Working Memory Impairments in Schizophrenics: A Bayesian Bivariate IRT Approach

Samantha Cook*
John Barnard
Yungtai Lo
Donald B. Rubin
Michael J. Coleman
Stephen Matthysse
Deborah L. Levy
Philip S. Holzman

February 28, 2002

*This paper has been submitted for publication in *Case Studies in Bayesian Statistics*, Volume 6.

Abstract

Several studies have shown that spatial working memory is impaired in schizophrenia patients. In our study, schizophrenia patients and normal controls participated in a memory test designed to measure both spatial and object working memory. The test items were designed to have differing levels of difficulty, making standard analyses inappropriate. The data were analyzed using a bivariate Bayesian Item Response Theory (IRT) model. Item response theory is a method for analyzing test scores in which the test items themselves are analyzed in addition to the test-takers' abilities. Analyzing the data in this way accounts for the fact that the questions were not all equally difficult, and also produces results which are more generalizable and less test-dependent. Earlier work in item response theory has focused on univariate abilities; here we extend these methods to include bivariate abilities. Using the Bayesian framework improves upon standard EM methods for IRT models by producing standard error estimates which more accurately represent uncertainty about the parameters. Using the Bayesian IRT model, we find that schizophrenia patients appear to be impaired on both object and spatial working memory, and that the degree of impairment is the same on average for both tasks.

1 Introduction

Schizophrenia is a psychiatric disorder characterized by delusions, hallucinations, thought disorder, and disturbance of emotion. Between 0.5 and 1 percent of people in the United States will be diagnosed with schizophrenia at some point in their lifetimes. Methods of treatment for schizophrenia patients include antipsychotic medication as well as cognitive-behavioral therapies and psychotherapy. Even with all of the available treatments, however, only about fifty percent of schizophrenia patients respond well to treatment (Hirsch and Weinberger, 1995).

Working memory is a limited capacity, short-term storage system for maintaining information as an active guide for adaptive actions. Several independent studies have reported that spatial working memory is impaired in schizophrenia patients (e.g. Keefe et al., 1995; Park and Holzman, 1992). Various response probes have been used to examine this deficit, including eye or hand movements in response to a visual target. To understand the role of this dysfunction in the pathophysiology of schizophrenia, it is important to assess whether the working memory impairment is specific to spatial information or involves other domains of working memory processes, such as object recognition. Our purposes here are to test whether object working memory is impaired in schizophrenia patients and, if so, to determine whether the two domains of memory are impaired to the same extent in schizophrenia patients.

Data from a test that measured object and spatial working memory abilities in schizophrenic and normal subjects are analyzed here under a bivariate Bayesian item response theory model. Item response theory (IRT; Lord, 1980) allows subjects' abilities to be analyzed while also analyzing and taking into account differences in the test items themselves. By studying these ability parameters we are able to quantify differences among subjects, in this case differences between normals and schizophrenics with respect to object and spatial working memory.

IRT models are used in a wide range of settings, particularly in educational testing. IRT is especially fitting in our situation because the test items were based on an existing paradigm and were known to have differing levels of difficulty, making standard analysis inappropriate. In addition, Bayesian analysis allows missing data to be dealt with in a straightforward and principled way, giving results that are not adversely affected by the fact that our data are incomplete. The analysis presented here is innovative in that the latent abilities under investigation are bivariate; to the best of our knowledge, previously published work in Bayesian IRT has used only univariate abilities.

Section 2 describes the experiment and presents summary statistics. Section 3 gives a brief introduction to and history of IRT models, and Section 4 discusses how we arrived at the IRT model used in this analysis. Sections 5 and 6 present our full Bayesian model and discuss its benefits. Results are given in Section 7, and Sections 8 and 9 present conclusions and final remarks.

2 Experiment

Participants were tested on a spatial and an object working memory task, based on the work of Smith et al. (Smith and Jonides, 1994; Smith et al., 1996; Smith et al., 1995). The tasks were administered using a personal computer equipped with a 17-inch monitor and a serial response box (Psychology Software Tools, Inc., 1996). To begin the experiment, a small fixation cross ($0.5°$ visual angle) appeared in the center of the screen with a prompt to press the space bar when ready. When the subject pressed the space bar, two irregular four-sided target polygons ($5.0°$ visual angle) appeared simultaneously on the screen for 500 milliseconds (see Figure 1). The fixation cross was present in the center of the screen throughout the trial. The two target polygons were randomly arranged in eight possible locations on an imaginary grid. The structure of this grid, chosen to inhibit mnemonic devices for remembering location, was not apparent to the participants. The irregular polygons had recognition and association values that were extremely low (Attneave and Arnoult, 1956; Vanderplas and Garvin, 1959) in order to inhibit mnemonic strategies for remembering the shapes of the polygons.

