Published in final edited form as:

Exp Psychol. 2015 ; 62(3): 143–152. doi:10.1027/1618-3169/a000287.

Hierarchical multinomial modeling approaches: An application to prospective memory and working memory

Nina R. Arnold,

Institute for Experimental Psychology, Heinrich-Heine-Universität Düsseldorf, Germany;
Department of Psychology of the University of Mannheim, Germany

Ute J. Bayen, and

Institute for Experimental Psychology, Heinrich-Heine-Universität Düsseldorf, Germany

Rebekah E. Smith

Department of Psychology, The University of Texas at San Antonio, U.S.A.

Abstract

Hierarchical extensions of multinomial processing tree (MPT) models have been developed to deal with heterogeneity in participants or items. In this study, the beta-MPT model (J. B. Smith & Batchelder, 2010) and the latent-trait approach (Klauer, 2010) were used to estimate individual model parameters for prospective and retrospective components of prospective memory (PM), which requires remembering to perform an action in the future. The data from two experiments investigating the relationship between PM and working memory (R. E. Smith & Bayen, 2005, Experiment 1; R. E. Smith, Persyn, & Butler, 2011) were reanalyzed using the two hierarchical modeling approaches, both of which provide parameter estimates for individual participants. The results showed a positive correlation of the prospective component of PM with working-memory span and provide the first direct comparisons of the two hierarchical extensions of an MPT model.

Keywords

prospective memory; multinomial modeling; working memory; individual differences; hierarchical modeling

Multinomial processing tree (MPT) models, which are nonlinear statistical models designed to estimate probability parameters that measure latent cognitive processes from observed categorical data (Batchelder & Riefer, 1999), are usually tailored to a specific research paradigm and have been applied to many cognitive psychology paradigms (see Erdfelder et al., 2009). In traditional MPT modeling, observations are aggregated over participants and items and are assumed to be independent and identically distributed (i.i.d.), thus ignoring differences between participants and items (e.g., Batchelder & Riefer, 1999; Riefer &

Correspondence concerning this article should be addressed to Nina R. Arnold, Experimental Psychology, Schloss, Ehrenhof Ost Room EO 266, D-68131 Mannheim, Germany. niarnold@mail.uni-mannheim.de. Phone number: +49- 621-181-2102.

Part of this work was part of the first author's doctoral dissertation research conducted at the Heinrich-Heine-Universität Düsseldorf under the supervision of the second author. Part of this research was presented at the "Tagung experimentell arbeitender Psychologen" (TeaP), Mannheim (2012), and the "International Conference on Cognitive Modeling" (ICCM), Berlin (2012).

Batchelder, 1988). However, J. B. Smith and Batchelder (2008) showed that this assumption of parameter homogeneity is often violated - even for carefully constructed item pools and relatively homogeneous groups of participants. This can result in biased parameter estimates (Klauer, 2006; J.B. Smith & Batchelder, 2008, 2010).

Recent attempts to deal with parameter heterogeneity in MPT modeling use a latent-class approach (Klauer, 2006) or Bayesian hierarchical modeling (e.g., Klauer, 2010; Matzke, Dolan, Batchelder, & Wagenmakers, 2013; J. B. Smith & Batchelder, 2010). The latent-class approach uses a finite-mixture model and assumes that participants fall into a finite number of latent classes. Participants in the same class are assumed to have the same parameters. Thus, this approach solves the problem of heterogeneity by dividing up the participants into smaller homogenous groups.

Hierarchical modeling approaches provide individual parameter estimates for each participant by defining an individual MPT model for each participant. The individual parameter estimates for these individual MPT models are assumed to arise from a common distribution. This hyperdistribution allows researchers to calculate group mean and variance for each parameter, as can be done for parameter estimates obtained in traditional MPT modeling. The participants' individual parameter estimates can be correlated with other performance measures to investigate how individual differences explain performance on cognitive tasks. In this study, we take advantage of hierarchical modeling to correlate individual model-based parameter estimates of different components of prospective memory (PM) with individual measures of working memory (WM) span.

An alternative single-level approach of calculating separate independent models for each participant to correlate individual model parameters with extraneous variables (e.g., Yonelinas, Regehr, & Jacoby, 1995), is often not feasible because some response categories are sparsely filled or empty and, therefore, the resulting parameter estimates lack reliability (Erdfelder et al., 2009). In addition, a correctly specified hierarchical model provides more accurate estimates by correcting for outliers (e.g., Rouder & Lu, 2005).

Prospective Memory

PM tasks require remembering to perform an action when there is a delay between forming the intention and the opportunity to carry out the intended action (for an overview, see McDaniel & Einstein, 2007). For event-based PM tasks the intended action must be completed when a certain event occurs; for example, mailing a letter when you see a mailbox. Successful PM performance involves a prospective component of remembering *that* you have to do something and a retrospective memory component of remembering *what* you wanted to do and *when*. R. E. Smith and Bayen (2004) proposed a MPT model of event-based PM that allows researchers to disentangle the prospective and the retrospective components. In this study we investigate the usefulness of hierarchical MPT modeling for research on individual differences in PM. We reanalyzed data from R. E. Smith et al. (2011) and R. E. Smith and Bayen (2005) to examine how individual differences in WM are related to the different components of PM. We first describe the MPT model of event-based PM. Second, we describe two hierarchical MPT modeling approaches, namely the latent-trait

approach and the beta-MPT approach. Third, we review previous findings on the relationship of PM and WM including the findings of the original studies. Finally, we present our reanalysis of the data and discuss the results. Specific predictions, based upon current PM theories, for the relationship between WM and each of the components of PM are discussed in the introduction to each experiment.

