This lets us conduct a finer-grained analysis of how recognition memory changes as AD progresses. Second, we use a simple variant of the standard SDT model that builds in an unequal-variance assumption. This is theoretically preferable, given empirical evidence that there is more variability in people's memory for studied than non-studied words. Third, we embed our SDT analyses with a hierarchical Bayesian framework for statistical inference. This lets us provide a coherent model-based account of variation, at both the level of individual patients, and the level of clinical sub-populations.

The plan of the paper is as follows. We begin by describing the clinical data, and then the unequal-variance SDT model we use. We show that the model provides a good account of the data, and show how inference about the model's parameters gives an interpretable account of changes in recognition memory with the progression of AD. We then extend the modeling to account explicitly for changes in decision bias, and conclude by discussing how our findings relate to the existing literature.

Clinical Data

Our data come from two neurology clinics where 1350 patients completed a standard old/new recognition memory test. The patient was shown a study list of 10 words to memorize, and was then tested on their ability to recognize the 10 studied *old* words from 10 unstudied *new* words. This means there are 20 test trials, on each of which the patient was shown a word and simply asked to decide whether or not the word was on the study list. Consequently, the patient's behavior on each trial naturally falls into one of the standard SDT classes of hits, misses, false alarms, and correct rejections. The words themselves were selected from the CERAD (Consortium to Establish a Registry for Alzheimer's Disease) word list (Shankle, Mangrola, Chan, & Hara, 2009).

Independent of patient performance on the recognition memory tests, a trained neurologist used the Functional Assessment Staging Test (FAST) to assess the severity of each patient's AD. The FAST (Reisberg, 1988) is a well-validated diagnostic tool used by clinicians to classify patients into one of the seven *stages* of AD, each of which corresponds to a level of functional impairment. Specifically, stage 1 corresponds to 'normal aging', stage 2 to 'possible mild cognitive impairment', stage 3 to 'mild cognitive impairment', stage 4 to 'mild dementia', stage 5 to 'moderate dementia', stage 6 to 'moderately severe dementia' and stage 7 to 'severe dementia'. We focus on only FAST Stages 1–5, because patients diagnosed into Stages 6 and 7 have very limited functional capabilities, and cannot necessarily understand and complete memory tasks. In our sample of 1350 patients, 288 were classified as Stage 1, 308 as Stage 2, 129 as Stage 3, 436 as Stage 4, and 189 at Stage 5.

Hierarchical SDT Model

In this section, we describe the hierarchical SDT model we use to analyze the clinical data. We start with a standard SDT model, and then describe how our hierarchical extensions add the capability to model individual differences and changes in bias. We then implement the model as a graphical model to allow Bayesian inference.

Signal Detection Theory

The basic SDT model shown in Figure 1 assumes that, on each trial, the presented word evokes some memory strength. The memory strengths of both old and new words are assumed to have Gaussian distributions, with the mean of the new distribution separated from the mean of the old distribution by a distance $d' > 0$. In this way, d' measures the *discriminability* of the old from the new words, and so represents the acuity of memory for the words.

Due to the assumed overlap of the old and new distributions, an individual needs a decision strategy for relating memory strength to responses in a recognition test. SDT models assume this is done using a criterion level of memory strength k below which the individual will respond studied and above which the individual will respond non-studied. The area h under the old distribution above the criterion corresponds to the hit rate, and the area f under the new distribution above the criterion corresponds to the false-alarm rate.

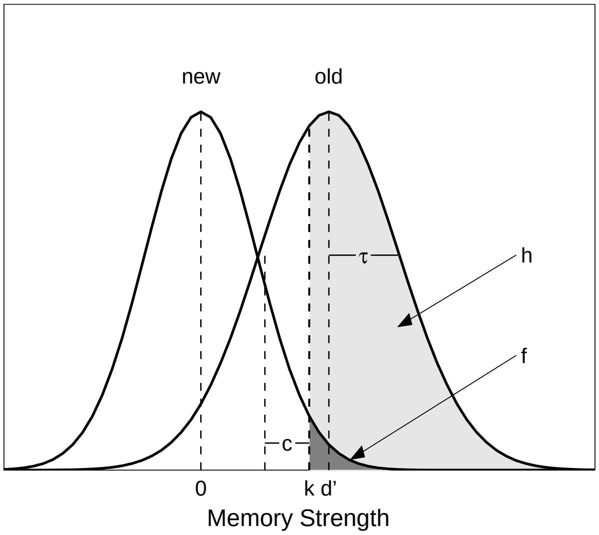

Figure 1: The unequal-variance SDT model and parameters.