After the disappearance of the target polygons, the screen was blank for three seconds, except for the fixation cross. This period was the retention interval, which was followed by the presentation of a single probe polygon that required a response by the participant. In the spatial task, the participants were instructed to decide whether the probe polygon was in the same location on the screen as either of the two target polygons, regardless of shape. In the object task, the participants were instructed to decide whether the probe polygon was the same shape as either of the target polygons, regardless of location. The subject's response was made on the serial response box by pressing one of two buttons labeled "yes" (same) or "no" (different). The screen then cleared in preparation for the next trial. Thus, the stimuli were identical for the two tasks; only the instructions differed.

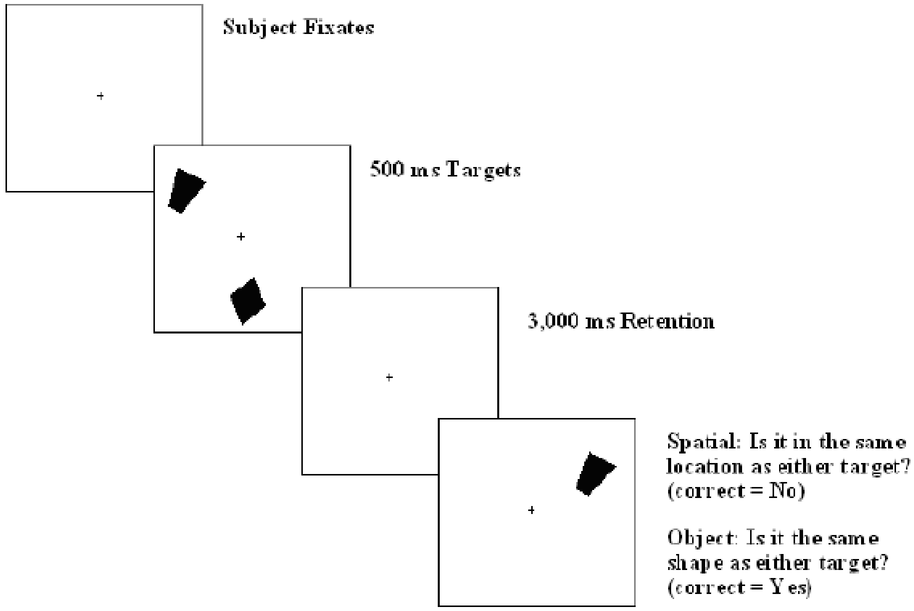

Figure 1: Test Schematic

The experiment called for 64 trials of the spatial working memory task and 64 trials of the object working memory task. For each task the trials were divided into four blocks of 16 trials each, with rest periods after each block. Each subject also completed several practice trials of both tasks before the experiment, whose purpose was to ensure that no learning effects appeared in the experimental data. In each task, half of the probes (32 trials) were *foils*, that is, probes that did not match either of the two target polygons. Half the foils in each task were hard foils and half were easy foils in the following sense. For the spatial task the hard foils were very near the location of one of the two target polygons but not in the exact location; the easy foils were far from both targets. For the object task, the hard foils were very similar to one of the target polygons but not an exact match; the easy foils were very different from both targets. The different types of questions were presented in randomized order, and the presentation order of all object and spatial tasks was counterbalanced.

The 28 schizophrenia patients tested were all outpatients at McLean Hospital in Belmont, MA, in various degrees of remission from psychosis. Schizophrenics and the 33 controls were matched on age and education, and controls who met DSM-IV criteria for a psychotic disorder or for schizotypal or schizoid personality disorders were excluded. Summary statistics are given in Table 1.