The MPT model of Event-Based Prospective Memory

Because PM tasks often involve interrupting a current activity, in standard laboratory tests of PM participants perform an ongoing task (e.g., lexical decision) that is interrupted to perform the PM task (e.g., press the “F1” key when they see the syllables “low” or “per” during lexical decisions; R. E. et al., 2011). The model (R. E. Smith & Bayen, 2004; see Figure 1) is designed for a binary ongoing task, such as lexical decisions with the response options “word” and “non-word.” PM targets can occur on word trials and non-word trials. This produces four trial types, represented by separate trees in Figure 1: (1) a PM target on a word trial, (2) a PM target on a non-word trial, (3) a word trial without a PM target, and (4) a non-word trial without a PM target.

The ability to distinguish between words and non-words is captured by Parameters C_1 and C_2 . On word trials (first and third tree in Figure 1), C_1 is the probability that the participant detects that the string is a word. On non-word trials (second and fourth tree), C_2 is the probability of recognizing that the string is not a word. In all trees, P is the probability of remembering that there is an additional task (i.e., the prospective component). In all trees, M is the probability of successfully discriminating between PM targets and non-targets (retrospective recognition component). On target trials (first and second tree), correct discrimination results in a PM response. On non-target trials (third and fourth tree), correct discrimination results in an ongoing-task response.¹

The model also includes guessing probabilities. If participants are unable to discriminate between PM targets and non-targets, they must guess whether the string is a PM target. Parameter g is the probability of guessing that the trial includes a PM target, resulting in a PM response. $1 - g$ is the probability of guessing that the trial does not include a PM target, resulting in a response to the ongoing lexical decision task. If participants do not remember that there is a PM task (with probability $1 - P$), this will also result in an ongoing-task response. If participants respond to the ongoing task but do not detect that a string is a word (with probability $1 - C_1$) or do not correctly detect that a string is a non-word (with probability $1 - C_2$), they can guess with probability c that the letter string is a word, and with probability $1 - c$ that it is not a word.

Because the model with seven free parameters is not identifiable, theoretically based parameter restrictions are imposed (Smith & Bayen, 2004). The four free parameter (P , M , C_1 , C_2) model is identifiable, has been validated, and fitted the aggregated response

¹Note that the retrospective component M in the model captures the recognition of the PM targets (i.e., *when* to perform the action) and not the recollection of the PM key (i.e., *what* action to perform). However, in the studies that we reanalyzed the PM action was very easy to remember (and participants who nonetheless did not remember it were excluded) such that retrospective memory for the action would not influence PM performance.

frequencies of participant groups in several studies (Horn, Bayen, R. E. Smith, & Boywitt, 2011; Pavawalla, Schmitter-Edgecombe, & R. E. Smith, 2012; Schnitzspahn, Horn, Bayen, & Kliegel, 2012; R. E. Smith & Bayen, 2005, 2006; R. E. Smith, Bayen, & Martin, 2010; R. E. Smith, Horn, & Bayen, 2012; R. E. Smith & Hunt, 2014; R. E. Smith et al., 2011; R. E. Smith, McConnel Rogers, McVay, Lopez, & Loft, 2014).

Hierarchical MPT Modeling

Two hierarchical models, differing mainly in the assumed underlying parameter distributions, have been proposed to handle participant heterogeneity in MPT models. The beta-MPT model (J. B. Smith & Batchelder, 2010) assumes that each participant's parameters are drawn independently from a multivariate distribution consisting of independent marginal beta distributions. The advantage of the beta distribution is that it lies in the interval (0,1) and, thus, in the natural parameter space of the model parameters which represent probabilities. The beta distribution is defined by the parameters (α , β) and is unimodal for $\alpha, \beta > 1$, U-shaped for $\alpha, \beta < 1$, and takes uniform shape for $\alpha, \beta = 1$ on the unit interval. J. B. Smith and Batchelder applied the beta-MPT approach to the MPT model of pair clustering (Batchelder & Riefer, 1986) and provided a WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000) implementation. We previously applied the beta-MPT approach to the MPT model of source monitoring (Arnold, Bayen, Kuhlmann, & Vaterrodt, 2013) and to the MPT model of PM (Arnold, Bayen, & Böhm, in press).

The latent-trait approach (Klauer, 2010) assumes that each participant's parameters are drawn from a multivariate normal distribution of probit transformed parameters. The probit link transforms parameters from the interval (0,1) to the real line. The latent-trait approach has the advantage of accounting for correlations between the parameters of a model. Matzke et al. (2013) provided a WinBUGS implementation for the latent-trait version of the pair clustering model, which we have adapted for the MPT model of PM.

Both approaches rely on Bayesian modeling, in which initial beliefs are represented by treating the parameters as random variables. A prior distribution, either very vague or more concrete depending on prior knowledge about the parameter distribution, is specified before starting the analysis. A posterior distribution given the data is calculated using Bayes' theorem. As the amount of data increases (e.g., due to more participants or more observations per participant), the effect of the choice of prior distribution diminishes. Bayesian confidence intervals (BCI) reflect the range of values in the posterior distribution where the true parameter is found with certain confidence (given the data and the prior).

Both hierarchical MPT models are computed using Markov Chain Monte Carlo (MCMC) methods. An MCMC sample consists of a large number of draws from the target distribution, from which one can obtain knowledge about the properties of the model parameters. The frequently used Gibbs sampler is a method of producing (serially dependent) draws from the target posterior distribution, under certain conditions, that is implemented in a number of software packages (e.g., WinBUGS, Lunn, et al., 2000).

