Hierarchical Multinomial Modeling of Individual Differences in Memory

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Nina Rebecca Arnold
aus Lahnstein

Düsseldorf, März 2014

aus dem Institut für Experimentelle Psychologie
der Heinrich-Heine Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Referentin: Prof. Ute J. Bayen, Ph.D.

Korreferent: Prof. Dr. Jochen Musch

Tag der mündlichen Prüfung: 17.4.2014

Contents

Zusammenfassung

Multinomiale Verarbeitungsbaummodelle sind statistisch motivierte Modelle, die die Trennung zugrundeliegender kognitiver Prozesse bei kategorialen Daten erlauben. Traditionell werden hierbei die Daten über Probanden und Items hinweg aggregiert und Heterogenität wird ignoriert. Diese Arbeit stellt drei Studien vor, die zwei hierarchische Erweiterungen anwenden, die Heterogenität von Probanden beinhalten: den beta-MPT-Ansatz (J. B. Smith & Batchelder, 2010) und den Latent-Trait-Ansatz (Klauer, 2010; Matzke, Dolan, Batchelder, & Wagenmakers, 2013). In der ersten Studie wurde der Beta-MPT-Ansatz auf das Zwei-Hochschwellenmodell des Quellengedächtnisses (Bayen, Murnane & Erdfelder, 1996) angewendet. Es zeigte sich ein signifikanter Zusammenhang zwischen der wahrgenommenen Quelle-Item-Kontingenz und dem Rateverhalten. Wenn die Quellenschemata erst vor dem Test aktiviert wurden, fand sich außerdem ein Zusammenhang zwischen der Abweichung der wahrgenommenen Quelle-Item-Kontingenz zur wahren Kontingenz und der Güte des Quellengedächtnisses. Dies unterstützt den Probability-Matching-Ansatz (Spaniol & Bayen, 2002). Dieser besagt, dass Leute ihr Rateverhalten an die wahrgenommene Quelle-Item-Kontingenz anpassen. In der zweiten Studie wurde der Beta-MPT-Ansatz auf das multinomiale Modell für ereignisbasiertes prospektives Gedächtnis (PG; R. E. Smith & Bayen, 2004) angewendet. PG beinhaltet die Erinnerung daran, in der Zukunft eine Aufgabe zu erledigen. Die prospektive Komponente – sich erinnern, dass man etwas tun muss – und die retrospektive Komponente – sich daran erinnern, wann man etwas tun muss – des PG wurde mit Angst- und Depressions-Fragebogenwerten korreliert. Nur Zustandsangst war negativ mit der prospektiven Komponente korreliert. Die dritte Studie verglich den Beta-MPT und den Latent-Trait-Ansatz durch Reanalyse der Daten dreier Experimente (Experiment 1 und 2 von R. E. Smith & Bayen, 2005; R. E. Smith, Persyn, & Butler, 2011), die den Zusammenhang zwischen PG und Arbeitsgedächtnis untersuchten.

Übereinstimmend zeigte sich eine positive Korrelation zwischen der prospektiven Komponente und der Arbeitsgedächtnisspanne. Die Parameterschätzungen beider Ansätze korrelierten hoch miteinander.

Abstract

Multinomial processing tree (MPT) models are a class of statistical models that can be applied to categorical data to separate underlying cognitive processes. Traditionally, MPT models use data that are aggregated over participants and items – thereby ignoring heterogeneity. This thesis presents three studies that applied two hierarchical extensions of MPT models that incorporate heterogeneity: The beta-MPT approach (J. B. Smith & Batchelder, 2010) and the latent-trait approach (Klauer, 2010, Matzke, Dolan, Batchelder, & Wagenmakers, 2013). In the first study, the beta-MPT version of the two-high-threshold model of source-monitoring (Bayen, Murnane, & Erdfelder, 1996) was applied to find significant correlations between the perceived source-item contingency and source guessing. When the source schema was not activated until retrieval, the correlation between the deviation of the guessing bias from the true contingency and the accuracy of source memory was also significant. This supports the probability-matching account (Spaniol & Bayen, 2002) which states that people adjust their source-guessing to the perceived source-item contingency. In the second study, the beta-MPT was applied to the multinomial model of event-based prospective memory (PM; R. E. Smith & Bayen, 2004). PM requires remembering to perform a task in the future. The prospective component – remembering that you have to do something – and the retrospective component – remembering when to do something – were correlated with questionnaire estimates for depression and anxiety. Only state anxiety was negatively correlated with the prospective component. The third study was a comparison of the beta-MPT and the latent-trait approach by a reanalysis of three experiments (R. E. Smith & Bayen, 2005, Experiments 1 and 2; R. E. Smith, Persyn, & Butler, 2011) that investigated the relationship between PM and working memory. The prospective component of PM was positively correlated with working-memory span for both hierarchical approaches. Parameter estimates of both approaches were highly correlated.

1 Introduction

Many, if not most cognitive processes are unconscious and not accessible by introspection. Additionally, psychological data result from multiple interacting processes (Batchelder & Riefer, 1999). Thus, they are also not assessable with simple behavioral measurements. However, most theories focus on these underlying processes to explain psychological phenomena (Riefer & Batchelder, 1988). Therefore, it is necessary to operationalize the theories and create auxiliary hypotheses to separate these processes.

Erdfelder (2000) stated that there are three main problems for empirical operationalism: (1) The problem of meaningfulness questions the validity of the measurements. The choice of the dependent variable can influence the results. If a certain construct is operationalized in two ways, results obtained by both operationalizations can either converge or diverge. If the results converge, the theory is said to be either approved (if the result is positive) or disapproved (if the result is negative). If the results diverge, the operationalizations apparently measure different constructs. This resembles construct validity of measurement models in the sense of Cronbach and Meehl (1955): Convergent validity means that measurements covariate with related constructs. Discriminant validity describes the lack of covariation with unrelated constructs. Measurements are said to be valid if they show both types of validity. (2) The second problem – error problem – according to Erdfelder (2000) is closely related to reliability. Usually, empirical data are error-prone due to variability in participants and items, confounding variables, and random variation. One way to deal with this is to enhance the precision of the empirical observation. Another solution is to explicitly model errors because even in a controlled setting there will be uncontrolled errors in human behavior. (3) The third problem is the decomposition problem. It is closely related to the second problem. As stated above, empirical measurements are never process-pure or purely represent a theoretical construct. Normally, they involve other psychological processes that are not related to the construct. If an empirical measurement varies between different conditions, this could be due to other (confounding) variables but not due to the theoretical construct of interest. Design-based approaches try to eliminate confounding variables. Stochastic model-based approaches, in contrast, incorporate confounding variables into the model to separate them from the processes of interest. This is their biggest advantage in comparison to the design-based approach.

Memory, like most cognitive abilities, is not described by a single process or system (e.g., Baddeley, Eysenck, & Andersen, 2009). For example, the number of hits in a recognition experiment does not only depend on the abilities to discriminate between presented and non-presented items but also on response biases (e.g., Snodgrass & Corwin, 1988). The comparison of different conditions in an experiment is likewise problematic because different measures of memory vary in sensitivity. Therefore, dissociation (i.e., an effect on memory measure A but not on memory measure B) can occur due to different reliabilities of those measures (Erdfelder, 2000).

Most observations in memory paradigms are categorical (e.g., hits and false alarms). This makes them easily applicable to multinomial processing tree (MPT) models. MPT models (or multinomial models) are statistically motivated models for categorical data that separate the underlying cognitive processes. MPT models solve the problems mentioned before because they force researchers to make their underlying assumptions explicit (thereby addressing the meaningfulness problem) and allow to model other processes as well (thereby addressing the error problem and the decomposition problem). Traditionally, they have been applied to aggregated data. However, there has been rising awareness for the usefulness of individual MPT parameter estimates and individual MPT models. The present thesis addresses the recently developed hierarchical MPT models that are based on Markov Chain Monte Carlo (MCMC) sampling. Hierarchical MPT models can deal with heterogeneity in participants and/or items and estimate individual model parameters. This provides the means to calculate correlations between the individual parameter estimates and other cognitive abilities. Usually, hierarchical MPT models are described for the case of participant heterogeneity. These models still assume that items are homogenous. However, the approaches can be modified to the case where items are assumed to be heterogeneous and participants are assumed to be homogeneous.

The present thesis applies the novel technique of hierarchical MPT modeling to different MPT models within the memory domain. After describing traditional MPT models, hierarchical MPT models including discrete and continuous hierarchical models as well as the associated methodological background are presented. Finally, three studies applying the beta-MPT approach and the latent-trait approach to selected MPT models are described and discussed.

2 Multinomial Processing Tree Modeling

MPT modeling is a nonlinear statistical method for analyzing observable events that can be described by categorical frequency data (e.g., Batchelder & Riefer, 1999). MPT models are very useful to draw conclusions about latent, non-observable variables and to estimate the likelihood of different underlying causes for an observable event. These underlying causes are represented by the parameters of an MPT model. The model parameters lie in the interval (0, 1) (e.g., Batchelder & Riefer, 1999; Hu & Batchelder, 1994) and can be interpreted as transition probabilities between the cognitive states.

MPT models assume a multinomial distribution on a finite set of mutually exclusive categories (e.g., Klauer, 2006; Matzke, Dolan, Batchelder, & Wagenmakers, 2013; Purdy & Batchelder, 2009). They make the underlying assumptions of the theory (or the application they are designed for) explicit by displaying the hypothesized sequences of cognitive events (Purdy & Batchelder, 2009). These sequences can be represented as a tree. A joint MPT model has several independent item classes and consists of different subtrees for each item class (Riefer & Batchelder, 1988).

MPT models are tailored to a specific research paradigm which makes them very flexible tools. Erdfelder et al. (2009) identified 70 MPT models in more than 20 research areas. Most MPT models deal with memory paradigms: for example, hindsight bias (Erdfelder & Buchner, 1998), prospective memory (R. E. Smith & Bayen, 2004), recognition memory (Batchelder & Riefer, 1990), source memory (Bayen, Murnane, & Erdfelder, 1996), and pair-clustering (Batchelder & Riefer, 1986). This is not surprising since the problem of decomposition is very prominent in memory research. Conclusions based on MPT modeling often do not concur with design-based conclusions. However, corresponding or diverging results of both methods do neither validate nor contradict the MPT model. The goodness of a MPT model depends only on its validity which is tested via systematically manipulating its parameters and check for construct validity (Batchelder & Riefer, 1999).

A special case of an MPT model is a binary MPT (BMPT) model. A BMPT model is a model that has only two choice alternatives at every node of the tree. Most MPT models can be transferred to a statistical equivalent BMPT model (Hu & Batchelder, 1994, Purdy & Batchelder, 2009). Only a few models like MPT models with order constraints in the parameters cannot be transferred into BMPTs (Knapp & Batchelder, 2004). The advantage of the notation as BMPT model is that it allows presentations that are computationally more efficient (Purdy & Batchelder, 2009). Formal properties of MPT models are described in Hu & Batchelder (1994), Purdy and Batchelder (2009), as well as Riefer and Batchelder (1988). I use the BMPT notation throughout this thesis and mostly follow the notation by J.B. Smith and Batchelder (2010). Table 1 provides summary of the meaning of the parameters.

An MPT model is described by a set of $K \geq 2$ categories $\boldsymbol{C} = \{C_1, \ldots, C_K\}$, and a vector of $S \geq 1$ parameters $\boldsymbol{\theta} = \langle \theta_s \rangle_{s=1}^{S}$ where each component has the parameter space (0, 1). The parameters are functionally independent, so the resulting parameter space is the cross-product of all parameter spaces, namely $\Lambda_{\boldsymbol{\theta}} = (0,1)^S$. A BMPT model consists of a finite set of branches. Each branch is a path from the root to one of the assigned categories. The probability of the ith branch that terminates in category C_k is called B_{ik}, where $i = 1,\ldots,I_k$, $k = 1,\ldots,K$. The probability of the ith branch is a product of the parameters θ_s and their complements $1 - \theta_s$, respectively, given by:

$$P(B_{ik}|\boldsymbol{\theta}) = \prod_{s=1}^{S} \theta_s^{u_{ik,s}} (1-\theta_s)^{v_{ik,s}}, \tag{1}$$

where $u_{ik,s} \geq 0$ and $v_{ik,s} \geq 0$ are the number of arrows on branch B_{ik} that are associated with θ_s and $(1 - \theta_s)$, respectively. The probability of a category C_k is calculated by adding all I_k branches that terminate in that category

$$P(C_k|\boldsymbol{\theta}) = \sum_{i=1}^{I_k} \prod_{s=1}^{S} \theta_s^{u_{ik,s}} (1-\theta_s)^{v_{ik,s}}. \tag{2}$$

It follows from these rules that

$$\forall\, \widetilde{\boldsymbol{\theta}} \in \Lambda_{\boldsymbol{\theta}}, \qquad \sum_{k=1}^{K} P\big(C_k \big| \widetilde{\boldsymbol{\theta}}\big) = 1. \tag{3}$$

Categorical data consist of observations from N participants responding to M items. Each of the observations falls into one of the C_k categories. The data of N participants responding to M items can be described by the matrix $\boldsymbol{D} = (x_{nm,k})_{NxMxK,}$ where $x_{nm,k} = 1$ if participant n responds to item m in category C_k and $x_{nm,k} = 0$ otherwise. Overall, this results in $N * M$ total observations.