	Spatial Working Memory Percent Correct (sd)	Object Working Memory Percent Correct (sd)
Normal Controls	.925 (.046)	.808 (.061)
Schizopnrenia Patients	.863 (.091)	.727 (.071)

Table 1: Summary Statistics

From Table 1, it appears that schizophrenics performed worse than normal controls on both the object and the spatial working memory tasks. However, due to the foil structure of the experiment, different items were known to have differing levels of difficulty, and therefore comparing percentages taken over all items ignores potentially important information. In addition, it is unclear how to compare schizophrenics' impairments relative to normal controls on the two tasks since the controls themselves performed much better on the spatial task than on the object task. Furthermore, not all subjects completed all questions, so these summary statistics are based on "available cases." Approximately five percent of the observations (400 out of 7808) were missing. Nine schizophrenics and nine normal controls dropped out of the experiment; the normal controls had a total of 160 missing observations and the schizophrenics had a total of 240 missing observations. Since schizophrenics tended to drop out earlier and more often than normals, and dropped out largely after poor performance, any analysis that ignores the missing data is clearly questionable. By using item response theory and multiple imputation, we are able to improve upon these results.

3 Item Response Model

Dichotomous item response models generally deal with data of the form $\{X_{ij} : i = 1, ..., I; j = 1, ..., J\}$ where X_{ij} represents the response of subject i to item j, with $X_{ij} = 1$ if the item was answered correctly and $X_{ij} = 0$ if answered incorrectly. The item response model assumes that the probability a single test item is answered correctly is a non-decreasing function of the subject's latent ability θ. These functions are referred to as item response functions (IRFs) or item-characteristic curves (ICCs; Tucker, 1946) and generally have the form of cumulative distribution functions of random variables. The exact item response function is determined by the values of the vector of item parameters β. Given θ and β, we assume that X depends only on θ and β, so that the effects of any covariates are contained in the model parameters. In particular, a schizophrenic and a normal control with the same latent ability θ are expected to perform equally; given θ, the state of being schizophrenic or not has no impact on performance. Given θ, we assume local independence, i.e. that responses are independent within subjects, and also that responses are independent across subjects (Patz and Junker, 1999). The fact that the subjects completed practice trials before the experiment makes the local independence assumption more believable in this setting. Furthermore, the autocorrelations of individual subjects' responses were very low, which provides additional support for the assumption of local independence.

The first IRT model developed was the normal-ogive model (Lord, 1952; van der Linden and Hambleton, 1996) which uses the standard normal CDF as the item response function:

$$p(\theta_i) = P(X_{ij} = 1|\theta_i, \beta_j = (\beta_{1j}, \beta_{2j})) = \int_{-\infty}^{\beta_{2j}(\theta_i - \beta_{1j})} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz;$$

β_{1j} is referred to as the difficulty parameter for item j, and is equal to the value of θ for which $p(\theta) = 0.5$; β_{2j} is the discrimination parameter for item j, and is proportional to the slope of the IRF at the point where θ equals β_1.

Birnbaum replaced the normal-ogive model with the CDF of the logistic distribution, which can be expressed in closed form, as the item response function. The parameters of the two-parameter logistic model (2PL; Birnbaum, 1968) have the same graphical interpretation as in the normal-ogive model:

$$p(\theta_i) = \frac{1}{1 + \exp\{-\beta_{2j}(\theta_i - \beta_{1j})\}}.$$

The 2PL model can also be expanded to account for guessing.

The three-parameter logistic model (3PL; Lord, 1980) is obtained by assuming that subjects know the correct answer with the same probability as in the 2PL model, and if they do not know the correct answer will guess correctly with probability β_{3j}, which is referred to as the guessing parameter and is the left asymptote of the item response function:

$$p(\theta_i) = \beta_{3j} + \frac{1 - \beta_{3j}}{1 + \exp\{-\beta_{2j}(\theta_i - \beta_{1j})\}}.$$

The one-parameter logistic model, or Rasch model (Rasch, 1960) is also sometimes used for IRT analysis. The Rasch model assumes all discrimination parameters are equal to one and has the following item response function:

$$p(\theta_i) = \frac{1}{1 + \exp\{\theta_i - \beta_{1j}\}}.$$

Ability parameters θ_i are often unidimensional, but may be vectors. Our analysis expands upon standard IRT models by using bivariate ability parameters: each subject has a separate parameter for object and for spatial working memory ability, and the model allows for correlation between these two parameters.

4 Model Selection

The 3PL model seemed most appropriate for our data since it includes the guessing parameter. Because subjects were instructed to answer each question whether they thought they knew the correct answer or not, there is reason to believe that guessing took place and should thus be accounted for.