Drawing statistical inference from MCMC chains requires evaluation of convergence (i.e., the chain reaches a stationary distribution), which can be assessed using the $\hat{R}$ statistic

(Gelman & Rubin, 1992). $\hat{R}$ compares variances within and between the chains and will be close to 1 under convergence. Because early draws often have poor convergence, a burn-in period is discarded and not used for parameter estimates or convergence estimation. Successful implementation of an MCMC chain results in a sample from the full posterior distribution, thus we can calculate statistics about all basic and hyperparameters and estimate the uncertainty of the estimates by using BCI or the MCMC error.

Limitations of Prior Investigations of PM and WM Using the MPT Model of PM

R. E. Smith and Bayen (2005) and R. E. Smith et al. (2011) used the MPT model of event-based PM to assess the relationship between the PM components and WM, but these studies have limitations. First, they used traditional MPT modeling based on data aggregated over participants, which may lead to biased estimates. Second, R. E. Smith et al. (2011) used an extreme-group design, in which they compared the 25% of the participants with the highest WM span scores to the 25% with the lowest WM span scores, thus excluding half of the data. We used beta-MPT (J. B. Smith & Batchelder, 2010) and latent-trait (Klauer, 2010) re-analyses of the original data from Experiment 1 by R. E. Smith and Bayen (2005) and R. E. Smith et al. (2011) to address these limitations.

Reanalysis of R.E. Smith and Bayen (2005)

In Smith & Bayen's (2005) Experiment 1, 20 adults completed an ongoing sentence verification task, a PM task of pressing the F1 key when one of four target words appeared, and a counting span test (Conway, 1998) as a measure of WM span. The data were aggregated within WM span groups formed via median split. Current theories of PM predict a correlation between the WM and the prospective component of PM. The PAM (Preparatory Attention and Memory) theory (Smith, 2003) proposes that the prospective component will be resource demanding and will therefore rely on WM. The multiprocess view (MPV; McDaniel & Einstein, 2007) also predicts a reliance on WM because multiple PM targets may encourage reliance on non-automatic processes for retrieving the intention. The retrospective component measures recognition memory, which more heavily depends on automatic familiarity processes (e.g., Wixted, 2007), thus the retrospective component is less likely to be related to WM. As expected, participants in the higher-WM group had a greater probability of remembering that they had to perform the PM task (prospective component; Parameter *P*), than participants in the lower-WM group, but the groups did not differ in the retrospective-memory component (Parameter *M*). For the following analyses, we hypothesized that the *P* parameter would correlate with WM span scores.

Reanalysis with Hierarchical MPT Models

Using the medians of the posterior distributions of the individual parameter estimates resulting from the beta-MPT and the latent-trait analyses, we computed correlations with WM. We estimated the parameters using Bayesian modeling. For consistency, we also report Bayesian statistics for the correlations. For a discussion of the advantages of Bayesian testing over classical frequentist null hypothesis testing, see, for example, Wagenmakers

(2007). The Bayes factor (BF ; e.g., Jeffreys, 1961; Morey & Wagenmakers, 2014; Wetzels & Wagenmakers, 2012) denotes the probability of the data under the null hypothesis relative to the alternative hypothesis. We used the Bayesian hypothesis test for correlations presented by Wetzels and Wagenmakers (2012). To calculate the BF , the correlation is conceptualized as a comparison between two regression models, namely one that uses r as regression coefficient and one that does not include a regression coefficient. The calculation uses a Jeffreys-Zellner-Siow prior (Liang, Paulo, Molina, Clyde, & Berger, 2008). However, these BF s are two-sided. Since we had clear expectations about the direction of the effect, we used order-restricted BF s (Morey & Wagenmakers, 2014). According to Jeffreys (1961), the BF must be greater than 3 to indicate substantial evidence for the H_1 , in our case for a correlation. A BF smaller than $1/3$ indicates substantial evidence for the absence of a correlation (H_0).

We report “traditional” p values in addition to BF s. With 20 participants and a (one-tailed) α of .05, the power to detect a medium effect of $r = .30$ and a large effect of $r = .50$ (Cohen, 1992) was .39 and .80, respectively (G*Power, Faul, Erdfelder, & Buchner, 2007). We computed correlations with WM span for the prospective component P and the retrospective component M , and the traditional measure of PM performance $PM\ hits$ (proportion of PM targets correctly responded to), see Table 1. We also report the parameter estimates for aggregated data estimated with MultiTree (Moshagen, 2010) in Table 2. For all tests, we used $\alpha = .05$.

In the latent-trait approach external covariates (e.g., WM scores) can be included directly into the model to compute a posterior distribution for the correlation (Klauer, 2010). This is not possible in the beta-MPT approach. Nevertheless, it is possible to compute the correlation on each iteration, thereby yielding a posterior distribution. For both approaches and all re-analyses, we present the means and BCIs for these correlation parameters in Table 1. If the BCI does not include zero there is evidence in favor of a correlation.

The beta-MPT approach—For beta-MPT modeling, we used a uniform distribution between 1 and 5000 as prior for each α_s and β_s of each parameter θ_s . This prior is very vague because of its wide range. However, α and β are greater than 1 which ensures that the beta distribution is bell shaped. We conducted 100,000 iterations with a thinning rate of 10, and discarded the first half of the iterations as burn-in period. For all parameter estimates, $\hat{R} < 1.05$. The posterior distributions of the hierarchical beta distributions are shown in Table 2.