Traditionally, parameter estimation is done by the expectation-maximization (EM) algorithm (Hu & Batchelder, 1994). It decreases the distances between observed and expected frequencies and chooses the parameter values that belong to that expected frequencies. The distance is usually measured by a member of the power divergence family (PD^{λ}, Read & Cressie, 1988), for example the χ^2 or the G^2-statistic (for a complete description see Moshagen, 2010). However, other members of PD^{λ} are also possible. All members of PD^{λ} are asymptotically χ^2-distributed. Depending on the number of λ, PD^{λ} may be undefined if empty cells occur. However, it is possible to add small values to all cells to avoid this. The PD^{λ} statistic is also used to assess the goodness-of-fit. Alternatively information criteria like the Akaike information criterion (AIC) or the Bayesian information criterion (BIC) can be used to assess model-fit. There are several computer programs and implementations that can be used for MPT modeling, for example, AppleTree (Rothkegel, 1999), GPT (Hu & Phillips, 1999) and HMMTree (Stahl & Klauer, 2007), MultiTree (Moshagen, 2010), and MPTinR (Singmann & Kellen, 2013).

Table 1

Notation

Symbol/Notation	Explanation
K	Number of categories
$\boldsymbol{C} = (C_1, \ldots, C_k)$	Categories
S	Number of parameters
N	Number of participants
M	Number of items
$\boldsymbol{\theta} = \langle \boldsymbol{\theta_s} \rangle_{s=1}^{S}$	Parameters (group-level)
$\boldsymbol{\theta_n} = \boldsymbol{\theta_{s,n}}$	Parameters (person-level)
I_k	Number of branches that terminate in category C_k
B_{ik}	ith branch that terminates in category C_k
$u_{ik,s}$	Number of branches that are associated with $\boldsymbol{\theta_s}$
$v_{ik,s}$	Number of branches that are associated with $(\boldsymbol{1} - \boldsymbol{\theta_s})$
$x_{nm,k}$	Response of participant n to item m in category C_k
$\boldsymbol{D}$	Data; matrix that entails responses of all participants to all items
$\boldsymbol{F_{n,k}}$	Number of items that fall in category C_k for participant n
L	Number of latent classes
λ_l	Size of latent class l
τ	Hyperparameter, defines hyperdistribution
α, β	Parameters that define the beta distribution
$\boldsymbol{\mu_s}$	probit transformed mean of parameter $\boldsymbol{\theta_s}$
S_{part}	Variance-covariance- matrix of the latent-trait approach
$\boldsymbol{\theta_{ns}^{prt}} = \boldsymbol{\Phi^{-1}}(\boldsymbol{\theta_{n,s}})$	Probit transformed parameters
$\boldsymbol{\rho_{ss'}}$	Correlation parameters
$\boldsymbol{\xi_{part}} = [\boldsymbol{\xi_{part_1}}, \ldots, \boldsymbol{\xi_{part_s}}]$	Scale parameters of scaled Inverse-Wishart

2.1 The Two-High-Threshold Model of Source-Monitoring

The two-high-threshold model of source monitoring (2HTSM; Bayen, et al., 1996) separates memory from guessing in source-monitoring; that is, judging the origin of information (e.g., Johnson, Hashtroudi, & Lindsay, 1993). In a typical source-monitoring task, participants are first presented with items from at least two different sources. In the following test they have to judge whether an item was presented by one of the sources or not (i.e., the item is new). If they indicate that the item was presented by one of the sources they have to judge by which one. The 2HTSM is presented in Figure 1.

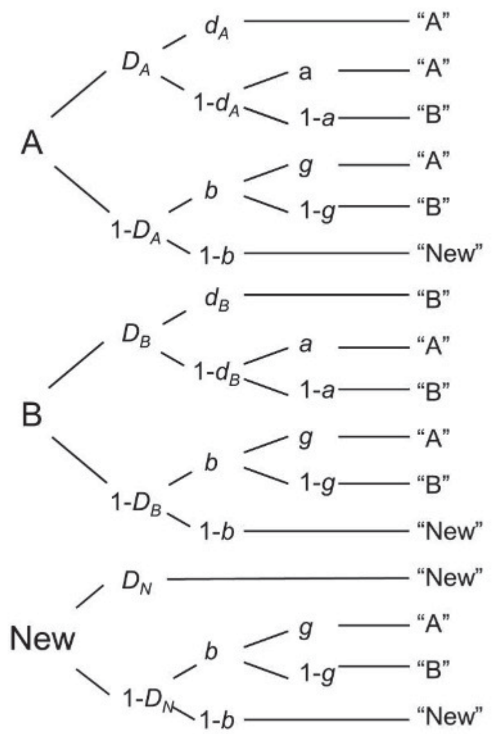


Figure 1. The two-high-threshold model of source monitoring. D_A = probability of detecting that an item presented by Source A is old; D_B = probability of detecting that an item presented by Source B is old; D_N = probability of detecting that an item is new; d_A = probability of correctly remembering that an item was presented by Source A; d_B = probability of correctly remembering that an item was presented by Source B; a = probability of guessing that an item that has been recognized as old is from Source A; g = probability of guessing that an item is from Source A if it was not recognizes as old; b = probability of guessing that an item is old. Adapted from "Source discrimination, item detection, and multinomial models of source monitoring," by U. J. Bayen, K. Murnane, and E. Erdfelder, 1996, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* p. 202. Copyright 1996 by the American Psychological Association.

Most often, the 2HTSM is applied to experiments with two sources but there are also versions for three or more sources (Keefe, Arnold, Bayen, McEvoy, & Wilson, 2002; Riefer, Hu, and Batchelder, 1994). In the two-source version, items are either presented by Source A or by Source B. Sources can be defined for example by location, mode of presentation, or different people "presenting" statements (Johnson et al., 1993). In Figure 1, the first tree represents items that have been presented by Source A. With probability D_A, participants correctly recognize that an item is old. With probability d_A, they also remember that the item was presented by Source A. With probability 1-d_A they cannot remember the source and have to guess. With probability a, they guess that the item is from Source A. With probability 1-a, they guess that the item is from Source B. With probability 1-D_A, participants cannot remember that the item is old. They guess with probability b that the item is old or, with probability 1-b, that it is new. It is assumed that they cannot remember the source of an item when they cannot recognize it as old and have to guess. With probability g, they guess that the item was presented by Source A; with probability 1-g, they guess that it was presented by Source B. The second tree represents items that have been presented by Source B. It is very similar to the first tree. Here, D_B defines the probability that participants correctly recognize that an item is old, and d_B describes the probability that they also remember that the item was presented by Source B. The third tree represents processes for new items that were not presented. D_N describes the probability that participants correctly notice that an item is new. Of course it is not possible to recognize the source of an item because it has not been presented before. However, if participants do not notice that the item is new with probability 1-D_N, they have to guess. If they guess with probability b that the item is old, they either guess with probability g that the item was presented by Source A or with probability 1-g that it was presented by Source B.

The model, as presented above, is not identifiable. Bayen et al. (1996) presented a nested hierarchy of all identifiable submodels. Submodel 4 is often used for the standard source-monitoring paradigm (e.g., Bayen & Kuhlmann, 2011; Bayen, Nakamura, Dupuis, & Yang, 2000; Kuhlmann, Vaterrodt, & Bayen, 2012). It assumes that item memory ($D_A = D_B$) as well as source memory ($d_A = d_B$) are equal for both sources and that the probability of noticing that an item is new is equal to the probability of recognizing an item as old ($D_A = D_B = D_N$).

2.2 The Multinomial Model of Event-Based Prospective Memory

The multinomial model of event-based prospective memory (PM) was designed by R. E. Smith and Bayen (2004) to analyze data from standard laboratory event-based PM tasks. PM describes the task of remembering to perform an action in the future. Event-based PM means that participants have to perform the action when a certain event occurs. In contrast, time-based PM requires performing the action at a particular point of time (e.g., McDaniel & Einstein, 2007). However, there is no MPT model for time-based PM. PM involves a prospective component, remembering *that* you have to do something, and a retrospective component that involves remembering *what* action to perform and *when* to perform it (Einstein & McDaniel, 1990). The model was designed to separate these two components. In a laboratory PM task, participants are engaged in an ongoing activity, for example a lexical-decision task (e.g., R. E. Smith, Persyn, & Butler, 2011), a sentence verification task (e.g., R. E. Smith & Bayen, 2005), or a color-matching task (e.g., R. E. Smith & Bayen, 2004). The PM task is to press a special key when a rare PM target occurs, for example a special word in a color-matching or sentence-verification task or specific letters or syllables in a lexical-decision task.

The MPT model (see Figure 2) can only be applied for binary ongoing task, such as the ones described above, and PM targets occurring on both options of the ongoing task. This results in four possible events that are represented by separate subtrees: (1) a PM target occurs on Option

1 of the ongoing task, (2) a PM target occurs on Option 2 of the ongoing task, (3) an Option 1 event occurs without a PM target, and (4) an Option 2 event occurs without a PM target.

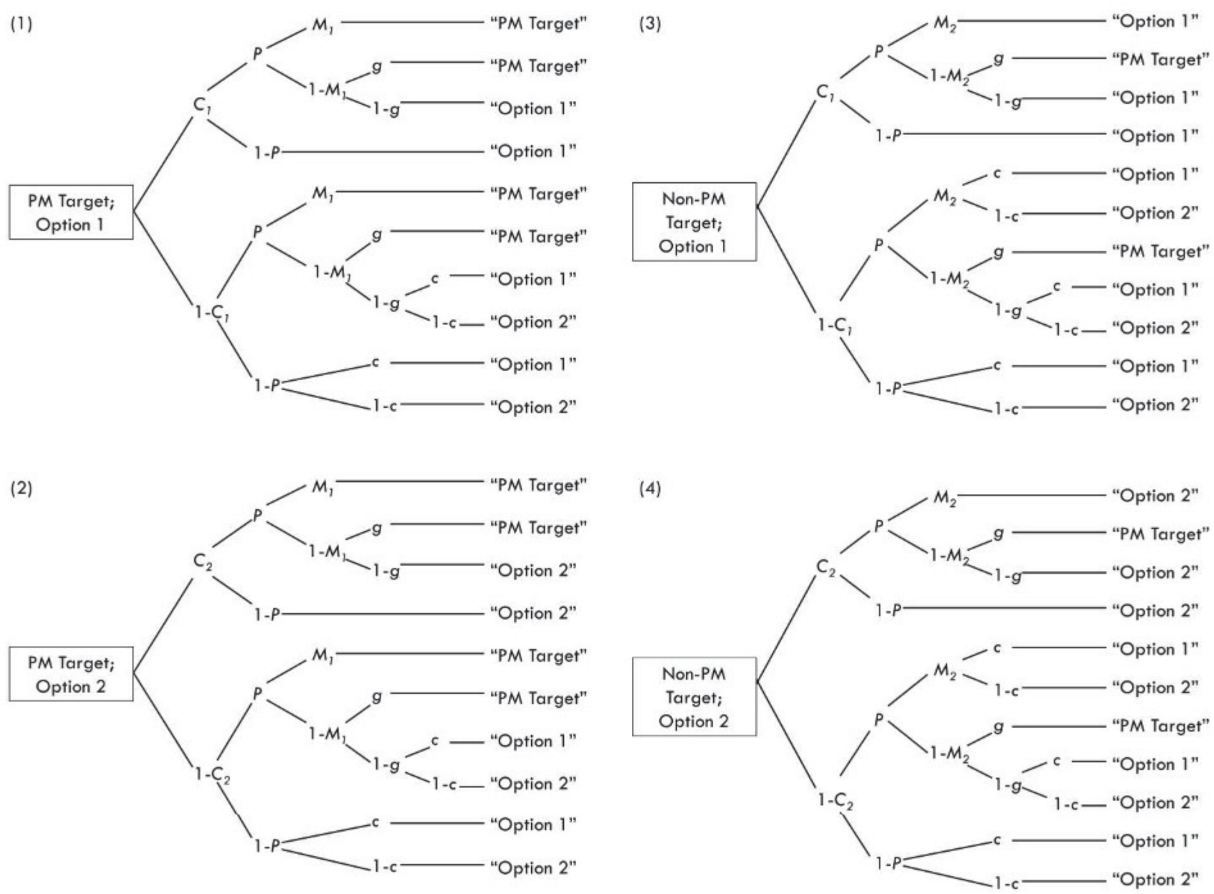


Figure 2. The Multinomial Model of Event-Based PM. PM = prospective memory; C_1 = probability of detecting Option 1 in the ongoing task; C_2 = probability of detecting Option 2 in the ongoing task; P = prospective component; M_1 = probability of recognizing PM targets; M_2 = probability of noticing that an event is not a PM-target; g = probability of guessing that the event is a target; c = probability of guessing that the answer to the ongoing task is Option 1. 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*, p. 758.

The ability to solve the ongoing task is captured by Parameters C_1 and C_2. On Option 1 trials (i.e., first and third tree in Figure 2), C_1 is the probability that the participant recognizes Option 1. The second and fourth tree represent Option 2 trials. Here, C_2 is the probability that the participant recognizes Option 2. P is the probability that the participant remembers that there is an additional task (i.e., the prospective component). On target trials (first and second tree) M_1 is the probability that a participant successfully recognizes a PM target. This results in a PM response. On non-target trials (third and fourth tree), M_2 is the probability that a participant successfully recognizes that an event is not a PM target. This results in an ongoing-task response. Note that although M_1 and M_2 represent the retrospective component in the model, they only capture 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).