However, 3PL models are often unstable. Because each item has both a difficulty and a guessing parameter to be estimated, it can be difficult to distinguish between the two from subjects' responses. For example, if subjects did very well on a given question, it could be that the question was easy, or it could be that the question was difficult but easy to guess, and the IRT model may not differentiate well between these two possibilities (Patz and Junker, 1999). This instability was apparent in the results from the 3PL analysis. Posterior means for β_3, the guessing parameter, ranged from 0.35 to 0.9 over the different types of questions; there was an apparent positive correlation between the difficulty and guessing parameters; and small changes in the data produced large changes in the results. Given the test design, there is no reason to believe that the questions would be guessed correctly with probability as high as 0.9 or as low as 0.35. With the data set at hand, we simply did not have enough information to effectively estimate the guessing parameters.

Due to the apparent instability of the three parameter model, the guessing parameter was fixed at 0.5, which is the probability with which we expect each question to be guessed correctly. The questions were presented in random order, so even if subjects did not guess randomly, we would still expect them to guess correctly fifty percent of the time. Our final model is based on the 3PL model with 0.5 substituted for the guessing parameter:

$$p(\theta_i) = 0.5 + \frac{0.5}{1 + \exp\{-\beta_{2j}\theta_i + \beta_{1j}\}},$$

where θ_i refers to either object or spatial working memory ability for subject i, depending on the question type. This expression is slightly different from the logistic models presented earlier; the reparameterization seen here resulted in more stable computations. Thus, our model was chosen because it made scientific sense based on the design of the test items, and the issue of further model selection was not considered.

5 Bayesian Model

We fit the following Bayesian model to our data set:

Likelihood (Object Working Memory Items)

$$X_{ij} \sim \text{Bern}\left(0.5 + \frac{0.5}{1+\exp\{-\beta_{2j}\theta_{i,1} + \beta_{1j}\}}\right) \quad i = 1,\ldots,61,\ j = 1,\ldots,64.$$

Likelihood (Spatial Working Memory Items)

$$X_{ij} \sim \text{Bern}\left(0.5 + \frac{0.5}{1+\exp\{-\beta_{2j}\theta_{i,2} + \beta_{1j}\}}\right) \quad i = 1,\ldots,61,\ j = 65,\ldots,128.$$

Item Parameter Priors

$$\begin{array}{ll} \beta_{1j} \sim \text{N}(\eta_1, \tau_1) & j = 1,\ldots,128 \\ \beta_{2j} \sim \text{N}(\eta_2, \tau_2) : \beta_{2j} > 0 & j = 1,\ldots,128. \end{array}$$

Ability Parameter Priors

$$\begin{array}{lll} \theta_i \sim \text{N}_2(\mu_N, \Sigma_N) & i = 1,\ldots,33 & \text{(Normals)} \\ \theta_i \sim \text{N}_2(\mu_S, \Sigma_S) & i = 34,\ldots,61 & \text{(Schizophrenics).} \end{array}$$

Hyper-priors

$$\eta_1, \eta_2 \sim \text{N}(0, 10) \qquad \tau_1, \tau_2 \sim \text{Gamma}(8, 1)$$

$$\mu_S[k] \sim \text{N}(0, 10) \quad \Sigma_S[k,k] \sim \text{Gamma}(4, 1) \quad k = 1, 2$$

$$\rho_k \equiv \text{cor}(\mu_{\text{k}}[1], \mu_{\text{k}}[2]) \sim \text{Unif}(-1, 1) \quad k = \text{S}, \text{N}.$$

The means and variances of normal subjects' abilities were anchored at zero and one, respectively (i.e. $\mu_N = (0,0)$, $\Sigma_N[1,1] = \Sigma_N[2,2] = 1$). Since item and ability parameters are both latent, the model needs to be anchored in some way in order to be identifiable. Varying the forms and parameters of these prior and hyper-prior distributions (for example, replacing Normal distributions with t distributions, putting Normal priors on the Fisher's z-transformations (Fisher, 1921) of the correlations, and inflating the variances of the Normal and Gamma hyper-prior distributions) had negligible effects on the results, indicating that our joint prior distribution was appropriately non-informative.