The distance c between this criterion and unbiased responding is commonly used as a measure of *response bias* due to its purported independence from d' (Snodgrass & Corwin, 1988). The response bias measures the tendency of an individual to give one response rather than another.

Extension for Unequal Variance

Most SDT modeling in psychology assumes that the standard deviations of the old and new distributions are equal, with $\sigma_{\text{old}} = \sigma_{\text{new}} = 1$ for convenience. Results of recognition memory experiments (e.g., Mickes, Wixted, & Wais, 2007), however, support a version of SDT in which the standard deviation of the old distribution is 25% larger than the standard deviation of the new distribution, so that $\sigma_{\text{new}}/\sigma_{\text{old}} = 0.8$. This finding is usually interpreted as coming from variability in the encoding of studied words. Our SDT model adopts an unequal-variance assumption, using the approach developed by Dennis, Lee, & Kinnell (2008).

Extension for Individual Differences

Most previous applications of SDT models to the recognition memory data of Alzheimer's patients have also ignored the issue of individual differences. To address this shortcoming, we apply hierarchical methods to extend the standard SDT model (e.g., Dennis, Lee, & Kinnell, 2008; Rouder & Lu, 2005). The basic idea is to introduce sub-populations at a group-level that allow for different parameter values for different levels of severity in AD. An individual patient's discriminability and response bias parameters are then drawn from the appropriate group-level distribution for their level of severity. In this way, the model allows freedom for different individuals to have different parameters, but still maintain a

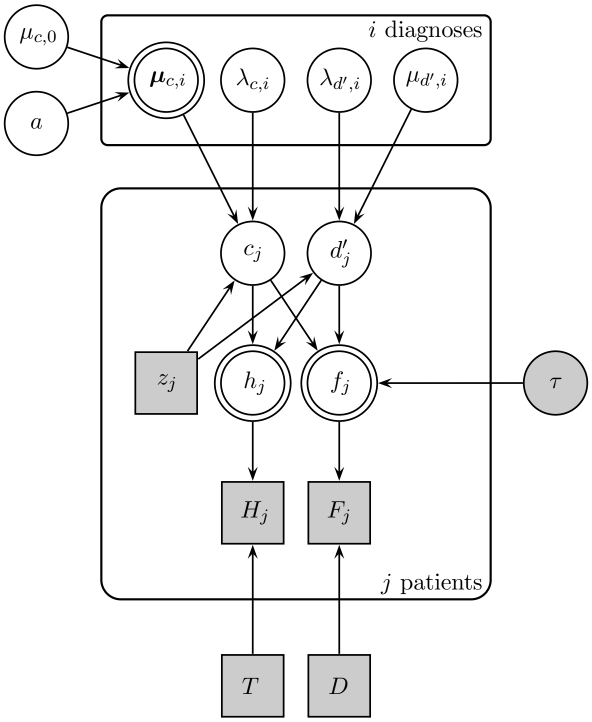

Figure 2: Graphical model implementation of the hierarchical SDT model.

similarity to other patients with a similar level of cognitive impairment.

Extension for Modeling Change

Most previous analyses focusing on changes in response bias with AD progression have taken a purely statistical approach. Typically, they have tested for significant differences in bias or criterion parameters, as inferred separately from AD and non-AD patients. We adopt a different approach based on cognitive modeling, building assumptions about how bias changes into the model itself. This is consistent with the basic idea of *generative* models, which try to provide formal accounts of how latent parameters produce and co-vary with observed behavior, and can be contrasted with the *discriminative* philosophy of post-hoc statistical tests. In the generative approach, a model of change is incorporated into the SDT model, with the goal of providing a complete and integrated account of how the criterion changes with the progression of AD.