The deviance information criterion (DIC) is a Bayesian method for model comparison similar to AIC or BIC . Lower DIC indicates better model fit. The DIC for the beta-MPT was 391.38. Additionally, to check for the recovery of the mean category frequencies, we calculated Klauer's (2010) test statistic T_1^2 for data generated by the hyperparameters as well as by the individual parameter estimates. We compared the statistics for 2000 draws from the posterior distribution and the observed data and found that 11.90 % of the draws

²For the reanalysis of Smith and Bayen's (2005) data, some draws for both hierarchical approaches produced empty categories for all participants. For these draws T_1 is not defined because of division by zero. In these cases, we defined $T_1 = 0$ because the nominator is zero.

generated from the hyperparameters and 27.90 % of the draws generated from the individual parameters had values larger than the observed data indicating that the model does describe the data adequately (Matzke et al., 2013). A graphical comparison of the predicted with the observed frequency data is presented in Online Supplement 1. The correlations along with means and standard deviations of the WM span measures are shown in Table 1. We found a significant correlation between WM span and PM hits, $r = .40$, $p = .04$ (one-tailed), and between WM span and the prospective component P , $r = .39$, $p = .04$ (one-tailed). The Bayes factors (BF) were 1.50 and 1.42 indicating anecdotal evidence to support a correlation. Additionally, the BCI for the posterior correlation did not include zero supporting the evidence. WM span and the retrospective-memory component M did not correlate significantly, and the $BF = 0.38$ indicated anecdotal evidence for the absence of a correlation. The BCI for the posterior correlation included zero. Thus, our results concur with those of Smith et al. (2011) who reported that participants with higher WM span scores had a higher prospective component P , but that there was no difference in the retrospective-memory component M .

The latent-trait approach—For the latent-trait approach, we used multivariate normal distributions with $\mu_{\mu_s} = 0$ and $\sigma_{\mu_s}^2 = 1$ as prior for the population level μ_s for each parameter s . This corresponds to a uniform distribution on the probability scale (Matzke et al., 2013; Rouder & Lu, 2005). We conducted 100,000 iterations with a thinning factor of 10 and discarded the first half of the iterations as burn-in period. For all parameter estimates, $\hat{R} < 1.05$. The posterior distributions are shown in Table 2.

DIC for the latent-trait approach was 321.07. We found that 27.40 % of 2000 draws generated from the hyperparameters of the posterior distribution and 43.95 % of draws generated from individual model parameters had values larger than the observed data indicating that the model adequately describes the data (Online Supplement 1 shows a graphical comparison). We show correlations with the resulting individual parameter estimates of the latent-trait approach in the lower half of Table 1. We found a similar pattern as with the beta-MPT approach. That is, we found a significant correlation between WM span and the prospective component P , $r = .41$, $p = .04$. The BF is 1.57, indicating anecdotal evidence for a correlation. However, the BCI of the correlation included zero and thus does not favor a correlation. WM span and the retrospective-memory component M did not correlate, $r = .17$, $p = .24$, $BF = 0.33$, and the BCI included zero.

Comparison of the approaches—We compared the parameter estimates obtained with both approaches. Table 2 shows that the BCI for the population level parameters overlap. Estimates are higher with the latent-trait approach than with the beta-MPT approach with the parameter estimates of the traditional approach lying in-between. However, on an individual level, the parameter estimates did not differ significantly, all $p > .05$, and all $BFs < 2/3$. No approach showed consistently larger correlations than the other approach (see Online Supplement 2 for parameter correlations). The parameter estimates obtained with both methods showed very high correlations, all $r > .98$, all $p < .001$, all $BF > 100$. In terms of DIC and T_1 , the latent-trait approach fit the model better than the beta-MPT approach.

Discussion—The correlations between the parameter estimates of the different approaches are very high, and the general pattern of results is the same for both approaches. The beta-MPT population level parameters were smaller than the traditional MultiTree parameter estimates which in turn were smaller than the latent-trait approach population level parameter estimates. However, the *BCIs* overlapped for all parameters. There was no advantage of either approach.

The results of our re-analyses concur with R. E. Smith and Bayen's (2005) findings. To yield higher power, we next reanalyzed the study by R. E. Smith et al. (2011) with 413 participants. Additionally, in this study, a different WM-span measure and different ongoing and PM tasks were used, thus giving us the opportunity to investigate if results replicate across different measures of the same constructs.

Reanalysis of R.E. Smith, Persyn, and Butler (2011)

Participants ($N=413$) completed an ongoing lexical decision task and a PM task of pressing the F1 key when the syllables “low” or “per” appeared. This was a non-focal PM task because the detection of the syllables required different processes than the ongoing lexical decision task, and therefore the prospective component was expected to rely on non-automatic processes according to the MPV (McDaniel & Einstein, 2007) and according to PAM (Smith, 2003). In addition, participants completed a symmetry span test (Unsworth, Redick, Heitz, Broadway, & Randall, 2009) as a measure of WM span. Only participants with the lowest 25% and the highest 25% of the span scores were included in the extreme groups analysis. Participants in the higher WM group had a higher probability of remembering that they had to perform the PM task (Parameter *P*) than did participants in the lower WM group. The retrospective-memory component (Parameter *M*) did not differ between groups. For our analyses, which included data from all 413 participants, we hypothesized a correlation of the *P* parameter with WM span scores.

Reanalysis with Hierarchical MPT Models

Again, we computed correlations of WM span with PM components using the individual parameter estimates resulting from the beta-MPT and the latent-trait analyses as well as the *BCIs* for the posterior distributions of the correlation (Table 1). In Table 2, we report parameter estimates obtained with the traditional approach based on aggregated data, and with both hierarchical approaches.