If participants do not recognize whether an item is a PM target ($1-M_1$) or not ($1-M_2$), they have to guess. Parameter g describes the probability of guessing that the trial includes a PM target, whereas $1–g$ denotes the probability of guessing that the trial does not include a PM target. In the latter case, participants respond to the ongoing task. If the participants do not remember that there is an additional PM task ($1 - P$), they simply respond to the ongoing task. If the participants cannot solve the ongoing task ($1-C_1$ respectively $1-C_2$), they guess with probability c for Option 1 and with probability $1- c$ for Option 2.

The model as described above is not identifiable. Therefore, the parameters are restricted based on theoretical assumptions (R. E. Smith & Bayen, 2004): Parameters M_1 and M_2 are set equal resulting in Parameter M that reflects the process of discriminating between PM targets and non-targets. The guessing parameter c and g are set according to the ratio of Option 1 and Option 2 items and the ratio of PM targets and non-PM-targets, respectively. This reflects the assumption that participants use probability-matching. The resulting model has four parameters: P, M, C_1, and C_2, and has been validated (Horn, Bayen, R. E. Smith, & Boywitt, 2011; R. E. Smith & Bayen, 2004).

3 Hierarchical MPT Modeling

In recent years, there has been rising awareness that the traditional method to evaluate MPT models is limited (e.g., Erfelder, 2000; Klauer, 2006, 2010; Matzke, et al., 2013; J. B. Smith & Batchelder, 2008, 2010). MPT models are usually applied to aggregated data assuming that all participants and items are independent and identically distributed (i.i.d.) over the categories which means that all participants and all items are exchangeable and follow the same distribution (i.e., they are homogeneous). It is assumed that neither participants nor items differ in terms of parameter estimates and category counts are a sample from the model (Matzke et al., 2013).

J. B. Smith and Batchelder (2008) showed that this assumption rarely holds – even for a relatively homogenous pool of participants like first year-psychology students and a carefully constructed item pool. If the assumption is violated, parameter estimates can be biased (Erdfelder, 2000; Klauer, 2006, 2010; J. B. Smith & Batchelder, 2008, 2010). Estimates based on aggregated data underestimate then the variance in data and lead to confidence intervals that are too narrow. Thus, goodness-of-fit tests become significant too often and models are falsely rejected (Klauer, 2006). Furthermore, overdispersion increases when heterogeneity and data points per participant increase (Klauer, 2006). Additionally, parameters may often be correlated and the pattern of the correlation can bias the parameter estimates (Matzke et al., 2013). Parameter correlations lead to an overestimation of variability and, therefore, to confidence interval that are too wide (Klauer, 2010). Hence, all kinds of biases are possible for aggregated data.

Individual differences are especially likely for special groups like children or older adults. Even if the group seems relatively homogenous, the research question may require individual differences measures. Cognitive psychometrics (e.g., Batchelder, 1998, Batchelder & Riefer, 1999) use well-validated MPT models as measurement tools for latent cognitive processing abilities in individuals. This is especially important in clinical assessments and differential diagnosis. Parameter estimates based on aggregated data cannot be used here.

The obvious solution would be to calculate separate models for each participant. Often categories are only sparsely filled and, therefore, cannot be evaluated with any degree of reliability. The usually used test statistics like G^2 are asymptotic tests that may produce biased results when the sample size is too small (Erdfelder et al., 2009). A rule of thumb is that there ought to be five observations or more per category (e.g., Hays, 1994). However, depending on the paradigm, some item classes rarely occur (e.g., PM targets in the multinomial model of event-based PM). In most cognitive experiments there are not enough observations to calculate reliable estimates and there may be even empty cells for some participants. Sometimes, a constant of one is added to all categories if there are zero counts in an analysis. However, this does still not meet the criteria of five or more counts per category and simply adding a constant does violate proportionality. Additionally, data obtained by small sample sizes are not reliable and have a large standard error of estimate (Erdfelder, 2000).

There are exact test (e.g., García-Pérez, 2000) and parametric bootstrap (Efron & Tibshirani, 1997, Moshagen, 2010) alternatives. MultiTree (Moshagen, 2010) offers a bootstrapping procedure. The advantage of bootstrapping is that is produces more reliable estimates of confidence intervals. Bootstrapping also allows for evaluating the exact distribution of the PD^{λ}-statistic if the assumptions are violated due to parameters at the boundaries (Moshagen, 2010) which also are often a result of individual parameter estimates.

Even if there are enough observations per participant to calculate separate MPT models there are advantages of incorporating heterogeneity into a common MPT model. A correctly specified hierarchical model provides more accurate estimates than separate parameter estimates because it corrects for outliers (e.g., Rouder & Lu, 2005). However, this effect diminishes as the amount of data per participant increases.

Hierarchical MPT models treat the parameters as random rather than fixed effects. The main difference to traditional MPT models is there is a core MPT model with potentially different parameter values for participants. Hierarchical MPT models specify a distribution of the individual parameters to model heterogeneity (Klauer, 2006). The MPT model is, hence, specified on two levels: On the base level (or group level), the group parameters are specified. This is comparable to the parameter estimates in traditional MPT models. The hierarchical level (or population level) is defined by hyperparameters and describes the variability in participants and/or items.

Different distributional forms for the population level, the so-called hyperdistribution, can be assumed, for example discrete distributions (latent-class approach), beta distributions (beta-MPT), or transformed normal distributions (latent-trait approach). Discrete hyperdistributions lead to finite-mixture models. If the hierarchical distribution is misspecified, the analysis of the data can be biased (J. B. Smith & Batchelder, 2008). However, in practice it has been shown that the group-level results can be interpreted even if the population-level distribution is misspecified (e.g., Agresti, Caffo, & Ohman-Strickland, 2004).

Methods for analyzing hierarchical models are computationally intensive because basic model parameters and hyperparameters have to be evaluated for each participant's data. Recently, there have been some developments that led to tractable methods. Programs like WinBUGS (Spiegelhalter, Thomas, Best, & Lunn, 2003) can estimate hierarchical models by using MCMC methods.

3.1 Tests of Parameter Homogeneity

Tests of parameter homogeneity assess whether the assumption of homogeneity is met and, thus, if parameter estimates on aggregated data can be used. If heterogeneity is present, it is inappropriate to use aggregated data since the results can be biased as pointed out above. However, models that incorporate heterogeneity are more complex and, therefore, capable of accounting for more data than models that do not include heterogeneity which could result in overfitting (J. B. Smith & Batchelder, 2008). Therefore, the decision whether to model heterogeneity is important. There are different methods to test for heterogeneity.

Klauer (2006) introduced the statistics S_1 and S_2 to test for variability among participants and parameter correlations. S_1 and S_2 can be computed from the output of traditional MPT analyses. Violations of parameter homogeneity should lead to overdispersion in the category counts and to inflated correlations between different category counts. S_1 and S_2 test whether variances and covariances are adequately described by the MPT under the assumption of homogeneity. Klauer (2006) showed that S_1 and S_2 detect heterogeneity almost with certainty.

J. B. Smith and Batchelder (2008) presented the model-free Monte Carlo permutation test for detecting heterogeneity in participants and/or items. Permutation tests condition on some aspect of the data and sample possible other aspects of the data to construct a reference distribution. The authors developed a test to check for item homogeneity when there is possible participant heterogeneity that can be easily adapted to test for participant heterogeneity. The method is based on the assumption that most statistical models result in overdispersion of some standard statistics of participants × items data array. They present program codes for R and MATLAB.

If the research question requires individual parameter estimates, another way to check for parameter heterogeneity is to check whether the credible intervals (i.e., Bayesian confidence intervals; BCIs) of the standard deviations for the parameters include zero (e.g., Matzke et al., 2013). If the BCIs do not include zero, parameter homogeneity (i.e., *SD* = 0) is very unlikely. Since BCIs unlike frequentist confidence intervals (CIs) stay within the parameter space, they will never be smaller than zero or larger than one.

3.2 Discrete Hierarchical MPT Modeling: The Latent-Class Approach

The latent-class approach (Klauer, 2006) is an extension of the traditional MPT approach that incorporates parameter heterogeneity to a certain extend but maintains the advantages of classical MPT modeling techniques. The main idea is that participants fall into a finite number of L mutually exclusive latent classes with size λ_l. Within a class, all participants are assumed to have the same parameter values θ_{sl}. Parameters can vary and even correlate across classes. Thus, there are L fixed parameter vectors θ_l with $l = 1,\ldots, L$.

To estimate the parameters of Klauer's (2006) approach, the EM-algorithm by Hu and Batchelder (1994) can be adapted for the maximum likelihood estimation of latent-class MPTs. Participants are assigned with a posterior probability of class membership. It is possible to calculate individual parameter estimates by computing the weighted means of the class parameters. However, these parameters are still assumed to stem from a discrete hyperdistribution. Theoretically, any hyperdistribution can be approximated by a suitable number of latent classes but latent-class MPT models are usually limited to a small number of classes due to identifiability problems (Klauer, 2006; J.B. Smith & Batchelder, 2010). A latent-class model is identified if there are at least $2L-1$ observations per person n and tree j (Klauer, 2006).

The computer program HMMTree (Stahl & Klauer, 2007) provides a graphical user interface for calculating parameter estimates, CIs, and goodness-of-fit statistics for latent-class and traditional MPT models. Moreover, it allows for testing the assumption of parameter homogeneity using the S_1 and S_2 statistics. It also computes the Fisher Information matrix as well as AIC and BIC. HMMTree computes parameter estimates by using a combination of the EM algorithm as proposed by Klauer (2006) and the conjugate-gradient method (e.g., Shewchuk, 1994).

There may be situations where the assumption of homogenous subgroups holds - for example, if the heterogeneity is due to unknown discrete background variables like education (Klauer, 2006). However, in most cases, a continuous distribution is more plausible.

3.3 Continuous Hierarchical MPT Modeling

In continuous hierarchical MPT modeling, participant and/or item parameters are drawn i.i.d. from a multivariate hyperdistribution $h(\theta, \tau)$, where τ is the hyperparameter that defines the hyperdistribution. I present hierarchical MPT models for heterogeneous participants, each responding to the same set of homogeneous items. Within participants, observations are assumed to be i.i.d. (or in Bayesian terms exchangeable). Between participants, observations are independent but may not be identically distributed. The data structure consists of category counts for each participant and/or item, $\mathbf{H} = \langle \mathbf{F}_{\mathrm{n}} \rangle_{\mathrm{n}=1}^{\mathrm{N}}$, where $\langle \mathbf{F}_{\mathrm{n,k}} \rangle_{\mathrm{k}=1}^{\mathrm{K}}$ and $\mathrm{F}_{\mathrm{n,k}} = \sum_{\mathrm{m}=1}^{\mathrm{M}} \mathrm{X}_{\mathrm{nm,k}}$ denote the number of items that fall into category C_k for participant n. Thus, there exist n times as many individual data categories as aggregated data categories. It would be pointless to run sufficient participants to conduct classical inference (J. B. Smith & Batchelder, 2010).

As a solution, three approaches have been proposed: J. B. Smith and Batchelder (2010) proposed the beta-MPT model where parameters are assumed to follow independent marginal beta distributions. In contrast, the latent-trait approach (Klauer, 2010) assumes transformed multivariate normal distributions. The beta-MPT approach and the latent-trait approach can account for either variability in participants *or* items. The third approach, the crossed-random effects approach, proposed by Matzke, et al. (2013) is an extension of the latent-trait approach which covers participant and item variability simultaneously.

3.3.1 Techniques for Hierarchical Modeling

In most cases, hierarchical models are not solvable with classical maximum-likelihood estimation (MLE) techniques like the EM algorithm (Hu & Batchelder, 1994) but there is a Bayesian solution using MCMC methods. Model fitting and hypothesis testing in the Bayesian framework require different methods from traditional null-hypothesis significance testing (NHST) in the frequentist framework.

3.3.1.1 Bayesian Modeling

One way to illustrate Bayesian statistics is to think of it as measurement of the opinions of ideally consistent people (Edwards, Lindman, & Savage, 1963). Statistical inference is the modification of these opinions in the light of new evidence. Bayes' theorem is the rule how these updates should be made. Thus, Bayesian statistics is a set of rules and techniques for expressing and updating one's opinion. All uncertainties are seen as probabilities. Although initial opinions can vary extremely, the rules for updating the opinion are always the same.

In Bayesian modeling initial beliefs are represented by prior distributions. Based on the existing knowledge, these prior distributions can be rather vague or very concrete. A posterior distribution is calculated using Bayes' theorem:

$$P(\boldsymbol{\theta}|\boldsymbol{D}) = \frac{P(\boldsymbol{\theta})P(\boldsymbol{D}|\boldsymbol{\theta})}{P(\boldsymbol{D})} \tag{4}$$

The theorem states that the likelihood of the parameters $\boldsymbol{\theta}$ given the data $\boldsymbol{D}$, $P(\boldsymbol{\theta}|\boldsymbol{D})$ is proportional to the prior probability of the parameters $P(\boldsymbol{\theta})$ (the initial beliefs) times the probability of the data given the parameters $P(\boldsymbol{D}|\boldsymbol{\theta})$.