Graphical Model Implementation

We implemented our hierarchical SDT model in the form of a Bayesian graphical model, a formalism widely used statistics and computer science (e.g., Jordan, 2004). In graphical models, nodes correspond to variables, and their interdependencies show the causal relationships between the variables. In particular, graphical models show how unobserved variables (i.e., parameters) generate observed variables (i.e., data). Details and tutorials are aimed at cognitive scientists are provided by Lee (2008) and Shiffrin, Lee, Kim, and Wagenmakers (2008). The practical advantage of graphical models is that sophisticated and relatively general-purpose Markov Chain Monte Carlo (MCMC) algorithms exist that can sample from the full joint posterior distribution of the parameters conditional on the observed data.

It is easiest to understand the graphical model in Figure 2 by starting with the d'_j and c_j nodes, which are the discriminability and bias parameters for the jth patient. These parameters can be used to generate the hit and false-alarm rates for that patient, according to the SDT model. The hit rate is $h_j = \Phi(d'_j/2 - c_j)$ and the false alarm rate is $f_j = \Phi(-(d'_j/2 + c_j)/\tau)$, where $\tau = 0.8$ gives the unequal-variance model advocated by Mickes, Wixted, and Wais (2007). Based on these hit and false alarm rates and the $O = 10$ old and $N = 10$ new words presented to all patients during the recognition tests, the jth patient produces $H_j \sim \text{Binomial}(h_j, T)$ hits and $F_j \sim \text{Binomial}(f_j, D)$ false-alarms.

The distributions of discriminability and bias for different AD diagnoses, at the group or sub-population level, are controlled by the mean μ and precision λ variables. There is a Gaussian group distribution for each group. If, for example, we use FAST stage diagnoses to define groups, and the jth patient belongs to stage z_j, then $d'_j \sim \text{Gaussian}(\mu_{d',z_j}, \lambda_{d',z_j})$ and $c_j \sim \text{Gaussian}(\mu_{c,z_j}, \lambda_{c,z_j})$.

Finally, the graphical model in Figure 2 implements a basic model of change for response bias. Following previous analyses (e.g., Snodgrass & Corwin, 1988), we just consider the change from non-AD to AD patients. The parameter $\mu_{c,0}$ measures the non-AD response bias, and a quantifies the change, so that $\mu_{c,1}, \mu_{c,2} = \mu_{c,0}$ and $\mu_{c,3}, \dots, \mu_{c,5} = \mu_{c,0} + a$.

Modeling Results

In order to perform Bayesian inference, we implemented the graphical models in WinBUGS (Spiegelhalter, Thomas, & Best, 2004. This software uses a range of MCMC computational methods to obtain samples from the posterior distributions of the relevant parameters (e.g., Mackay, 2003). All of our analyses are based on 10,000 posterior samples collected following a burn-in of 1000 samples, using multiple chains to check convergence.

Assessing Model Fit

Posterior predictive distributions provide an intuitive and principled to assessing the descriptive adequacy of a Bayesian model (Gelman, Carlin, Stern, & Rubin, 2004, pp. 165–172). A posterior prediction corresponds to the data the model expects, based on the parameter values it has inferred, and naturally takes into account uncertainty in those parameter estimates.