The beta-MPT approach—For the beta-MPT modeling, we used a uniform distribution between 1 and 5000 as prior for each α_s and β_s of each parameter θ_s . Due to slower convergence, we conducted 500,000 iterations with a thinning rate of 500, and discarded the first half of the iterations as burn-in period. For all parameter estimates, $\hat{R} < 1.05$.

DIC was 8553.17. Klauer's (2010) test statistic T_1 with 2000 draws generated from the hyperparameters and the individual parameters of the posterior distribution revealed that 26.25 % and 35.15 % of the draws, respectively, had values larger than the observed data indicating that the model adequately describes the data (Matzke et al., 2013).

We found a significant correlation of WM span with PM hits, $r = .11$, $p = .01$, $BF = 1.12$ and with the prospective component P , $r = .15$, $p < .01$. The BF was 7.33 indicating substantial support for a correlation. The BCI did not include zero. WM span and the retrospective-memory component M did not correlate significantly, and the BF did not support a correlation, and the BCI included zero. Thus, our results concur with Smith et al.'s (2011) who reported a higher prospective component P in participant with higher WM span, but no difference in the retrospective-memory component M .

The latent-trait approach—For the latent-trait approach, we used multivariate normal distributions with $\mu_{\text{us}} = 0$ and $\sigma_{\text{us}}^2 = 1$ as prior for the population level μ_s for each parameter s . As the latent-trait approach showed faster convergence than the beta-MPT approach, we conducted 100,000 iterations with a thinning factor of 100 and discarded the first half of the iterations as burn-in period. For all parameter estimates, $\hat{R} < 1.05$.

DIC was 7393.41. The model describes the data well. That is, 49.90 % of the draws generated from the hyperparameters of the posterior distribution and 43.75 % of draws generated from individual model parameters had T_1 values larger than the observed data (Online Supplement 1 shows a graphical comparison). We found a similar pattern as with the beta-MPT approach: the correlation between WM span and the prospective component P was significant, $r = .14$, $p < .01$. The BF is 5.41, indicating evidence for the presence of a correlation. The BCI did not include zero. WM span and the retrospective-memory component M did not correlate, $r = .04$, $p = .40$, $BF = 0.07$, and the BCI included zero.

Comparison of the approaches—Table 2 shows that the BCI for the population level parameters do not overlap. Estimates are higher with the latent-trait approach than with the beta-MPT approach, with the traditional parameter estimates lying in-between. However, on an individual level, the parameter estimates do not differ significantly for C_1 , C_2 , and P , all $ps > .05$, and all $BFs < 1/8$. Only the retrospective component M shows higher beta-MPT estimates than latent-trait estimates, $t(412) = 7.47$, $p < .001$, $BF > 1,000,000$. Note that the difference is opposite to the difference for population level parameters. No approach showed consistently larger correlations than the other approach (see Table 1). The correlations between the parameters obtained with both methods were very high, all $r > .91$, all $p < .001$, all $BF > 100$ (see Online Supplement 3 for parameter correlations). Correlations between the individual model parameters within each approach were higher in the latent-trait approach than in the beta-MPT approach. In terms of DIC , the beta-MPT approach fit the model better than the latent-trait approach. Klauer's (2010) test statistic T_1 showed an advantage for the latent-trait approach.

Discussion—In this reanalysis, the correlations between the parameter estimates of the different approaches was again very high, and the general pattern of results was the same for both approaches. The beta-MPT population level parameters were smaller than the traditional MultiTree parameter estimates which were smaller than the latent-trait approach population level parameter estimates. That is, we found the same pattern as in the reanalyses of R. E. Smith and Bayen's (2005) data. This time, however, the $BCIs$ did not overlap. The results concur with the original findings (Smith et al., 2011).

General Discussion

We investigated individual differences in prospective and retrospective components of PM and their relationship with WM by reanalyzing data from R. E. Smith and Bayen (2005) and R. E. Smith et al. (2011). With the hierarchical MPT modeling framework, we were able to avoid biased parameter estimates and to incorporate the data from all participants. Most importantly, we were able to analyze the data on an individual basis. WM span correlated with the prospective component of PM, but not with the retrospective memory component.³ This is in line with current theories of PM (Smith, in press) that predict that the prospective component of the PM tasks used in these studies are resource demanding and, therefore, require WM capacity. Although our reanalysis has a number of advantages, limitations remained for the reanalyses of R. E. Smith and Bayen's (2005) data in that sample sizes were very small leading to limited power to detect a medium-effect sized correlation, and the retrospective-component parameter M was very high and estimated with the latent-trait approach even approached 1. Therefore, we cannot rule out that restriction of range resulting from a ceiling effect may have impeded the emergence of significant correlations. However, the study by R. E. Smith et al (2011) had a larger sample size, and the estimates of the retrospective component are much lower. This experiment also shows the expected pattern of results.

This is the first time the beta-MPT and the latent-trait approach have been compared with the same data. Both approaches showed the same pattern of correlations and neither approach showed consistently larger correlations with WM span than the other approach. Correlations between the individual parameter estimates (e.g., correlation of P with M) within each approach were higher in the latent-trait approach than in the beta-MPT approach. The latent-trait approach incorporates correlations already in the prior distributions. Of course, it is possible to calculate correlations between the posterior parameter estimates for both approaches. The advantage of the latent-trait approach is that the correlations between parameters are explicitly modeled and it is therefore possible to put a strong prior on the correlations if a correlation of a specific magnitude is expected. However, in our case, we had a uniform prior on the parameter correlations assuming that all correlation values were equally likely. Even if we assume a priori that the parameters are uncorrelated, sufficiently informative data will overwhelm the prior, and we can compute correlations from the posteriors. However, for small datasets this may lead to biased parameter correlations (e.g., Rouder et al., 2007).