Bayes' theorem can be reformulated to apply to continuous data (Edwards et al., 1963; Rouder & Lu, 2005):

$$f(\boldsymbol{\theta}|\boldsymbol{D}) = \frac{f(\boldsymbol{\theta})f(\boldsymbol{D}|\boldsymbol{\theta})}{f(\boldsymbol{D})} \tag{5}$$

The goal of Bayesian statistics is to find the parameters $\boldsymbol{\theta}$ that best describe the data $\boldsymbol{D}$. The left side of Equation 5, $f(\boldsymbol{\theta}|\boldsymbol{D})$, describes the distribution of the parameters given the data and is called the *posterior distribution*. The mean or the median of the posterior distribution can be used as point estimate for $\boldsymbol{\theta}$. $f(\boldsymbol{D}|\boldsymbol{\theta})$ describes the function of the data $\boldsymbol{D}$ for known parameters $\boldsymbol{\theta}$ and is the probability mass function. It represents the probability of any observation $\boldsymbol{D}$ given a certain set of parameters $\boldsymbol{\theta}$. It is also referred to as likelihood function because it describes the likelihood of the data give that the parameters are true. $f(\boldsymbol{\theta})$ is called *prior distribution* and reflects the initial belief of the researcher about the true value of the parameters $\boldsymbol{\theta}$. The researcher must specify a prior distribution to perform Bayesian analysis. Finally, $f(\boldsymbol{D})$ reflects the distribution of the data given the model. However, since we are interested in $\boldsymbol{\theta}$ and $f(\boldsymbol{D})$ does not depend on it, it can be seen as a normalizing constant that ensures that the posterior density integrates to 1. Therefore, another formulation of Bayes' theorem is

$$f(\boldsymbol{\theta}|\boldsymbol{D}) \propto f(\boldsymbol{D}|\boldsymbol{\theta})\, f(\boldsymbol{\theta}) \tag{6}$$

The symbol "∝" denotes proportionality and reads as "is proportional to".

If the integral of a probability distribution is finite, this distribution is called proper (Rouder & Lu, 2005). In Bayesian analysis, the posterior distribution has to be proper to conduct inferences. A proper prior leads to a proper posterior distribution (sufficient condition). Conjugate priors are prior distributions that lead to posterior distributions that are necessarily of the same family of distributions (Edwards et al., 1963). This makes it easier to estimate posterior distributions. Beta priors and normal priors are examples for conjugate proper priors.

As described above, in a hierarchical MPT model, it is assumed that participants vary in their parameters according to a specified hierarchical distribution which is estimated from the data. In Bayesian analysis this means that a hierarchical prior has to be defined. The hierarchical distribution serves as prior for each individual. The hierarchical distribution is called first-stage prior. Second-stage priors are placed on the hierarchical distribution. The result of the hierarchical structure is that extreme estimates that can occur in traditional frequentist modeling, and mostly reflect noise, are closer to the mean in the hierarchical model (Rouder & Lu, 2005). However, as the number of observation increases, the effect of the priors decreases. Many researchers prefer non-informative priors. These are priors that are only vaguely informative to reduce the influence of the chosen prior. However, sometimes it may be rational to choose a more informative prior – especially if there is well-established knowledge about the parameters (e.g., from previous research). Priors can even serve as a filter for extreme data (Rouder & Lu, 2005).

Many authors prefer Bayesian analysis over traditional frequentist NHST analysis because of the underlying rationale for hypothesis testing (e.g., Edwards et al., 1963; Lee & Wagenmakers, 2005; Wagenmakers, 2007). Although appearing very subjective at the first glance, the prior has several advantages: (1) A researcher may have particular reasons for assuming that the parameters lie (or do not lie) in a special region. He or she can assign a very high prior probability to this region. Also, the researcher can assign low probabilities to rare events. Those events do nevertheless occur though very rarely. By assigning low probabilities to these events, they are not omitted entirely (Edwards et al., 1963). (2) Sometimes researchers confuse the null hypothesis (which is the probability of the data given the hypothesis) with the probability of the hypothesis given the data. This is made more explicit in the Bayesian framework (Wagenmakers, 2007). (3) The classical significance levels of $\alpha = .01$ and $\alpha = .05$ are arbitrary values. In most cases there is no justification for the choice of α (Wagenmakers, 2007). (4) Traditional inference sometimes tests against a null hypothesis that nobody would believe in (Erdwards et al., 1963). Even if the null hypothesis is reasonable, in real world applications it will never be exactly true and, therefore, it will always be rejected if the sample size is large enough

(Wagenmakers, 2007). (5) The *p* value depends on the hypothetical null hypothesis. That is, it depends on data that have never been observed (Wagenmakers, 2007).

In contrast to frequentist hypothesis testing, Bayesian inference has several advantages: It does not depend on the intention with which the data were collected. They can be collected until there is enough evidence in favor or against a hypothesis. The method is consistent and conceptually straightforward (Wagenmakers, 2007). Sequential updating is no problem, since the posterior distribution of the first observations becomes the prior for the later observations. It is conditioned on the data $\boldsymbol{D}$ that have been observed and does not depend on imaginary data. Another advantage of Bayesian modeling is that it comes with BCIs (sometimes called credible intervals) which represent what many people assume traditional CIs to be. A CI refers to the proportion of CIs that will include the real parameter with an infinite number of samples. They treat the parameter of interest as random and the CI as fixed. In contrast, BCIs treat the interval as random and the parameter of interest as fixed. The BCI represents the certainty that the parameter lies within the interval (e.g., J. B. Smith & Batchelder, 2010).

However, most of the reasons against traditional inference can be omitted by conducting power analyses to define an appropriate sample size and by reporting effect sizes as demanded by APA standards (American Psychological Association, 2010). In hierarchical modeling, the reasons to use Bayesian statistics are not only philosophical but also practical in nature: We know how to analyze hierarchical models only in the Bayesian framework (Rouder & Lu, 2005). Bayesian nonlinear hierarchical models are hardly tractable and very difficult to implement because they require complex integration that is not solvable analytically and computationally intensive (Rouder & Lu, 2005; Wagenmakers, 2007). However, in the last decades new techniques have been developed that allow for the analysis of previous intractable models including nonlinear hierarchical models. MCMC sampling made Bayesian analysis more tractable (Rouder & Lu, 2005).

3.3.1.2 MCMC sampling and WinBUGS

MCMC sampling is a method to find the marginal posterior distribution. A MCMC sample consists of a large number of draws from the target distribution. From this sample, knowledge about the model parameters can be obtained.

Explained very informally: Imaging you compete in a trampolining championship. As in many other sports, the results depend on the degree of difficulty of the routine (depending on how many turns you perform) and the form score (depending on how accurate you perform your routine). Trampolining has an additional score called the "time of flight" (TOF; a score that is added depending on the duration of the routine to foster athletes who can maintain greater height). The degree of difficulty does not vary within a routine but the form score and the TOF usually do. There is a relationship between the form score and TOF but it is not linear. You need a certain minimum height to perform a routine accurately. However, if you exceed a maximum height, the routines gets sloppy. If you perform the routines many, many times and count the form score and the TOF, this knowledge will help you to choose the routine that enables you to receive the highest scores and win the competition. The routine can be seen as the items that are used to assess your trampolining skills while the form score and TOF depend on the person.

Gibbs sampling is a special MCMC technique (e.g., Geman & Geman, 1984). It breaks up the target distribution into a series of conditional distributions and samples from these distributions. For example, if we want to obtain a sample from a posterior distribution specified by μ and σ, a Gibbs sampler could proceed like the following sequence:

1. We choose an arbitrary value for σ "$[\sigma]_1$" and sample $[\mu]_1$ from $\sigma|\mu,\boldsymbol{D}$

[choosing an arbitrary TOF score and sample the form score from looking at the form scores given this TOF score and your routine scores]

2. With $[\mu]_1$ we sample $[\sigma]_2$ from $\mu|\sigma,\boldsymbol{D}$

[take the form score and sample the TOF score from looking at the TOF scores given this form score and your routine scores]

3. With $[\sigma]_2$ we sample $[\mu]_2$ from $\sigma|\mu,\boldsymbol{D}$

[take this TOF score and sample the form score from looking at the form scores given this TOF score and your routine scores]

4. With $[\mu]_2$ we sample $[\sigma]_3$ from $\mu|\sigma,\boldsymbol{D}$

[take the form score and sample the TOF score from looking at the TOF scores given this form score and your routine scores]

etc.

The sampling is repeated many times and with different starting values (i.e., running different chains) until some basic regularities hold. Often, early draws of an MCMC chain depend on the starting values and therefore show poor convergence. Usually, early draws are discarded as a burn-in period which is not used for parameter estimation. It has been shown that under mild conditions (see Gilks, Richardson, & Spiegelhalter, 1996) these later draws represent samples from the posterior distribution. For drawing statistical inference from MCMC chains, it is necessary that the chains (starting from different starting values) have converged (i.e., the chain has reached a stationary distribution). Convergence can be checked using the $\hat{R}$ statistic (Gelman & Rubin, 1992). It compares the variances within the chains and between the chains. Under convergence, $\hat{R}$ is close to 1. Mostly $\hat{R} < 1.05$ is used as criterion for convergence (e.g., Klauer, 2010; Matzke et. al., 2013).

Gibbs sampling can be implemented via the program WinBUGS (Spiegelhalter et al., 2003). "BUGS" stands for Bayesian Inference Using Gibbs Sampling; the prefix "Win" is because it is only available for Windows. There are some similar programs like JAGS (Just Another Gibbs Sampler; Plummer, 2003), or OpenBUGS (Lunn, Spiegelhalter, Thomas, & Best, 2009). The implementations of the hierarchical MPT models presented here are WinBUGS implementations. $\hat{R}$ is provided by several programs like the R2WinBUGS package (Sturtz, Ligges, & Gelman, 2005) which also serves as an interface between R and WinBUGS.

WinBUGS is not tailored to particular models. Therefore, WinBUGS may converge very slowly if there is high autocorrelation between the chains (Matzke et al., 2013). Autocorrelation means that successive MCMC samples highly depend on each other (Rouder & Lu, 2005). This can happen if the data contain little information about a parameter, for example when parameters are close to the boundaries of the probability space. If there is high autocorrelation, one solution is to use a long burn-in period and to run relatively long chains and thin each chain (Matzke et al., 2013). Thinning by a factor t means that only samples from every tth iteration are retained. Another solution in case of high autocorrelation is to use a handmade Gibbs sampler that uses block-wise sampling for groups of correlated parameters (e.g., Klauer, 2010; Rouder, et al., 2007). Additionally, J. B. Smith and Batchelder (2010) as well as Matzke et al. (2013) provided WinBUGS implementations of the beta-MPT approach and the latent-trait approach, respectively. These implementations can be used and adapted as long as the convergence of the MCMC chains is monitored. A successful implementation of a MCMC chain results in a sample from the full posterior distribution from which model parameters can be calculated. WinBUGS provides means, standard deviations, and quantiles for all requested hierarchical, base, and individual parameters.

3.3.1.3 Model Fitting and Hypothesis Testing

In MPT modeling it is common to check whether the model fits the data by MLE or similar methods. The concept of goodness-of-fit cannot be adopted exactly for Bayesian hierarchical MPT modeling. However, the question whether the model accounts for the data and the comparison of different models can be assessed by several methods.

Posterior predictive checking is a method to assess whether the model describes the data well. Therefore, data are generated from the posterior distribution. Either the hyperdistributions or the person-level parameters can be used (Klauer, 2010). The predicted data are then compared to the data that have been observed. Similarly, Klauer proposed the statistics T_1 and T_2 to check for recovery of the observed frequencies and covariance structure. Additionally, small BCIs indicate that a parameter is very constraint by the data, whereas large BCIs reveal a larger uncertainty. However, model checks comparable to traditional goodness-of-fit tests are not available.

Models can be compared directly using the deviance information criterion (DIC; e.g., Spiegelhalter, Best, Carlin, & van der Linde, 2002). DIC is a Bayesian method for model comparison similar to AIC or BIC. All three information criteria trade off model fit and model complexity. Lower values indicate better model fit. The DIC is provided by WinBUGS. DIC, AIC, and BIC can be used to compare models in the Bayesian framework, but they do not to indicate whether a model describes the data well. They only access which of the models relatively fits better. Fit is not evaluated in absolute terms. Klauer (2010) also proposed another statistic called T_3. Model comparisons can be done by computing Bayes factors of two different models. For the Bayes factor (BF) it is necessary to compute the odds of one hypothesis relative to another hypothesis. Odds in favor of an event A are the probability that the event will happen devided by the probability that the event will not happen. Odds and probabilities can be translated into one another (Edwards et al., 1963)

$$\Omega(A) = \frac{P(A)}{1-P(A)} = \frac{P(A)}{P(\bar{A})}. \tag{7}$$

According to Bayes' theorem it holds that

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} \tag{8}$$

$$P(\bar{\theta}\big|D) = \frac{P(D|\bar{\theta})P(\bar{\theta})}{P(D)}. \tag{9}$$

This leads to

$$\Omega(\theta|D) = \frac{P(\theta|D)}{P(\bar{\theta}\big|D)} = \frac{\frac{P(D|\theta)P(\theta)}{P(D)}}{\frac{P(D\big|\bar{\theta})P(\bar{\theta})}{P(D)}}$$

$$= \frac{P(D|\theta)P(\theta)P(D)}{P(D)P(D\big|\bar{\theta})P(\bar{\theta})} = \frac{P(D|\theta)P(\theta)}{P(D\big|\bar{\theta})P(\bar{\theta})}$$

$$= \frac{P(D|\theta)}{P(D|\bar{\theta})}\Omega(\theta) = L(\theta; D)\Omega(\theta). \tag{10}$$

The posterior odds in favor of the parameters $\boldsymbol{\theta}$ given the data $\boldsymbol{D}$ are the prior odds multiplied by the ratio of the (conditional) probabilities of the data given the parameters and given the parameters' negation. The ratio of conditional probabilities $L(\boldsymbol{\theta}; \boldsymbol{D})$ is called the likelihood ratio in favor of the parameters $\boldsymbol{\theta}$ on the basis of the data $\boldsymbol{D}$ (Edwards et al., 1963). However, it is difficult to calculate $P(\boldsymbol{D}|\boldsymbol{\theta})$. This is the reason why a Bayesian hypothesis test always involves at least two different models (Wagenmakers, 2007).