Figure 3 shows a posterior predictive analysis for the

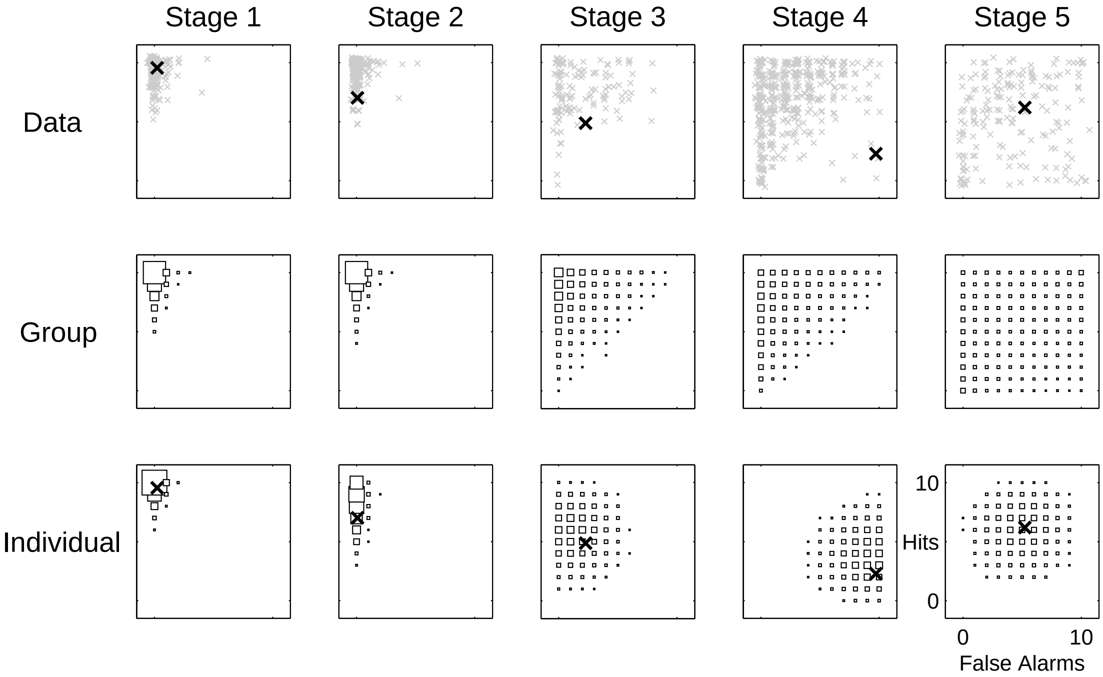

Figure 3: Posterior predictive assessment of the fit of the hierarchical SDT model. The first row shows the hit and false-alarm counts for each patient, according to their FAST stage, with the counts for a randomly selected patient shown in bold. The second and third rows show the corresponding posterior predictive distributions for hit and false alarm counts for the group data and for the individual patient data. In the posterior predictive panels, the box sizes are proportional to the mass of the posterior predictive distribution for that combination of hits and false alarms.

hierarchical SDT model. The first row corresponds to the behavioral data, the second row to the group-level inferences of the model, and the third-level to the individual-level inferences of the model. The columns correspond to the five FAST stages. Each panel shows the distribution of data or predicted data in terms of hit and false-alarm counts, as in standard Receiver Operation Characteristic (ROC) analysis (e.g., MacMillan Creelman, 2004).

The observed data for all patients are shown as gray crosses, except for one highlighted individual—selected out to test the individual-level predictions of the model—shown by a black cross. For the group level, the model's posterior predictions are shown by squares, with areas proportional to predictive mass. It is clear that the group-level predictions match the data, and show a degradation in performance, with fewer hits and more false-alarms, as the severity of AD progresses. In this sense, the model provides an accurate description of the similarities and differences between clinical sub-populations. In the individual-level model predictions, the area of the squares again correspond to predictive mass, and provide accurate fits to the observed data. We note that several of the individuals were deliberately chosen to be outliers within their clinical sub-population. The ability of the model for describe these individuals well, while simultaneously describing group-level performance, highlights the advantages of the hierarchical approach we have taken to modeling individual differences.

Assessing Discriminability and Bias

Figure 4 shows the joint and marginal posterior distributions for both discriminability and bias, at the level of the FAST stage groups. The main panel shows samples from the joint distribution for each of the five FAST stages. The side panels show the marginal distributions for both discriminability and bias.

As would be expected, discriminability decreases as AD severity progresses, starting around $d' = 4$ for non-AD patients in the first two stages, and decreasing to $d' < 1$ for patients in stage 5. The pattern change in recognition bias across the stages is more revealing. Patients in the non-AD stages start with a conservative bias, with $c > 0$, meaning they are more likely to fail to recognize studied words than to false-alarm to non-studied words. This bias changes significantly for the AD patients, and becomes much more liberal, shifting to a position almost consistent with unbiased responding at $c = 0$.