According to *DIC*, there was no clear advantage of either approach, and the results were similar. For the study by R. E. Smith and Bayen (2005), the latent-trait approach showed better model fit, and for the study by R. E. Smith et al. (2011), the beta-MPT showed better fit. Klauer's (2010) test statistic T_1 indicated better parameter recovery for the latent-trait

³We also reanalyzed the data from Smith and Bayen's (2005) second experiment and, contrary to expectations, we did not find correlations of model parameters with WM span. However, the sample size was very small, limiting the statistical power to detect medium-size effects to less than .50. Likewise, for the Bayesian approach, we did not have enough information to report evidence in favor of either the absence or the presence of correlations (except for the correlation of the parameter estimates obtained with both methods), and the BCIs for the posterior correlations were extremely large. Because of the lack of clear results, likely due to limited sample size, we have omitted a detailed discussion of these re-analyses.

approach for both studies. However, all models described the data reasonably well. We, thus, have no straightforward recommendation in favor of either approach. However, incorporating correlations between parameters seems reasonable in many cases, because basic cognitive abilities or motivational variables may simultaneously affect several model parameters. An example by Matzke et al. (2013) is that two cognitive abilities that both reflect aspects of memory retrieval are likely related. The possibility to incorporate between-parameter correlations, thus, speaks in favor of the latent-trait approach.

Both approaches allow for either participant or item heterogeneity. Matzke et al. (2013) introduced the crossed-random effects approach which is an extension of the latent-trait approach and incorporates participant as well as item heterogeneity. It assumes that participant and item effects combine additively on the probit scale (e.g., Rouder & Lu, 2005; Rouder et al., 2007). In this approach, participant variability is assumed to follow a multivariate normal distribution, whereas item heterogeneity is assumed to follow independent normal distributions. However, the approach is not adaptable to MPT models such as the model of event-based PM, where parameter constraints are required between different model trees and items appear in only one of the model trees (Matzke et al., 2013).

PM tasks usually have few targets, and error rates (in terms of incorrect PM responses on non-target trials) are often low. Therefore, a general problem of parameter estimation for the MPT model of event-based PM are low frequencies in those response categories that are relevant for the estimation of the parameters of greatest interest, that is, the prospective component P and the retrospective component M . To use a χ^2 -statistic such as G^2 to measure goodness of fit, the frequencies in each of the categories should be 5 at least (e.g., Hays, 1994), which is often not the case for individuals.

In Bayesian hierarchical modeling, the precision of the parameter estimates is indicated by the BCI . While the ongoing-task parameters C_1 and C_2 have very small BCI , the BCI for the prospective component P and the retrospective component M are wider. Still, they are reasonably accurate. An advantage of BCI compared to classical confidence intervals (CIs) is that they always stay within the boundaries of the parameter space. CIs in MPT modeling can exceed 1 or even be negative depending on the data structure.

The present application showed that Bayesian hierarchical models are very useful for applying MPT modeling to an individual-differences approach. Using hierarchical MPT modeling, we were able to examine correlations between WM span with different cognitive components that underlie PM performance. WM span was related to the prospective component P supporting the theoretical notion that monitoring for the occurrence of PM target taxes WM, at least in cases such as these in which the PM task is non-focal or involves multiple target events. The investigation of possible contributions of other individual-difference variables to PM via hierarchical MPT modeling is a fruitful venue for the future.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The research was supported in part by Grant AG034965 from the National Institute on Aging of the US National Institutes of Health. We thank Karl Christoph Klauer for help with incorporating external covariates in the latent-trait approach.

References

- Arnold NR, Bayen UJ, Böhm MF. Is prospective memory related to depression and anxiety? A hierarchical MPT modelling approach. *Memory*. in press.
- Arnold NR, Bayen UJ, Kuhlmann BG, Vaterrodt B. Hierarchical modeling of contingency-based source monitoring: A test of the probability-matching account. *Psychonomic Bulletin & Review*. 2013; 20:326–333. [PubMed: 23224987]
- Batchelder WH, Riefer DM. The statistical analysis of a model for storage and retrieval processes in human memory. *British Journal of Mathematical & Statistical Psychology*. 1986; 39:120–149.
- Batchelder WH, Riefer DM. Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*. 1999; 6:57–86. [PubMed: 12199315]
- Brewer GA, Knight JB, Marsh RL, Unsworth N. Individual differences in event-based prospective memory: Evidence for multiple processes supporting cue detection. *Memory & Cognition*. 2010; 38:304–311. [PubMed: 20234020]
- Cherry KE, LeCompte DC. Age and individual differences influence prospective memory. *Psychology and Aging*. 1999; 14:60–76. [PubMed: 10224632]
- Cohen J. A power primer. *Psychological Bulletin*. 1992; 112:155–159. [PubMed: 19565683]
- Conway, ARA. Counting Span Task [Computer program]. Chicago, IL: University of Illinois at Chicago; 1998.
- Craik FIM, Govoni R, Naveh-Benjamin M, Anderson N. The Effect of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*. 1996; 125:159–180. [PubMed: 8683192]
- Einstein GO, McDaniel MA. Prospective memory and what costs do not reveal about retrieval processes: A commentary on Smith et al. (2007). *Journal of Experimental Psychology Learning, Memory, and Cognition*. 2010; 36:1082–1088.
- Einstein GO, McDaniel MA, Thomas RA, Mayfield S, Shank H, Morrisette N, Breneiser J. Multiple processes in prospective memory retrieval: Factors determining monitoring versus spontaneous retrieval. *Journal of Experimental Psychology: General*. 2005; 134:327–342. [PubMed: 16131267]
- Erdfelder E, Auer TS, Hilbig BE, Aßfalg A, Moshagen M, Nadarevic L. Multinomial processing tree models: A review of the literature. *Journal of Psychology*. 2009; 217:108–124.
- Faul F, Erdfelder E, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*. 2007; 39:175–191. [PubMed: 17695343]
- Fernandez MA, Moscovitch M. Divided attention and memory: evidence of substantial interference effects at retrieval and encoding. *Journal of Experimental Psychology: General*. 2000; 129:155–176. [PubMed: 10868332]
- Gelman A, Rubin D. Inference from iterative simulation using multiple sequences. *Statistical Science*. 1992; 7:457–472.
- Hays, WL. *Statistics*. 5th. Fort Worth, TX: Harcourt Brace College Publishers; 1994.
- Horn SS, Bayen UJ, Smith RE, Boywitt CD. The multinomial model of prospective memory: Validity of ongoing-task parameters. *Experimental Psychology*. 2011; 58:247–255. [PubMed: 21106476]
- Jeffreys, H. *Theory of probability*. Oxford: UK: Oxford University Press; 1961.
- Klauer KC. Hierarchical multinomial processing tree models: A latent-class approach. *Psychometrika*. 2006; 71:7–31.
- Klauer KC. Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika*. 2010; 75:70–98.