One of the criticisms of Bayesian statistics on traditional frequentist statistics is that an unlikely null hypothesis is not sufficient reason to reject it, because the data may be even more unlikely under the alternative hypothesis. The posterior odds in favor of the null hypothesis H_0 versus the alternative hypothesis H_1 are given by

$$\frac{\mathrm{P}(H_0|\boldsymbol{D})}{\mathrm{P}(H_1|\boldsymbol{D})} = \frac{\mathrm{P}(\boldsymbol{D}|H_0)}{\mathrm{P}(\boldsymbol{D}|H_1)}\frac{\mathrm{P}(H_0)}{\mathrm{P}(H_1)} \tag{11}$$

The BF (e.g., Jeffreys, 1961, Wagenmakers, 2007) describes evidence coming from the data, that is, the change in odds from the prior to the posterior. It is indicated by the ratio $P(\boldsymbol{D}|H_0)/P(\boldsymbol{D}|H_1)$. A BF > 1 indicates evidence in favor of the null hypothesis H_0 (or the hypothesis that is placed in the numerator). A BF < 1 indicates evidence in favor of the alternative hypothesis H_1 (or the hypothesis that is placed in the denominator). According to Jeffreys (1961), the *BF* must be greater than 3 to indicate substantial evidence for the H_0, and smaller than 1/3 to indicate substantial evidence for the H_1.

To check for group differences, according to J. B. Smith and Batchelder (2010), independent samples can be compared by modeling separate hierarchical distributions for both groups and sampling the expected difference between both groups. If zero is extreme, this indicates a group difference. For dependent samples, J. B. Smith and Batchelder propose to model both groups within the same hierarchical model and afterwards compare the groups via a repeated measures ANOVA or *t*-test. However, the individual parameter estimates may be biased towards the combined sample mean. Additionally, individual estimates yielded by hierarchical modeling are no longer independent because the value of one individual depends on the values of the other individuals in the sample. This violates the assumptions of ANOVA and similar methods (Rouder & Lu, 2005). Therefore, it is recommended to use *BF*s.

3.3.2 Continuous Hierarchical MPT Approaches

As outlined above, continuous hierarchical MPT models assume that participants' (and/or items') parameters are drawn from a multivariate hyperdistribution $h(\boldsymbol{\theta}, \tau)$. In the following, I present three continuous hierarchical MPT approaches: The beta MPT approach (J. B. Smith & Batchelder, 2010), the latent-trait approach (Klauer, 2010), and the crossed-random effects approach (Matzke et al., 2013). The beta-MPT and the latent-trait approach both assume a single source of heterogeneity (participants or items). The crossed-random effects approach varies both items and participants. However, the crossed-random effects approach is not adaptable for my purposes.

3.3.2.1 Beta-MPT Approach

In the beta-MPT approach (J. B. Smith & Batchelder, 2010) participants' parameters are drawn from a multivariate distribution consisting of independent marginal beta distributions. The beta distribution is a very flexible distribution that is defined by the parameters α and β ranging from zero to infinity. The range of the beta distribution is (0, 1). This is very convenient because the model parameters represent probabilities which also lie between zero and one.

The density of the beta distribution for one parameter θ_s is defined by

$$g(\theta_s|\alpha_s, \beta_s) = \frac{\Gamma(\alpha_s+\beta_s)}{\Gamma(\alpha_s)\Gamma(\beta_s)}\theta_s^{\alpha_s-1}(1-\theta_s)^{\beta_s-1}, \tag{12}$$

where Γ(x) is the gamma function which is $\Gamma(x) = (x-1)!$ for positive integers. The multivariate hyperdistributions are given by

$$g(\boldsymbol{\theta}) = \langle\theta_s\rangle_{s=1}^S|\tau = \langle\alpha_s\beta_s\rangle_{s=1}^S = \prod_{s=1}^S g(\theta_s|\alpha_s, \beta_s). \tag{13}$$

The hyperparameter τ has parameter space $\Lambda_\tau = (0, \infty)^{2S}$.

When α and β are both greater than one, the distribution is unimodal. When both parameters equal one, it is uniform. When they are smaller than one, the distribution is u-shaped (see Figure 3). When either α > 1 or β > 1, and the other one is smaller, the distribution is monotonically increasing or decreasing respectively. However, in my implementations, I restricted both parameters to be greater than one, so that the distribution is unimodal.

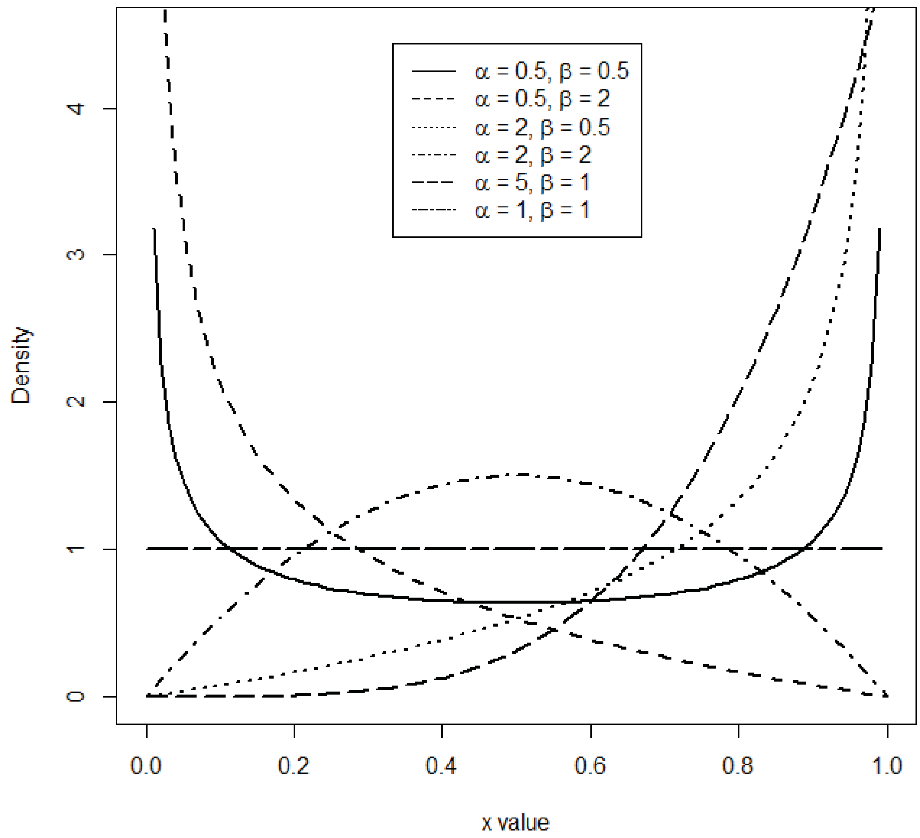


Figure 3. Probability density functions of the beta distribution depending on different values for α and β.

The beta distribution has mean

$$E(\theta_s) = \mu_s = \frac{\alpha_s}{\alpha_s+\beta_s} \tag{14}$$

and variance

$$Var(\theta_s) = \sqrt{\sigma_s} = \frac{E(\theta_s)[1-E(\theta_s)]}{\alpha_s+\beta_s+1} = \frac{\frac{\alpha_s}{\alpha_s+\beta_s}\left[1-\frac{\alpha_s}{\alpha_s+\beta_s}\right]}{\alpha_s+\beta_s+1} = \frac{\frac{\alpha_s}{\alpha_s+\beta_s}\left[1-\frac{\alpha_s}{\alpha_s+\beta_s}\right](\alpha+\beta)^2}{(\alpha_s+\beta_s+1)(\alpha+\beta)^2}$$

$$= \frac{\left[1-\frac{\alpha_s}{\alpha_s+\beta_s}\right]\alpha_s(\alpha_s+\beta_s)}{(\alpha_s+\beta_s+1)(\alpha_s+\beta_s)^2} = \frac{\alpha_s(\alpha_s+\beta_s)-\alpha_s^2}{(\alpha_s+\beta_s+1)(\alpha_s+\beta_s)^2} = \frac{\alpha_s^2+\alpha_s\beta_s-\alpha_s^2}{(\alpha_s+\beta_s+1)(\alpha_s+\beta_s)^2} = \frac{\alpha_s\beta_s}{(\alpha_s+\beta_s+1)(\alpha_s+\beta_s)^2} \tag{15}$$

which are comparable to traditionally obtained mean and variance for aggregated data. Most importantly, every participant has an own set of parameters.

The variance approaches zero when α and β approach infinity. This is a problem for parameter estimation. To avoid this problem and to improve the interpretability of the hyperdistribution, α and β can be reformulated to mean and standard deviation using equations (14) and (15). This leads to a new hyperparameter $\boldsymbol{\tau}^* = \langle \boldsymbol{\mu}, \boldsymbol{\sigma} \rangle$ with $\boldsymbol{\mu} = \langle \mu_s \rangle_{s=1}^{S}$ and $\boldsymbol{\sigma} = \langle \sigma_s \rangle_{s=1}^{S}$ and a new hyperdistribution $g^*(\boldsymbol{\Theta}|\tau^*) = \prod_{s=1}^{S} g^*(\theta_s|\mu_s\sigma_s)$ which is the reparameterized beta distribution.

J. B. Smith and Batchelder (2010) provided a frequentist and a Bayesian approach for parameter estimation within the beta-MPT framework. Both approaches assume that the underlying hyperparameter τ generated the base model parameters $\boldsymbol{\theta}$ through Equation (13) and from that the individual parameters. In the frequentist approach, the goal is to find the MLE that maximizes the likelihood function for the observed data. However, the frequentist marginal likelihood function involves high-dimensional integrals that do not possess an analytical solution in most cases.

With the Bayesian framework an analytical solution is also impossible for most applications but there exist solutions using MCMC methods. The hyperprior π(τ*) is interpreted as the initial belief about the hyperparameters τ^* – prior to observing the data $\boldsymbol{D}$. The posterior distribution is computed by applying Bayes' theorem as well as the properties of the MPT model and the hyperdistribution. Of course, the posterior distribution depends on the hyperprior $\pi(\tau *)$. Therefore, the researcher must choose suitable numbers for α and β that reflect the belief about the possible values before the data are collected (e.g., Rouder & Lu, 2005). However, the influence of the prior distribution diminishes as the amount of data increases, for example due to more participants or more observations per participant. When α = β = 1, the researcher beliefs all values of Θ being equally likely before the data are collected.

J. B. Smith and Batchelder (2010) provided a WinBUGS code for the pair-clustering model. The Gibbs sampler used by WinBUGS sometimes shows large autocorrelation when the variance is small. This slows the process of convergence. Therefore, convergence has to be monitored carefully.

3.3.2.2 Latent-Trait Approach

Cognitive abilities do not only vary between participants and items but also they are often highly correlated (Matzke et al., 2013). Therefore, it is necessary to incorporate parameter correlations in the hierarchical modeling approaches. The beta-MPT approach does not incorporate parameter correlations. Therefore, parameter estimates can still be biased (Klauer, 2010; Matzke et al., 2013).

The latent-trait approach (Klauer, 2010) takes into account that parameters can be correlated. It assumes that the parameters are drawn from a multivariate normal distribution. A normal distribution ranges from minus infinity to infinity and does not have the same space as the MPT model parameters. Therefore, the approach transforms parameters from the interval (0, 1) to the real line and assumes that these transformed parameters follow a multivariate normal distribution with mean and covariance matrix to be estimated from the data. The transformation from the interval (0, 1) to the real line is done by a probit link. The model is therefore reparameterized by means of the new transformed parameters with $\theta_{ns}^{prt} = \Phi^{-1}(\theta_{n,s})$, where Φ is the cumulative distribution function of the standard normal distribution. The transformed parameters are assumed to follow a multivariate normal distribution with mean $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{S_{part}}$ and are reparametrized as follows $\theta_{ip}^{prt} = \mu_s + \delta_{part_{ns}}$, where $\mu_s$is the group mean (on the real line) for parameter θ_s and $\delta_{part_{ns}}$is the *n*th participant's deviation from it. The δ_{part_n} parameters are drawn from a zero-centered multivariate distribution with variance-covariance matrix S_{part}.

The latent-trait approach was originally formulated by Klauer (2010). However, he did not present an implementation. Recently, Matzke et al. (2013) published a WinBUGS implementation for the pair-clustering model. Their version shows slight modifications from Klauer's approach. I will mention them wherever they exist.