6 Model Fitting

Parameter estimation for IRT models is a difficult problem because both item and ability parameters are generally unknown. If either item or ability parameters were fixed at known values, estimating the other group of parameters would be straightforward. Because of this, iterative algorithms are often used for parameter estimation. Joint maximum likelihood (JML) estimates can be obtained by iteratively maximizing the likelihood over either item or ability parameters while treating the other group of parameters as fixed at their current values (Baker, 1992). Marginal maximum likelihood (MML) estimates for item parameters may be calculated by treating the ability parameters as missing data and using the EM algorithm (Dempster, Laird, and Rubin, 1977; Bock and Aitken, 1981). Both of these ML strategies have significant drawbacks. JML estimates are generally not consistent since the number of parameters grows with the sample size (Bock and Aitken, 1981). The MML approach may be useful in certain situations, as IRT is often used in settings where the ability parameters are not of direct interest, for example when calibrating items for standardized tests. MML estimates of item parameters are consistent as the number of examinees, I, goes to infinity (Baker, 1992). The marginal ML approach does not, however, provide a way to properly estimate ability parameters. The MML estimates of item parameters can be treated as fixed and then used to estimate ability parameters; however, this will give standard errors for ability parameters that are too low since this method ignores the variability in the item parameter estimates (Tsutakawa and Soltys, 1988).

Our Bayesian model formulation avoids these problems by using the posterior distributions of the parameters to make inferences. Gibbs sampling (Geman and Geman, 1984) was used to obtain draws of the parameters. Because the posterior distribution is the joint distribution of all parameters, estimates based on the posterior distribution automatically take into account uncertainty about all other parameters. As long as we are not estimating joint functions of θ and β, MML consistency properties apply to moment estimates based on the posterior distribution (Patz and Junker, 1999).

Bayesian analysis and Gibbs sampling provide a convenient way to handle missing data using multiple imputation (Rubin, 1987). The missing values are simply treated as another group of parameters and drawn at each iteration. The parameter draws are then based on completed data. The missing values were imputed under the assumption that they were missing at random (MAR; Rubin, 1976). This condition implies that the probability that an observation is missing may be related to other observed quantities, but may not depend on the value of the missing observation or on any other missing quantities, conditional on observed values. When missing data are assumed to be missing at random, the missing values can be imputed from their conditional distributions, given the observed data and values of the parameters, with no additional modeling required (Rubin, 1976).

The MAR assumption seems appropriate in this case because subjects tended to drop out after poor performance, which was observed. If the MAR assumption were not correct in this situation, the Bayesian IRT model should give results that are conservative. Under the MAR assumption, if two subjects performed similarly up to a point where one dropped out and the other did not, the imputed values would be similar to the values observed in the subject with complete data. It is possible that subjects who dropped out, had they continued, would have done worse than predicted by MAR and therefore values imputed under MAR would show better performance than the true underlying values. In this case, schizophrenics would be imputed to appear less impaired than they actually are, since schizophrenics had more missing data and therefore more imputed values. Thus, by using this imputation model we should obtain results which are at worst conservative.

Gibbs sampling was carried out using winBUGS (Spiegelhalter, Thomas, Best, and Gilks, 2000) software, which is available free-of-charge at *http://www.mrc-bsu.cam.ac.uk/bugs*. WinBUGS imputes missing values from their conditional distributions. Since we assumed MAR, no extra work was required to handle the missing data.

7 Results

Posterior means, standard errors, and 95% intervals are given in Table 2, based on a sample of 50,000 draws from the posterior distribution of the parameters. Five chains were run from different starting values until the Gelman-Rubin statistic $\sqrt{\hat{R}}$ (Gelman and Rubin, 1992) was less than 1.03 for all parameters, and then a sample of 10,000 was kept from each chain. Gelman, Carlin, Stern, and Rubin (1995) recommend running all chains until $\sqrt{\hat{R}} < 1.2$ for all scalar quantities of interest and keeping draws after that point; $\sqrt{\hat{R}} < 1.03$ is therefore a conservative convergence criterion.

The main parameters of interest, the ability parameters of schizophrenics, indicate that schizophrenics are impaired on both spatial and object working memory. The posterior means for schizophrenics' ability parameters are -1.10 and -1.17 for object and spatial working memory, respectively, both of which are more than one standard deviation away from the means of the normal controls. Furthermore, neither of the two 95% intervals contains zero, indicating a significant difference in object and spatial working memory ability between normals and schizophrenics at the five percent level. The differences in abilities between schizophrenics and normals appear essentially identical on average for the two types of working memory studied here. The 95% posterior interval for the difference between the schizophrenics' ability parameters is centered at zero, indicating that the means of the two ability parameters are indistinguishable at the five percent level.