- Liang F, Paulo R, Molina G, Clyde M, Berger J. Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association*. 2008; 103:410.
- Lunn D, Thomas A, Best N, Spiegelhalter D. WinBUGS – A Bayesian modeling framework: Concepts, structure, and extensibility. *Statistics and Computing*. 2000; 10:325–337.
- Matzke D, Dolan CV, Batchelder WH, Wagenmakers EJ. Bayesian estimation of multinomial processing tree models with heterogeneity in participants and items. *Psychometrika*. 2013;100:7/S11336-013-9374-9
- McDaniel, MA.; Einstein, GO. *Prospective memory: An overview and synthesis of an emerging field*. Thousand Oaks, CA: Sage; 2007.
- Moshagen M. multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*. 2010; 42:42–54. [PubMed: 20160285]
- Pavawalla SP, Schmitter-Edgecombe M, Smith RE. Prospective memory following moderate-to-severe traumatic brain injury: A multinomial modeling approach. *Neuropsychology*. 2012; 26:91–101. [PubMed: 21988127]
- Reese CM, Cherry KE. The effects of age, ability, and memory monitoring on prospective memory task performance. *Aging, Neuropsychology, and Cognition*. 2002; 9:98–113.
- Riefer DM, Batchelder WH. Multinomial modeling and the measurement of cognitive processes. *Psychological Review*. 1988; 95:318–339.
- Rouder JN, Lu J. An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*. 2005; 12:573–604. [PubMed: 16447374]
- Rouder J, Lu J, Sun D, Speckman P, Morey R, Naveh-Benjamin M. Signal detection models with random participant and item effects. *Psychometrika*. 2007; 72:621–642.
- Schnitzspahn KM, Horn S, Bayen UJ, Kliegel M. Age effects in emotional prospective memory: Cue valence differentially affects the prospective and retrospective component. *Psychology and Aging*. 2012; 27:498–509. [PubMed: 21843006]
- Smith JB, Batchelder WH. Assessing individual differences in categorical data. *Psychonomic Bulletin & Review*. 2008; 15:713–731. [PubMed: 18792498]
- Smith JB, Batchelder WH. Beta-MPT: Multinomial processing tree models for addressing individual differences. *Journal of Mathematical Psychology*. 2010; 54:167–183.
- Smith RE. The cost of remembering to remember in event-based prospective memory: Investigating the capacity demands of delayed intention performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2003; 29:347–361.
- Smith, RE. Prospective memory: A framework for research on metaintentions. In: Dunlosky, J.; Tauber, S., editors. *The Oxford Handbook of Metamemory*. New York, NY: Oxford University Press; in press
- Smith RE, Bayen UJ, Martin C. The cognitive processes underlying event-based prospective memory in school age children and young adults: A formal model-based study. *Developmental Psychology*. 2010; 46:230–244. [PubMed: 20053020]
- Smith RE, Bayen UJ. A multinomial model of event-based prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2004; 30:756–777.
- Smith RE, Bayen UJ. The effects of working memory resource availability on prospective memory: A formal modeling approach. *Experimental Psychology*. 2005; 52:243–256. [PubMed: 16304724]
- Smith RE, Bayen UJ. The source of adult age differences in prospective memory: A multinomial modeling approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2006; 32:623–635.
- Smith RE, Horn SS, Bayen UJ. Prospective memory in young and older adults: The effects of ongoing task load. *Aging, Neuropsychology, and Cognition*. 2012; 19:495–514.
- Smith RE, Hunt RR. Prospective memory in young and older adults: The effects of task importance and ongoing task load. *Aging, Neuropsychology, and Cognition*. 2014; 21:411–431.
- Smith RE, Persyn D, Butler P. Prospective memory, personality, and working memory: A formal modeling approach. *Zeitschrift für Psychologie / Journal of Psychology*. 2011; 219:108–116. [PubMed: 21822501]

- Smith RE, McConnell Rogers MD, McVay JC, Lopez JA, Loft S. Investigating how implementation intentions improve non-focal prospective memory tasks. *Consciousness and Cognition*. 2014; 27:213–230. [PubMed: 24929276]
- Unsworth N, Redick TS, Heitz RP, Broadway JM, Engle RW. Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory*. 2009; 17:635–654. [PubMed: 19536691]
- Wagenmakers EJ. A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*. 2007; 14:779–804. [PubMed: 18087943]
- Wetzels R, Wagenmakers EJ. A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review*. 2012; 19:1057–1064. [PubMed: 22798023]
- Wixted J. Dual process theory and signal detection theory of recognition memory. *Psychological Review*. 2007; 114:152–176. [PubMed: 17227185]
- Yonelinas AP, Regehr G, Jacoby LL. Incorporating response bias in a dual-process theory of memory. *Journal of Memory and Language*. 1995; 34:821–835.