Like the beta-MPT approach, the latent-trait approach uses Bayesian inference and MCMC chains to estimate the parameters. Traditional maximum-likelihood approaches would be (in most cases) too computationally intense. For Bayesian inference, it is necessary to define prior distributions: The hyperprior distribution represents the prior beliefs about the hyperdistribution, that is μ_s and S_{part}. Klauer (2010) assumes independent normal distributions with $\mu_{\mu_s} = 0$ and $\sigma^2_{\mu_s} = 100$. However, Matzke et al. (2013) chose independent normal distributions with $\mu_{\mu_s} = 0$ and $\sigma^2_{\mu_s} = 1$ because it corresponds to a uniform distribution on the probability scale (Rouder & Lu, 2005). For the covariance matrix S_{part}, Klauer as well as Matzke et al. used a scaled Inverse-Wishart distribution (Gelman & Hill, 2007) which imposes a uniform distribution for the correlation coefficients and allows for a more free estimation of covariances than an Inverse-Wishart distribution. The advantage of the Inverse-Wishart distribution is that it results in an uninformative uniform prior distribution between -1 and 1 for the correlation parameters $\rho_{ss'}$. The disadvantage is that the Inverse–Wishart distribution imposes a very restrictive prior on the standard deviations. The scaled Inverse-Wishart distribution has scale parameters, $\xi_{part} = [\xi_{part_1,\ldots,}\xi_{part_s}]$ (Gelman & Hill, 2007) and still implies a uniform prior distribution for the correlation parameters. The variance–covariance matrix S_{part} is then reformulated into $S_{part} = Diag(\xi_{part})T_{part}Diag(\xi_{part})$, where $Diag(\xi_{part})$ is a diagonal matrix containing the scale parameters. T_{part} follows an Inverse-Wishart distribution with 1+3 degrees of freedom and with a scale matrix that is set to the 3 × 3 identity matrix. The standard deviations can be obtained by

$$\sigma_{part_p} = |\xi_{part_p}| \times \sqrt{T_{part_{pp'}}} \tag{16}$$

The correlation parameters are given by

$$\rho_{part_{ss'}} = \frac{\xi_{part_s}\xi_{part_{s'}}T_{part_{ss'}}}{|\xi_{part_s}|\sqrt{T_{part_{ss}}}\times|\xi_{part_{s'}}|\sqrt{T_{part_{s's'}}}} \tag{17}$$

The ξ_s parameters are given uniform distributions ranging from 0 to 100 (e.g., Gelman & Hill, 2007). Note that Klauer (2010) used normal distributions with a mean of one and a variance of 100 as prior for the scaling parameters, but this results in convergence problems for the variance and the correlation parameters in WinBUGS (Matzke et al., 2013). From these prior distributions the individual parameters θ_{ns}^{prt} are estimated which are still on the real line. Matzke et al. (2013) conducted a parameter recovery study that indicated that the WinBUGS version of the latent-trait pair-clustering model – like Klauer's original version – adequately recovered the true parameter values.

3.3.2.3 Crossed-Random Effects Approach

The beta-MPT approach (J. B. Smith & Batchelder, 2010) and the latent-trait approach (Klauer, 2010) both allow either participants or items to be variable. Matzke et al. (2013) introduced the crossed-random effects pair-clustering model that is an extension of the latent-trait approach and incorporates both participant and item heterogeneity simultaneously. This requires separate parameters for each participant–item combination resulting in problems of model identification. To reduce the number of required parameters, Matzke et al assume that participant and item effects combine additively on the probit scale (e.g., Rouder & Lu, 2005; Rouder et al., 2007; Rouder, Lu, Morey, Sun, & Speckman, 2008):

$$\theta_{nmsk}^{prt} = \mu_{sk} + \delta_{part_{nsk}} + \delta_{part_{msk}} \tag{18}$$

Participant variability is assumed to follow a multivariate normal distribution while item heterogeneity is assumed to follow independent normal distributions. Thus, participant effects are allowed to be correlated a priori, whereas item effects are not.

Matzke et al. (2013) also present a WinBUGS implementation for this approach. Moreover, they show in a simulation study that the WinBUGS implementation of the crossed-random effects pair-clustering model recovers the true parameter values well. However, the approach is not adaptable to most other MPT models. In particular, MPT models where parameter constraints are required between different subtrees have the necessary condition that each item occurs in the relevant trees to use across-subtree constraints for the item effects. This is not the case for the MPT model of event-based PM (R.E. Smith & Bayen, 2004) and the 2HTSM (Bayen et al. 1996). Hence, we cannot use the crossed-random effects approach for the applications presented here.

4 Overview of the Studies

This thesis uses hierarchical MPT modeling to discover relationships that would not have been discovered without hierarchical modeling. In all studies, we use median and standard deviation to summarize the posterior distributions because the posterior median is less sensitive to outliers than the posterior mean and especially preferable for non-symmetric distributions like the beta distribution (Klauer, 2010; Matzke et al. 2013).

4.1 Study 1: Beta-MPT Modeling of the 2HTSM: Testing the Probability-Matching Account

Source monitoring describes determining the source of an item. If people do not remember the source, they have to guess. Guessing strategies in source monitoring are manifold. People may rely on schemas (*schema-based guessing*; Bayen, et al., 2000) which means they rely on general schematic knowledge to guess the source of an item. For example, a doctor talking about medicine is more likely than a lawyer talking about medicine. The probability-matching account of source guessing (Spaniol & Bayen 2002) states that people guess according to the perceived contingency between items and sources. For example, if in an experiment the doctor says 75 % of statements that are typical for a lawyer and the lawyer says 75% of the doctor-typical statements, participants match their guessing behavior accordingly (Bayen & Kuhlmann, 2011). Only if they do not have a contingency representation, they rely on schema-based guessing. For example, if participants are not able to develop a contingency representation due to a second task in the experiment, they rely on schema-based guessing (Bayen & Kuhlmann, 2011). Probability matching has been observed in source monitoring (e.g., Bayen & Kuhlmann, 2011; Erdfelder & Bredenkamp, 1998) and other tasks (e.g., Buchner, Erdfelder, & Vaterrodt-Plünnecke, 1995; Ehrenberg & Klauer, 2005; Estes & Straughan, 1954). The probability-matching account predicts that people guess according to their *perceived* contingency which does not necessarily equal the true contingency. Hence, people who differ in their perceived contingency should also differ in their source guessing behavior. Therefore, the probability-matching account also predicts individual differences in source-guessing. This has already been suggested as an explanation (Spaniol & Bayen, 2002) but has never been tested. The 2HTSM (Bayen et al., 1996) as described in Chapter 2.1 separates memory and guessing in source monitoring. Using the beta-MPT approach we were able to account for individual differences in source-guessing. We hypothesized a positive correlation between perceived contingency and source-guessing parameter g of the 2HTSM. Additionally, we hypothesized that participants with bad source memory should be less able to realize the actual contingency and therefore should be further apart from the actual contingency. Hence, there should be a negative correlation between source-memory parameter d and the difference from perceived contingency pc to actual contingency of .5, $|pc - .5|$.

Forty-eight participants took part in a standard laboratory source-monitoring experiment containing established doctor-lawyer materials. In this experiment, both the doctor and the lawyer presented half of the doctor-typical statements and have of the lawyer-typical statements. Thus, there was a true contingency of zero between expectedness of the statement and source. In the first condition, participants were told about the profession of the sources before the encoding phase started (*encoding* condition). In the second condition, participants were not told about the profession of the sources after encoding but before the test started (*retrieval* condition).

Perceived contingency pc was measured by asking participants how many of the doctor-statements had been presented by the doctor and how many of the lawyer-statements had been presented by the lawyer. These absolute judgments were combined and transformed into relative judgments for the expected source. Thus, a perceived contingency of $pc = .5$ matches the true contingency. Perceived contingencies greater than .5 mean that participants perceived a higher contingency between items and their expected sources, whereas values smaller than .5 mean that participants perceived a higher contingency between items and their non-expected sources.

Parameter estimates were conducted with MultiTree and with the beta-MPT version of Submodel 4 of the 2HTSM. Submodel 4 assumes that item as well as source memory are equal for both sources and that the probability of noticing that an item is new is equal to the probability of recognizing an item as old. For the analysis, doctor and lawyer materials were combined into expected and unexpected statements and sources.

The results showed strong evidence for heterogeneity, especially for the source-guessing parameter *g* as indicated by the credible intervals of the standard deviation in Table 2.

Table 2

Posterior distributions of the parameters of the hierarchical beta distributions

	Encoding Condition				*Retrieval* Condition			
	M	*SD* [95% BCI]	α	β	*M*	*SD* [95% BCI]	α	β
D	.25	.04 [.01−.11]	230.88	678.19	.33	.15 [.09−.22]	3.10	6.38
d	.67	.10 [.01−.25]	55.50	28.05	.22	.14 [.02−.24]	5.59	34.64
b	.45	.18 [.13−.24]	3.17	3.90	.52	.15 [.10−.21]	5.73	5.20
g	.58	.20 [.15−.26]	3.01	2.15	.73	.19 [.15−.24]	3.17	1.11

Note. D = probability of item recognition; *d* = probability of remembering the source; *b* = probability of guessing that an item is old; *g* = probability of guessing that an item was presented by the schematically expected source; BCI = Bayesian confidence interval of the standard deviation. Adapted from "Hierarchical modeling of contingency-based source monitoring: A test of the probability-matching account," by N. R. Arnold, U. J. Bayen, B. G. Kuhlmann, and B. Vaterrodt, 2013, *Psychonomic Bulletin & Review, 20,* p. 331. Copyright 2013 by Springer.

Replicating previous studies, participants in the retrieval condition showed a significantly larger guessing bias than participants in the encoding condition with traditional analyses $G^2(1)$ =

97.05, $p < .01$. However, as shown in Table 3, credible intervals do overlap for the beta-MPT analysis, thereby indicating no substantial difference between the conditions.

Table 3

Group parameter estimates for schematically expected items with traditional MPT and beta-MPT

	Encoding Condition				Retrieval Condition			
	Traditional MPT		Beta-MPT		Traditional MPT		Beta-MPT	
	M(SD)	95% CI	*M (SD)*	95% BCI	*M (SD)*	95% CI	*M (SD)*	95% BCI
D	.24(.02)	[.20−.30]	.25(.03)	[.20−.30]	.33(.02)	[.28−.38]	.33(.04)	[.25−.41]
d	.70(.12)	[.46−.92]	.67(.11)	[.46−.89]	.28(.08)	[.13−.43]	.22(.09)	[.07−.40]
b	.45(.02)	[.41−.48]	.45(.04)	[.37−.53]	.53(.02)	[.49−.56]	.52(.04)	[.45−.60]
g	.60(.02)	[.55−.64]	.58(.05)	[.49−.67]	.78(.02)	[.75−.81]	.73(.04)	[.64−.81]

Note. The parameters represent probability estimates. *D* = probability of item recognition; *d* = probability of remembering the source; *b* = probability of guessing that an item is old (chance level is .5); *g* = probability of guessing that an item was presented by the schematically expected source (estimates higher than the chance level of .5 indicate guessing bias towards the schematically expected source; estimates lower than .5 indicate guessing bias towards the schematically unexpected source); CI = confidence interval; BCI = Bayesian confidence interval. Adapted from “Hierarchical modeling of contingency-based source monitoring: A test of the probability-matching account,” by N. R. Arnold, U. J. Bayen, B. G. Kuhlmann, and B. Vaterrodt, 2013, *Psychonomic Bulletin & Review, 20,* p. 331. Copyright 2013 by Springer.

The beta-MPT approach enabled us to calculate individual parameter estimates which allowed us to test our hypotheses using a correlational approach. The first hypothesis was confirmed: We found a positive correlation between perceived contingency and the source-guessing parameter *g*, (*encoding* condition: $r = .45$, $p = .02$, one-tailed; *retrieval* condition: $r = .55$, $p < .01$, one-tailed). The higher participants perceived the contingency between items and expected sources, the higher was the probability to guess that an item was presented by the expected source. This finding is important support for the probability-matching account.

Second, we hypothesized a negative correlation between source-memory parameter *d* and the difference from perceived contingency *pc* to actual contingency of .5, $|pc-.5|$. This was confirmed only in the *retrieval* condition, $r = -.42$, $p = .04$ (one-tailed) but did not reach significance in the *encoding* condition, $r = .05$, $p = .81$ (one-tailed). Source guessing, thus, was independent of source memory in the encoding condition but not in the retrieval condition. It is possible that, participants in the *encoding* condition learned about the contingency during the study phase and did not need a good memory to adjust their guessing. For participants in the retrieval condition, it was difficult to recognize the contingency during encoding. Thus, it might have been easier to learn about the contingency when schemas were provided.

Both findings underscore the importance of an individual differences approach. We were, for the first time, able to show the implications of the probability-matching account on the individual level.

4.2 Study 2: Beta-MPT Modeling of PM: PM and mental health

In the second study, we applied the beta-MPT approach to the multinomial model of event-based PM. The model is explained in Chapter 2.2. The PM paradigm is especially suitable for the hierarchical modeling approach. In a standard event-based laboratory task, participants work on an ongoing task. In this study they worked on a color-matching task. In a color-matching task, they see four colored rectangles in a row. Thereafter, a colored word appears, and participants have to judge whether it has the same color as one of the rectangles presented before. This is a very resource-demanding ongoing task. While working through the ongoing task, participants also have to work on the PM task; that is, if a PM target (in the present study a special word) appears, participants are asked not to answer the ongoing task but instead press a special key. The PM targets occur very rarely. In our study, 10 % of the items were PM targets. Since there are only few PM targets, there are often response categories for the MPT model (e.g., erroneous answers for PM targets) that are sparsely filled. For individual participants, some categories may even be empty. Therefore, it is even more inappropriate to calculate a separate model for each participant. With the hierarchical structure imposed by the beta-MPT model it is possible to calculate group parameters and individual parameter estimates for each participant.