The variance of object working memory ability appears very similar in schizophrenics and normal controls. Spatial working memory appears much more variable in schizophrenics relative to normal controls. Although the 95% interval for the difference between the schizophrenics' two variance parameters contains zero, the posterior probability that the variance is higher for spatial than for object working memory is 0.86, which supports the proposition that spatial ability is more variable than object ability in schizophrenics relative to normal controls. Posterior means for the correlation between object and spatial abilities suggest that the two have very low correlation in normals, but are more correlated in schizophrenics. The 95% posterior intervals are wide and contain zero for both correlation parameters, indicating a high posterior uncertainty about this correlation. However, the posterior probability that this correlation is higher for schizophrenics than for normals is 0.75, which provides evidence that the correlation differs between the two groups.

VARIABLE	Mean	SE	2.5%	97.5%
$\Sigma_S[1,1]$	1.2500	0.6760	0.3710	2.9300
$\Sigma_S[2,2]$	2.5500	1.0900	0.9800	5.1300
$\Sigma_S[2,2] - \Sigma_S[1,1]$	1.3000	1.268	-1.044	4.017
$\mu_S[1]$	-1.1000	0.4090	-1.9900	-0.3820
$\mu_S[2]$	-1.1700	0.4160	-2.0300	-0.3970
$\mu_S[1] - \mu_S[2]$	0.0700	0.5144	-0.9476	1.084
ρ_N	0.1300	0.2960	-0.4660	0.6550
ρ_S	0.4700	0.2710	-0.1540	0.8960
$\rho_S - \rho_N$	0.3400	0.392	-0.45	1.01

Table 2: IRT Results

$\Sigma_S[1,1]$ and $\Sigma_S[2,2]$ are the variances of object and spatial ability, respectively, for schizophrenics. $\mu_S[1]$ and $\mu_S[2]$ are mean ability parameters for schizophrenics on object and spatial tasks, and ρ_N and ρ_S are the correlations between object and spatial abilities for normals and schizophrenics, respectively.

Model results for item parameters are less relevant to the scientific questions we hope to answer here and are omitted from this paper for brevity. Examining the item parameter estimates did confirm our prior beliefs that object items were harder than spatial items, and that within each type of task, hard foils were in fact harder than easy foils.

8 Scientific Conclusions

The results from this study provide strong evidence that working memory is an area in which schizophrenics show impairment relative to nonschizophrenics, with schizophrenics exhibiting significant deficits in both spatial and object working memory. The increased variance in schizophrenics' spatial working memory ability compared to normals is also consistent with the results from other tasks which measure specific schizophrenia deficits, such as eye-tracking (Levy, Holzman, Matthysse, and Mendell, 1993).

There is no discernible difference in the mean degree of impairment for schizophrenics on the two domains of working memory studied here, although the interval for the difference between schizophrenics' mean abilities is fairly wide. Further studies with more participants could give a more precise estimate for this difference. Because the two domains appear impaired to the same extent on average, the deficit could be due to non-specific interferences affecting both domains, such as poor motivation or impaired rule-following behavior. However, the difference in variances of the two abilities indicates that whatever is impairing working memory in schizophrenia patients has a different effect on the two domains and is therefore likely to be specific for the two realms of working memory.

The correlation estimates indicate that the two domains of working memory are nearly independent in normals and are more correlated in schizophrenics. Again, however, the posterior intervals are wide and further research with larger samples could refine these estimates. In addition, a test in which each item had both a spatial and an object component could provide better estimates of their correlation: in our study, object and spatial working memory abilities were not observed simultaneously.

We are currently developing experiments and analyses that disentangle problems involving attention and rule-following from specific working memory impairments in schizophrenia patients. In addition, further research will include clinically unaffected first-degree relatives of schizophrenics to explore whether there are genetic contributions to the functional impairments in working memory seen in the present study.

9 Final Statistical Remarks

Here we developed a bivariate Bayesian model for IRT analysis, implemented using Gibbs sampling to obtain draws from the posterior distribution of the parameters. Likelihood analysis was shown to be insufficient for estimating ability parameters even in simple settings. In our situation, likelihood analysis would be even more infeasible because the ability parameters are bivariate, making our model more complex than standard IRT models. With the availability of software such as winBUGS, Bayesian model fitting remained straightforward even for these complex models.