Assessing Change in Recognition Criterion

Figure 4 shows that the change in criterion is sudden and sustained. At FAST stage 3—which is the first AD

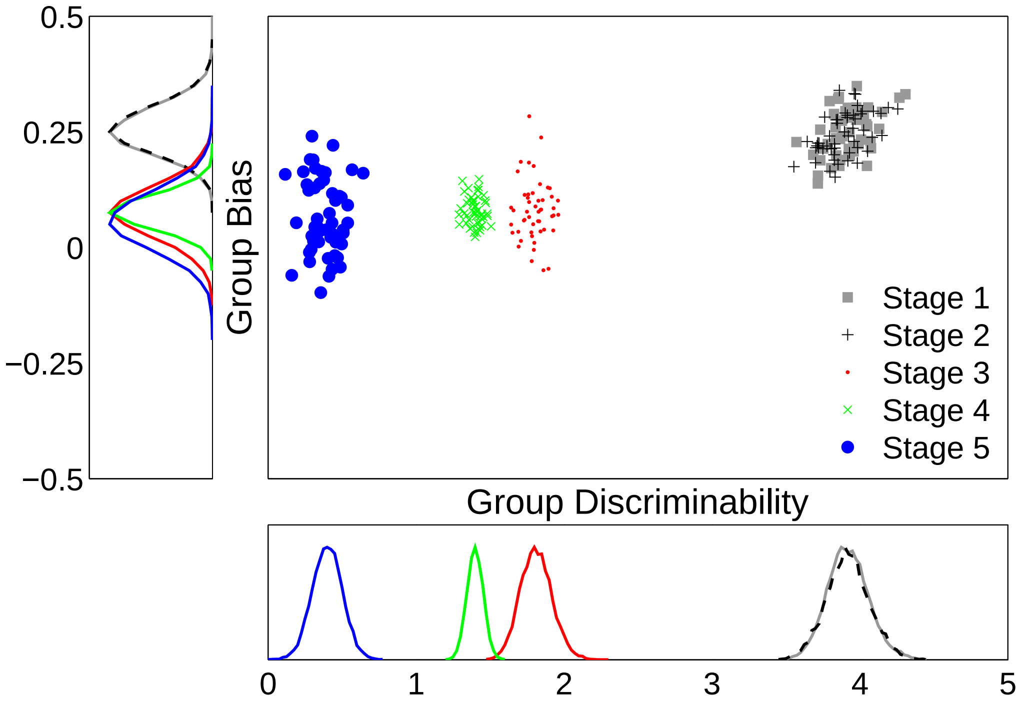

Figure 4: Joint and marginal posterior distributions for the group-level discriminability and bias parameters, for each of the five FAST stages.

stage—the distribution of individual response bias moves to a smaller value, and it sustains approximately the same distribution over subsequent progression through stages 4 and 5.

Our generative model of change allows an immediate inference about the significance of these apparent change in response bias, via the posterior distribution of the a parameter. This is the parameter that control the step-change in response bias between AD and non-AD diagnoses. Its posterior distribution is shown in Figure 5, and is clearly negative, and does not include zero, confirming the liberal change in bias at the onset of AD.

Discussion

Our results are largely consistent with previous findings, but are not identical. We have corroborated the most important existing finding, which is that the onset of AD leads to a liberalization in response bias in recognition memory tasks. Our results, however, extend the previous understanding of the change in response bias, through using a clinical data set with more FAST stage information about AD progression. Using this more detailed measure we found, perhaps surprisingly, that the change in response bias seems to involve a sudden shift at the onset of AD, rather than gradual change over its progression.

Unlike most previous studies, we found non-AD patients starting from a conservative criterion setting—being more likely to miss than to false-alarm—and so the liberalization actually leads to more unbiased decision-making in the AD patients. There are many possible reasons for this difference, which are worth further investigation. One possibility involves methodological issues, including details of the assessment tasks, such as differences in the word lists used. Another possibility relates to more fundamental theoretical and modeling differences in our analysis. We have introduced a number of innovations, any (or all) of which might lead to different findings from more standard analyses.

We think the modeling approach we have used has some clear advantages over previous work. As AD progresses, memory capabilities and decision strategies change in important and interpretable ways. But there remains variability in the characteristics of individual patients, even though they can appropriately be classified within groups like FAST stages. Our hierarchical approach naturally incorporates this interplay between clinical sub-populations and individual patients, making it suitable for both broad characterization of AD progression and for individual diagnosis.

Throughout our modeling, we used a simple extension of the standard SDT model to allow for unequal-variances between studied and non-studied words. We think this theoretically preferable, although we did not observe very different results when we repeated the current analyses with equal-variance SDT. Perhaps the most striking difference was that the posterior for the response bias parameter in Figure 5 showed a much stronger change in bias for the non-AD versus AD comparison.