PM involves remembering to perform an action in the future and is crucial for everyday life functioning. It includes remembering appointments or taking medicine and is therefore important for mental health. It is very important to separate the retrospective and the prospective component. If the prospective component – remembering *that* you have to do something – is impaired, patients can be taught how to use different reminders or special routines. If the retrospective component – remembering *when* to perform the action – is impaired, patients can be taught mnemonic techniques.

We were especially interested in the effects of anxiety and depression on PM. According to the resource allocation model of depression (Ellis & Ashbrook, 1988), depression limits the cognitive capacity that is available. This should impair particularly resource-demanding self-initiating processes. However, previous findings regarding depression-related PM impairments are mixed. Depression has been found to influence time-based PM tasks (e.g. Kliegel & Jäger, 2006) which require more self-initiation. Impairments in event-based PM have only been found with multiple PM targets and non-focal tasks (Altgassen, Kliegel, & Martin, 2009). Focality describes whether the PM target requires the same cognitive processes as the ongoing task. Non-focal tasks require more cognitive resources (e.g., McDaniel & Einstein, 2000). We therefore used a non-focal task with multiple targets which is very resource demanding. Because depression has been shown to impair recognition memory (Brand, Jolles, & Gispen-de Wied, 1992; Hertel & Milan, 1994; Ramponi, Murphy, Calder, & Barnard, 2010; Watts, Morris, & MacLeod, 1987), we hypothesized a negative correlation with both memory components of PM.

We also examined the influence of state and trait anxiety on PM. It has been shown that trait and state anxiety impair working memory capacity (Eysenck, 1985; MacLeod & Donnellan, 1993; Stout & Rokke, 2010). Working memory impairments have been found to impair the prospective component of PM (R.E. Smith & Bayen, 2005; R.E. Smith, et al., 2011). Since we used a highly resource-demanding task, we hypothesized a negative correlation between the prospective component and both trait and state anxiety. The findings on the influence of anxiety on recognition memory are rather sparse but do not indicate an anxiety-related impairment for recognition memory for both kinds of anxiety (Beato, Pulido, Pinho, & Gozalo, 2013). Therefore, we did not hypothesize a correlation between state and trait anxiety and the retrospective component *M*.

One hundred twenty-nine students took part in the study. The ongoing task was a color matching task and the PM task was to press the space bar when special words appeared. To gain enough data especially for PM targets, participants saw 336 colored words. Thirty of them were PM targets which appeared equally often in match and non-match trials. Following this task, participants filled out several questionnaires. Depression was measured by the Beck Depression Inventory II (BDI-II; Beck, Steer, & Brown, 1996; German translation by Hautzinger, Keller, & Kühner, 2006) and the depression subscale of the Hospital Anxiety Depression Scale (HADS-D; Herrmann, Buss, & Snaith, 1995). Trait anxiety was measured by the anxiety scale of the HADS-D and the trait anxiety subscale of the State-Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, & Lushene, 1971; German translation by Laux, Glanzmann, Schaffner, & Spielberger, 1981). State anxiety was measured by the state scale of the STAI. Parameter estimates obtained with the beta-MPT version of the multinomial model for event-based PM (R.E. Smith & Bayen, 2004) can be found in Table 4. As indicated by the credible intervals of the standard deviation of the posterior distribution (that do not approach zero), there is strong evidence for parameter heterogeneity.

Table 4

Parameters of the Hierarchical Beta Distributions

	M [95% BCI]	*SD* [95% BCI]	α	β
C_1	.71 [.67 − .74]	.20 [.18 − .23]	2.82	1.17
C_2	.89 [.88 − .91]	.07 [.06 − .08]	16.55	1.96
P	.76 [.73 − .79]	.18 [.16 − .20]	3.58	1.10
M	.89 [.86 − .91]	.10 [.08 − .12]	8.24	1.10

Note. C_1 = probability of detecting a color match; C_2 = probability of detecting a color nonmatch; *P* = prospective component of PM; *M* = retrospective component of PM; BCI = Bayesian confidence interval. *M* = Mean of the hierarchical beta-distribution; *SD* = standard deviation of the hierarchical beta-distribution. Posterior distributions describe the probability distributions of the parameters conditional on the data. Adapted from "Is prospective memory related to depression and anxiety? by N. R. Arnold, U. J. Bayen, and M F. Böhm, 2014, *manuscript submitted for publication.*

The correlations between the individual parameter estimates and the questionnaire scores are shown in Table 5. The hypothesis that depression is negatively correlated with the prospective and the retrospective component was not confirmed for both measures of depression. Equally, both measures of trait anxiety showed no correlation with any of the components. Only state anxiety correlated with the prospective component *P* as hypothesized, $r = -.18$, $p = .02$ (one-tailed). Traditionally, PM is measured as the number of PM targets that were correctly responded to, called *PM hits*. Like the prospective component, *PM hits* was only related to state anxiety, $r = -.19$, $p = .02$ (one-tailed).

Table 5

Correlations between Parameters M *and* P *of the Beta-MPT Model, Prospective-Memory Performance and Depression and Anxiety Scores*

Test	N		*P*	*M*	PM Hits
BDI-II	128	*r*	−.08	.10	−.03
		p	.18	.14	.36
HADS-D	129	*r*	−.04	.03	−.03
depression		*p*	.34	.37	.37
HADS-D	129	*r*	−.01	-.01	−.01
anxiety		*p*	.47	.48	.46
STAI state	129	*r*	−.18*	−.07	−.19*
		p	.02	.21	.02
STAI trait	129	*r*	−.08	.05	−.06
		p	.18	.31	.27

Note. PM = prospective memory; C_1 = probability of detecting a color-match; C_2 = probability of detecting a color-nonmatch; *P* = prospective component of PM; *M* = retrospective component of PM; BDI-II = Beck Depression Inventory; HADS-D = Hospital Anxiety and Depression Scale – German version; STAI = State-Trait Anxiety Inventory. * $p < .05$ (one-tailed). Adapted from "Is prospective memory related to depression and anxiety? by N. R. Arnold, U. J. Bayen, and M F. Böhm, 2014, *manuscript submitted for publication.*

The results from our study are consistent with the results obtained by Kliegel and Jäger (2006): We did not find a relationship between depression and event-based PM but between state anxiety and event-based PM with a similar task. The results are, however, not consistent with Altgassen et al. (2009) who found a relationship with event-based PM but with a clinically depressed sample. The difference between the relationship of trait and state anxiety with PM underlines the necessity of differentiating between state and trait anxiety. Our study is limited in the way that we used a student sample. Depression scores may not have been high enough to impair PM performance. However, obtaining a clinical sample of an appropriate sample size is very difficult to achieve. Additionally, the group parameter for the retrospective component *M* was very high with .89. Some participants answered all PM targets correctly and therefore there may have been a restriction of range.

Again, these findings highlight the importance of an individual differences approach. The traditional measure of PM hits cannot provide information about the different components of PM. However, the distinction between the prospective and the retrospective component is crucial for interventions. The MPT separates these components. Moreover, hierarchical MPT modeling is especially useful for MPT models that are prone to sparsely filled categories.

4.3 Study 3: Beta-MPT and latent-trait approach: PM and working memory capacity

The third study included reanalyzes of three PM experiments from two studies (R. E. Smith & Bayen, 2004; R. E. Smith, et al., 2011) investigating the relationship between PM and working memory (WM). Individual differences in WM span contribute to the variability in PM performance. This holds especially for non-focal tasks that require more cognitive resources (e.g., Brewer, Knight, Marsh, & Unsworth, 2010). Indeed, several studies found a relationship between WM capacity and PM performance (e.g., Ball, Knight, Dewitt, & Brewer, 2013; Brewer et al., 2010; Cherry & LeCompte, 1999; Einstein, McDaniel, Manzi, Cochran, & Baker, 2000; Reese & Cherry, 2002; R. E. Smith, 2003; R. E. Smith & Bayen, 2005; R. E. Smith et al., 2011; West & Craik, 2001) but only those of R. E. Smith and Bayen (2005) and R. E. Smith et al. (2011) had the appropriate data structure to use the MPT model of event-based PM. Both studies already used MPT modeling but with aggregated data. We argue that it is more useful to use hierarchical modeling because the parameter estimates are less biased and it is not necessary to fall back on extreme group analysis or median splits.

We reanalyzed these experiments using the beta-MPT approach (J. B. Smith & Batchelder, 2010) and with the latent-trait approach (Klauer, 2010). Although the latent-trait approach was published by Klauer in 2010, Matzke et al. (2013) only recently published a WinBUGS implementation that can be adapted to other MPT models. Thus, this is the first application of the latent-trait approach apart from the examples in the papers of Klauer (2010) and Matzke et al. (2013). Furthermore, it is the first time both approaches are applied to the same data. We report correlations with traditional *p* values and with BFs to be more consistent with the Bayesian inference used for parameter estimation.

Twenty participants took part in the first experiment by R. E. Smith and Bayen (2005). The ongoing task was a sentence-verification task and the PM task was to press the F1 key when specific words appeared. With traditional NHST, we found a significant correlation between WM span and the prospective component *P* for both approaches (beta-MPT: $r = .40$, $p = .04$ (one-tailed); latent-trait: $r = .41$, $p = .04$ (one-tailed)). However, the BFs were 0.74 and 0.82, respectively, thereby indicating neither support for the presence nor for the absence of a correlation. For the retrospective component *M*, the correlations did not reach significance for both approaches (beta-MPT: $r = .19$, $p = .21$ (one-tailed); latent-trait: $r = .17$, $p = .24$ (one-tailed)). The BFs were smaller than 1/3 thereby lending support for the absence of a correlation.

The second experiment by R. E. Smith and Bayen (2005) used the same task but included high WM load. Participants had to repeat the last word of each of the last four sentences every fourth sentence. With traditional NHST, we did not find significant correlations between WM span and any PM component, all $r < .35$, all $p > .06$ (one-tailed). However, none of the BFs was large or small enough to indicate decisive evidence either in favor or against a correlation. Therefore, this time, we did not concur with the previous results obtained by R. E. Smith and Bayen (2005).

The previous experiments just reported a small number of participants. Therefore, the power to detect a medium effect with traditional NHST was less than .50. For the Bayesian approach, we also did not have enough information to report evidence that supports either presence or absence of a correlation. The study of R. E. Smith et al. (2011) was conducted with 413 participants. This leads to more information and to a much greater power. The power was larger than .99 to detect a medium effect and .65 to detect a small effect. The study used a lexical decision task as ongoing task and the PM task was to press the F1 key when specific syllables appeared. The authors also used another measure of WM span. This gave us the opportunity to replicate the results with different measures. R. E. Smith et al. conducted an extreme group analysis and thereby omitted half of the data. Our analysis included data from all participants.

With NHST, we found a significant correlation between WM span and the prospective component P for both hierarchical modeling approaches (beta-MPT: $r = .15$, $p < .01$, one-tailed; latent-trait: $r = .15$, $p < .01$, one-tailed). The BF was 3.67 for the beta-MPT approach, indicating substantial support for a correlation. The BF for the latent-trait approach was 2.71, failing to reflect substantial support for a correlation. For the retrospective component M, the correlations failed to reach significance for both approaches (beta-MPT: $r = .01$, $p = .92$, one-tailed; latent-trait: $r = .04$, $p = .40$, one-tailed). The BFs were smaller than 1/3 thereby indicating support for the absence of a correlation. Thus, the results replicate the results of the first experiment with a larger sample size and concur with the results obtained by R. E. Smith et al. (2011).

For all three experiments reported here, we also compared the hierarchical modeling approaches in terms of model fit and differences in the parameter estimates both on group and on individual level. Posterior distributions of the population level parameters as well as traditional MPT parameters obtained using MultiTree as well as model fits can be found in Table 6.

For the experiments of R. E. Smith and Bayen (2005) the BCIs overlap for all parameters indicating no difference between the parameter estimates. For the experiment of R. E. Smith et al. (2011) BCIs do not overlap for any of the parameters meaning that the parameter estimates of the approaches differ. However, this experiment had much more observations due to more participants. This leads to much smaller BCIs and CIs. On an individual level, parameter estimates did not differ significantly between the beta-MPT and the latent-trait approach, all $p <$ .05 and $BF < 2/3$. Only the retrospective component in the experiment by R. E. Smith et al. (2011) shows higher beta-MPT estimates than latent-trait parameter estimates, $t(412) = 7.47$, $p < .01$, $BF >$ 1,000,000. No approach showed consistently larger or smaller correlations than the other approach in any of the experiments. In terms of DIC, in the two experiments of R. E. Smith and Bayen (2005) the latent-trait approach fit the data better than the beta-MPT approach. However, in the study of R. E. Smith et al. (2011), the beta-MPT showed a smaller DIC indicating better model fit than the latent-trait approach. We therefore conclude that there is no clear advantage of any of the two approaches. However, it is possible to explicitly model parameter correlations with the latent-trait approach. This speaks in favor of the latent-trait approach.