Acknowledgments

This research was supported by grants from the National Institute of Mental Health R01 MH31340, R01 MH49487, P01 MH31154, K05 MH01021, and the Roy Hunt Foundation. The authors would like to thank Courtney McCollum, Sapna Kundamal, Olga Krastoshevsky, and Professor William K. Estes for their valuable contributions to this paper.

References

Attneave, F., and Arnoult, M.D. (1956). Methodological considerations in the quantitative study of shape and pattern perception. *Psychological Bulletin* **53**, 452-471.

Baker, F.B. (1992). *Item Response Theory.* Marcel Dekker, Inc. New York.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord and M.R. Novick (Eds.), *Statistical theories of mental test scores*, 395-479. Reading, MA: Addison-Wesley.

Bock, R.D., and Aitken, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika* **46**, 443-459.

Dempster, A., Laird, N., and Rubin, D.B. (1977). Maximum Likelihood Estimation from Incomplete Data Using the EM Algorithm. *Journal of the Royal Statistical Society*, Series B, **39**, 1-38 (with discussion).

Fisher, R.A. (1921). On the probable error of a correlation coefficient... Reprinted in *Collected Papers of R.A. Fisher,* Vol. I (1971) (ed. J.H. Bennet). University of Adelaide Press, Adelaide, South Australia.

Gelman, A., Carlin, J., Stern, H., and Rubin, D.B. (1995). *Bayesian Data Analysis.* Chapman and Hall, London.

Gelman, A. and Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science* **7**, 457-511.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721-741.

Hirsch, S.R., and Weinberger, D.R. (1995). *Schizophrenia.* Blackwell Science Ltd. Oxford.

Keefe, R.S.E., Lees Roitman, S.E., Harvey, P.D., Blum, C.S., DuPre, R.L., Prieto, D.M., Davidson, M., and Davis, K.L. (1995). A pen-and-paper human analogue of a monkey prefrontal cortex activation task: spatial working memory in patients with schizophrenia. *Schizophrenia Research* **17**, 25-33.

Levy, D.L., Holzman, P.S., Matthysse, S., and Mendell, N.R. (1993). Eye tracking and schizophrenia: A critical perspective. *Schizophrenia Bulletin* **19**, 461-536.

Lord, F.M. (1952). A Theory of Test Scores. *Psychometric Monographs,* No. 7.

Lord, F.M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, New Jersey: Erlbaum.

Park, S. and Holzman, P.S. (1992). Schizophrenics show spatial working memory deficits. *Archives of General Psychiatry* **49**, 975-982.

Patz, R.J., and Junker, B.W. (1999). A Straightforward Approach to Markov Chain Monte Carlo Methods for Item Response Models. *Journal of Educational and Behavioral Statistics* **24**, 146-178.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. *Proceedings of the IV Berkeley Symposium on Mathematical Statistics and Probability* **4**, 321-333. Berkeley, CA: University of California.

Rubin, D.B. (1976). Inference and missing data. *Biometrika* **63**, 581-592.

Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys.* New York: Wiley.

Smith, E.E. and Jonides, J. (1994). Working memory in humans: Neuropsychological evidence. In M. Gazzaniga (Ed.), *The cognitive sciences* 1009-1020. Cambridge: MIT Press.

Smith, E.E., Jonides, J., and Koeppe, R.A. (1996). Dissociating verbal and spatial working memory using PET. *Cerebral Cortex* **6**, 11-20.

Smith, E.E., Jonides, J., Koeppe, R.A., Awh, E., Schumacher, E.H., and Minoshima, S. (1995). Spatial versus object working memory: PET investigations. *Journal of Cognitive Neuroscience* **7(3)**, 337-356.

Spiegelhalter, D.J., Thomas, A., Best, N.G., and Gilks, W.R. (2000). WinBUGS Version 1.3 User Manual. MRC Biostatistics Unit.

Tsutakawa, R.K., and Soltys, M.J. (1988). Approximation for Bayesian ability estimation. *Journal of Educational Statistics.* **13**, 117-130.

Tucker, L. (1946). Maximum validity of a test with equivalent items. *Psychometrika* **11**, 1-13.

van der Linden, W.J., and Hambleton, R.K. (Eds.) (1997). *Handbook of modern item response theory.* New York: Springer-Verlag.

Vanderplas, J.M., and Garvin, E.A. (1959). The association value of random shapes. *Journal of Experimental Psychology* **57(3)**, 147-154.