Figure 1.

Multinomial model of event-based prospective memory. PM = prospective memory; C_1 = probability of detecting that a letter string is a word in lexical decision (or that a sentence is true in sentence verification); C_2 = probability of detecting that a letter string is a non-word in lexical decision (or that a sentence is false in sentence verification); P = prospective component; M = probability of distinguishing PM targets and non-targets (retrospective component); g = probability of guessing that a word is a target; c = probability of guessing that a letter string is a word in lexical decision (or that a sentence is true in sentence verification). Adapted from "A Multinomial Model of Event-Based Prospective Memory" by R. E. Smith and U. J. Bayen, 2004, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(4), pp. 756-777. Copyright 2004 by the American Psychological Association.

Table 1
Means and Standard Deviations (SD) for WM and PM Measures, and correlations of PM with WM measures

		Smith & Bayen (2005)	Smith, Persyn, & Butler (2011)
		Exp 1	
Working Memory	<i>Mean</i>	22.15	28.67
Span	<i>SD</i>	10.82	7.39
PM Hits	<i>Mean</i>	.78	.55
	<i>SD</i>	.21	.33
	<i>r</i>	.40	.11
	<i>p</i>	.04	.02
	<i>BF</i>	1.50	1.12
Beta-MPT Prospective	<i>Mean</i>	.81	.64
Component <i>P</i>	<i>SD</i>	.15	.01
	<i>r</i>	.39	.15
	<i>p</i>	.04	< .01
	<i>BF</i>	1.42	7.33
	<i>posterior</i>	.35 [-.13 - .52]	.14 [-.10 - .17]
Beta-MPT Retrospective	<i>Mean</i>	.96	.75
Component <i>M</i>	<i>SD</i>	.02	.01
	<i>r</i>	.19	.01
	<i>p</i>	.21	.92
	<i>BF</i>	0.38	0.01
	<i>posterior</i>	.12 [-.29 - .42]	.01 [-.04 - .06]
Latent-Trait Prospective	<i>Mean</i>	.82	.72
Component <i>P</i>	<i>SD</i>	.17	.02
	<i>r</i>	.41	.143
	<i>p</i>	.04	< .01
	<i>BF</i>	1.57	5.41
	<i>posterior</i>	.40 [-.15 - .82]	.16 [-.06 - .27]
Latent-Trait Retrospective	<i>Mean</i>	.97	.82
Component <i>M</i>	<i>SD</i>	.05	.02
	<i>r</i>	.17	.041
	<i>p</i>	.24	.40
	<i>BF</i>	0.33	0.07
	<i>posterior</i>	.12 [-.50 - .74]	.01 [-.10 - .12]

Note.

PM Hits = rate of PM targets correctly responded to; *P* = prospective component of PM; *M* = retrospective component of PM. Parameter estimates *P* and *M* represent the mean of the medians of participant-level posterior distributions. WM = working memory span; *r* = Pearson correlation coefficient; *p* = *p* value for the test of the correlation (one-tailed); *p* values smaller than .05 are significant. *BF* = Bayes factor; *BF* smaller than 0.33 or greater than 3.00 are relevant. *posterior* = posterior distribution of the correlation with mean and 95% BCI.

Table 2
Posterior Distributions of the Population Level Parameters of the Hierarchical Distributions and Parameters Estimated from Aggregated Data via MultiTree

	Beta-MPT		Latent-Trait		MultiTree		
	Mean	[95 % <i>BCI</i>]	Mean	[95 % <i>BCI</i>]	Mean	[95 % <i>CI</i>]	
2005	<i>P</i>	.79	[.70 - .86]	.84	[.73 - .94]	.81	[.76 - .85]
	<i>M</i>	.96	[.92 - .98]	.98	[.94 - 1]	.97	[.94 - .99]
	<i>C</i> ₁	.90	[.85 - .93]	.92	[.87 - .95]	.90	[.88 - .93]
	<i>C</i> ₂	.77	[.71 - .82]	.78	[.72 - .84]	.78	[.74 - .81]
2011	<i>P</i>	.64	[.61 - .66]	.72	[.68 - .76]	.67	[.66 - .69]
	<i>M</i>	.75	[.72 - .77]	.82	[.78 - .86]	.80	[.79 - .82]
	<i>C</i> ₁	.94	[.93 - .94]	.95	[.95 - .96]	.94	[.94 - .94]
	<i>C</i> ₂	.95	[.94 - .95]	.96	[.96 - .97]	.95	[.94 - .95]

Note.

2005 2005 = R. E. Smith and Bayen (2005), Experiment 1; 2011 = R. E. Smith, Persyn, and Butler (2011); *P* = prospective component of PM; *M* = retrospective memory component of PM; *C*₁ = probability to detect that a letter string is a word in lexical decision, or that a sentence is true in sentence verification; *C*₂ = probability to detect that a letter string is a non-word in lexical decision, or that a sentence is false in sentence verification; *BCI* = Bayesian confidence interval; *CI* = (traditional) confidence interval.