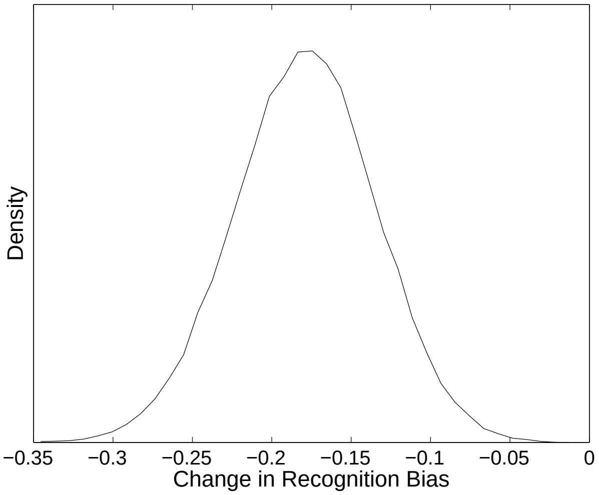

Figure 5: Posterior distribution for the a parameter, which controls the change in response bias from non-AD to AD patients.

It may be that equal-variance SDT overstates the change in decision strategies.

We believe the framework for modeling change we have introduced also has great potential, but realize we have only taken the smallest first step. The key idea is that group-level parameters like discriminability and bias can now be inter-related across diagnoses or classifications like FAST stages. We used a simple step function between non-AD and AD patients, but much more sophisticated functional relationships could be modeled, expressing a theory of how key psychological variables change throughout AD progression. Even more generally, graphical models provide a natural vehicle for modeling and evaluating changes in these variables due to external factors like treatments in clinical trials, or for expressing these variables in terms of causal or co-variate information like demographic or other properties of people. These sorts of extended possibilities highlight the potential of using cognitive models like SDT and hierarchical Bayesian analysis to understand Alzheimer's Disease.

Acknowledgments

This research was supported by award NIRG-08-90460 from the Alzheimer's Association.

References

Budson, A. E., Wolk, D. A., Chong, H., & Waring, J. D. (2006). Episodic memory deficits in Alzheimer's disease: Separating response bias from discrimination. *Neuropsychologia, 44*, 2222–2232.

Dennis, S., Lee, M., & Kinnell, A. (2008). Bayesian analysis of recognition memory: The case of the list-length effect. *Journal of Memory and Language, 59*, 361-376.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Jordan, M. I. (2004). Graphical models. *Statistical Science, 19*, 140-155.

Lee, M. D. (2008). Three case studies in the Bayesian analysis of cognitive models. *Psychonomic Bulletin & Review, 15*, 1-15.

MacKay, D. J. C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge: Cambridge University Press.

MacMillan, N., & Creelman, C. D. (2004). *Detection theory: A users guide* (2nd ed.). Hillsdale, NJ: Erlbaum.

Mickes, L., Wixted, J. T., & Wais, P. E. (2007). A direct test of the unequal-variance signal-detection model of recognition memory. *Psychonomic Bulletin & Review, 14*, 858-865.

Morris, J. C., Heyman, A., & Mohs, R. C. (1989). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD) Part I. Clinical and neuropsychological assessment of Alzheimers disease. *Neurology, 39*, 1159-1165.

Norman, K. A., Detre, G. J., & Polyn, S. M. (2008). Computational models of episodic memory. In R. Sun (Ed.), *The Cambridge handbook of computational psychology* (pp. 189-224). New York: Cambridge University Press.

Reisberg, B. (1988). Functional assessment staging (FAST). *Psychopharmacology Bulletin, 24*, 653–659.

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.

Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review, 12*, 573–604.

Shankle, W. R., Mangrola, T., Chan, T., & Hara, J. (2009). The CERAD wordlist memory performance index: Development and validation. *Alzheimer's & Dementia, 5*, 295-306.

Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E.-J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science, 32*, 1248-1284.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*, 34–50.

Spiegelhalter, D. J., Thomas, A., & Best, N. G. (2004). *WinBUGS Version 1.4 User Manual*. Cambridge, UK: Medical Research Council Biostatistics Unit.