Table 6

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

		Beta-MPT		*Latent-Trait*		*MultiTree*	
		M	[95% *BCI*]	*M*	[95% *BCI*]	*M*	[95% *CI*]
	P	.79	[.70 − .86]	.84	[.73 − .94]	.81	[.76 − .85]
2005	*M*	.96	[.92 − .98]	.98	[.94 − 1]	.97	[.94 − .99]
Exp 1	C_1	.90	[.85 − .93]	.92	[.87 − .95]	.90	[.88 − .93]
	C_2	.77	[.71 − .82]	.78	[.72 − .84]	.78	[.74 − .81]
			DIC = 421.5		DIC = 376.0		$G^2(4) = 2.79$
	P	.61	[.50 − .71]	.70	[.46 − .88]	.65	[.59 − .70]
2005	*M*	.95	[.91 − .98]	.97	[.92 − 1]	.96	[.93 − .99]
Exp 2	C_1	.81	[.75 − .85]	.82	[.77 − .87]	.81	[.78 − .84]
	C_2	.77	[.70 − .83]	.76	[.72 − .86]	.78	[.74 − .81]
			DIC = 474.8		DIC = 435.1		$G^2(4) = 2.98$
	P	.64	[.61 − .66]	.72	[.68 − .76]	.67	[.66 − .69]
	M	.75	[.72 − .77]	.82	[.78 − .86]	.80	[.79 − .82]
2011	C_1	.94	[.93 − .94]	.95	[.95 − .96]	.94	[.94 − .94]
	C_2	.95	[.94 − .95]	.96	[.96 − .97]	.95	[.94 − .95]
			DIC = 8553.17		DIC = 8590.94		$G^2(4) = 43.81$

Note. 2005 Exp 1 = R. E. Smith and Bayen (2005), Experiment 1; 2005 Exp 2 = R. E. Smith and Bayen (2005), Experiment 2; 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, or that a sentence is true in sentence verification; C_2 = probability to detect that a letter string is a non-word or that a sentence is false in sentence verification; BCI = Bayesian confidence interval; CI = (traditional) confidence interval. Adapted from "Hierarchical multinomial modeling approaches: An application to prospective memory and working memory," by N. R. Arnold, U. J. Bayen, and R. E. Smith, 2014, *manuscript submitted for publication.*

4.4 Overall Discussion

I presented three studies that demonstrated the usefulness of hierarchical MPT modeling with two different MPT models. With the beta-MPT version of the 2HTSM we were able to support the probability-matching account on an individual level by showing that individual differences in contingency perception influenced guessing behavior. Thereby we were able to contribute to a theoretical debate.

The other two studies examined individual differences in PM. It is very important to distinguish between the prospective and the retrospective component of PM. This has for example practical implications for clinical interventions. Due to the special features of the paradigm, even when participants undergo many trials, some categories are only sparsely filled. Thus, even with many data points, it is problematic to calculate individual parameter estimates. The first PM study deals with the relationship between PM and mental health. Previous findings are mixed and do not include separate measures for prospective and retrospective components. In this study, we showed that, at least with a non-clinical sample, depression is not related to event-based PM even if a resource-demanding task is used. We found no relationship with trait anxiety, but state anxiety was related to PM performance. This relationship was due to a negative correlation with the prospective component. This finding underlines the importance of distinguishing between state and trait anxiety as well as between the prospective and the retrospective component of PM. State anxiety has been shown to limit WM capacity.

The relationship between WM span and PM was the subject of Study 3. Like state anxiety, WM span was related to the prospective component P but not to the retrospective component M. Additionally, Study 3 involved a comparison between the beta-MPT approach and the latent-trait approach. We did not find substantial differences between the approaches. However, the latent-trait approach incorporates parameter correlations a priori, whereas the beta-

MPT approach assumes independent beta distributions for the parameters. This assumption is not very realistic (Erdfelder, 2000; Klauer, 2010; Matzke et al. 2013). However, it is possible to calculate posterior correlations for both approaches.

Another approach that has been proposed is the latent-class approach (Klauer, 2006). Participants are assumed to fall into a finite number of latent classes. Participants in the same class are assumed to have the same parameter values. This assumption only holds approximately for very few applications. In most cases, a continuous distribution is more realistic. Navarro, Griffiths, Steyvers, and Lee (2006) proposed Dirichlet-process modeling to incorporate heterogeneity for an infinite groups model. However, different kinds of hierarchical structures are possible. The Dirichlet distribution is a multivariate version of the beta distribution and would therefore be suitable as continuous hierarchical distribution as well.

All of the approaches presented here deal with either participant or item heterogeneity but not both. Matzke et al. (2013) presented the crossed-random effects approach that assumes heterogeneity in both participants and items. However, this approach is only applicable to very few MPT modeling paradigms because all items have to appear in all subtrees. This is not always possible.

MPT models can be used to solve the decomposition problem by explicitly modeling cognitive processes that are not related to the construct of interest. The extension to hierarchical MPT models has opened the door for new applications like cognitive psychometrics. Of course, not all problems in cognitive psychology or memory research can be solved by (hierarchical) MPT modeling. Observations for MPT models have to be categorical. Indeed, all continuous data can be transformed to categorical data and this is done for example in the hindsight bias MPT model (Erdfelder & Buchner, 1998). Still, this results in a loss of information. Another example for important continuous information is reaction time data. For reaction time data, other models like diffusion models (Ratcliff, 1978) or other random walk based models (e.g., Luce 1986) are to be preferred.

Multinomial models have to be carefully validated. The meaning that is assigned to the parameters cannot be controlled statistically. In most cases, there is another statistically equivalent MPT model that matches the data equally well. However, this can also be seen as an advantage of MPT models because it forces researchers to make there assumptions explicit and test them. Last but not least the quality of a study does not only depend on the MPT model but also on the quality of the experiment and the data to which the MPT model is applied (Erdfelder, 2000).

Taken together, MPT models are very useful and reliable tools in cognitive psychology. Incorporating individual differences makes them applicable to even more research questions. Unfortunately, user-friendly programs to conduct hierarchical MPT modeling do not exist yet. For traditional MPT modeling, the implementation of modeling programs has greatly improved the applicableness of MPT models for researchers. The only program that includes a hierarchical MPT model is HMMTree (Stahl & Klauer, 2007) which can model the latent-trait approach. User-friendly programs for continuous hierarchical MPT modeling are still missing. There exist WinBUGS implementations (Matzke et al., 2013; J. B. Smith & Batchelder, 2010) but this still requires a certain amount of programming knowledge in WinBUGS and other programs like R or Matlab.

References

Agresti, A., Caffo, B., & Ohman-Strickland, P. (2004). Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies. *Computational Statistics & Data Analysis, 47,* 639–653.

Altgassen, M., Kliegel, M., & Martin, M. (2009). Event-based prospective memory in depression: The impact of cue focality. *Cognition and Emotion, 23*, 1041–1055.

American Psychological Association (2010*). Publication manual of the American Psychological Association (6th ed.).* Washington, DC: American Psychological Association.

Baddeley, A., Eysenck, M. W., & Anderson, M. C. (Eds.) (2009). *Memory.* Hove, UK and New York: Psychology Press.

Ball, B. H., Knight, J. B., Dewitt, M. R., & Brewer, G. A. (in press). Individual differences in the delayed execution of prospective memories. *The Quarterly Journal of Experimental Psychology, 66,* 2411-2425.

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

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

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

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

Bayen, U. J., & Kuhlmann, B. G. (2011). Influences of source–item contingency and schematic knowledge on source monitoring: Tests of the probability-matching account. *Journal of Memory and Language,* 64, 1–17.

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 197–215.

Bayen, U. J., Nakamura, G. V., Dupuis, S. E., & Yang, C.-L. (2000). The use of schematic knowledge about sources in source monitoring. *Memory & Cognition, 28,* 480–500.

Beato, M. S., Pulido, R. F., Pinho, M. S., & Gozalo, M. (2013). Reconocimiento falso y ansiedad estado/rasgo [False recognition and state/trait anxiety]. *Psicológica*, *34*, 299–311.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II.* San Antonio, TX: Psychological Corporation.

Brand, A. N., Jolles, J., & Gispen-de Wied, C. (1992). Recall and recognition memory deficits in depression. *Journal of Affective Disorders, 25,* 77–86.

Brewer, G. A., Knight, J. B., Marsh, R. L., & Unsworth, N. (2010). Individual differences in event-based prospective memory: Evidence for multiple processes supporting cue detection. *Memory & Cognition, 38*, 304–311.

Buchner, A., Erdfelder, E., & Vaterrodt-Plünnecke, B. (1995). Toward unbiased measurement of conscious and unconscious memory processes within the process dissociation framework. *Journal of Experimental Psychology: General, 124,* 137–160.

Cherry, K. E., & LeCompte, D. C. (1999). Age and individual differences influence prospective memory. *Psychology and Aging, 14,* 60–76.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302.

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review, 70,* 193–242.

Efron, B., & Tibshirani, R. J. (1997). *An introduction to the bootstrap.* New York: Chapman & Hall.

Ehrenberg, K., & Klauer, K. C. (2005). Flexible use of source information: Processing components of the inconsistency effect in person memory. *Journal of Experimental Social Psychology, 41,* 369–387.

Einstein, G. O., & McDaniel, M. A. (1990). Normal aging and prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 717–726.

Einstein, G. O., McDaniel, M. A., Manzi, M., Cochran, B., & Baker, M. (2000). Prospective memory and aging: Forgetting intentions over short delays. *Psychology and Aging, 15,* 671–683.

Ellis, H.C. & Ashbrook, P.W. (1988). Resource allocation model of the effects of depressed mood states on memory. In K. Fiedler & J. Forgas (eds.), *Affect, cognition and social behavior* (pp.25–42). Toronto: Hogrefe.

Erdfelder, E. (2000). *Multinomiale Modelle in der kognitiven Psychologie [Multinomial models in cognitive psychology].* Unpublished habilitation thesis, Psychologisches Institut der Universität Bonn, Germany.

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Journal of Psychology, 217*, 108–124.

Erdfelder, E., & Bredenkamp, J. (1998). Recognition of script-typical versus script-atypical information: Effects of cognitive elaboration. *Memory & Cognition, 26,* 922–938.

Erdfelder, E. & Buchner, A. (1998). Decomposing the hindsight bias: A multinomial processing tree model for separating recollection and reconstruction in hindsight. *Journal of Experimental Psychology - Learning, Memory, and Cognition, 24,* 387–413.

Estes, W. K., & Straughan, J. H. (1954). Analysis of a verbal conditioning situation in terms of statistical learning theory. Journal of Experimental Psychology, 47, 225–234.

Eysenck, M. W. (1985). Anxiety and cognitive-task performance. *Personality and Individual Differences, 6*, 579–586.

García-Pérez, M. A. (1994). Parameter estimation and goodness-of-fit testing in multinomial models. *British Journal of Mathematical & Statistical Psychology, 47,* 247–282.

Gelman, A., & Hill, J. (2007). Data analysis using regression and multilevel/hierarchical models. Cambridge: Cambridge University Press.

Gelman, A., & Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7,* 457–472.

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

Gilks, W. R., Richardson, S. E., & Spiegelhalter, D. J. (1996). *Markov chain Monte Carlo in practice.* London: Chapman & Hall.

Hautzinger, M., Keller, F., & Kühner, C. (2006). *BDI II: Beck Depressions-Inventar - Manual* [BDI II: Beck Depression-Inventory – Manual]. Frankfurt am Main: Harcourt Test Services.

Hays, W. L. (1994). *Statistics* (5th ed.). Fort Worth, TX: Harcourt Brace College Publishers.

Herrmann, C., Buss, U., & Snaith, R. P. (1995). *HADS-D: Hospital Anxiety and Depression Scale – Deutsche Version – Ein Fragebogen zur Erfassung von Angst und Depressivität in* *der somatischen Medizin* [HADS-D: Hospital Anxiety and Depression Scale – German Version – A Questionnaire Measuring Anxiety and Depression in Psychosomatic Medicine]. Bern: Verlag Hans Huber.

Hertel, P. T., & Milan, S. (1994). Depressive deficits in recognition: Dissociation of recollection and familiarity. *Journal of Abnormal Psychology, 103*, 736–742.

Horn, S. S., Bayen, U. J., Smith, R. E., & Boywitt, C. D. (2011). The multinomial model of prospective memory: Validity of ongoing-task parameters. *Experimental Psychology*, *58*, 247–255.

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

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

Jeffreys, H. (1961). *Theory of probability.* Oxford: UK Oxford University Press.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3–28.

Keefe, R. S. E., Arnold, M. C., Bayen, U. J., McEvoy, J. P., & Wilson, W. H. (2002). Source-monitoring deficits for self-generated stimuli in schizophrenia: Multinomial modeling of data from three sources. *Schizophrenia Research, 57,* 51–67.

Klauer, K. C. (2006). Hierarchical multinomial processing tree models: A latent-class approach. *Psychometrika, 71*, 7–31.

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika. 75,* 70–98.

Kliegel, M., & Jäger, T. (2006). The influence of negative emotions on prospective memory: A review and new data. *International Journal of Computational Cognition, 4,* 1-17.

Knapp, B. R., & Batchelder, W. H. (2004). Representing parametric order constraints in multi-trial applications of multinomial processing tree models. *Journal of Mathematical Psychology*, *15*, 215–229.

Kuhlmann, B. G., Vaterrodt, B., & Bayen, U. J. (2012). Schema-bias in source monitoring varies with encoding conditions: Support for a probability-matching account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 1365–1376.

Laux, L., Glanzmann, P., Schaffner, P., & Spielberger, C. D. (1981). *STAI: Das State-Trait Angstinventar* [STAI: The State-Trait Anxiety Inventory]. Weinheim: Beltz Testgesellschaft.

Lee, M. D., & Wagenmakers, E. J. (2005). Bayesian statistical inference in psychology: Comment on Trafimow (2003). *Psychological Review, 112,* 662–668.

Luce, R. D. (1986). *Response times. Their role in infering elementary mental organization.* New York: Oxford University Press.

Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The BUGS project: Evolution, critique and future directions. *Statistics in Medicine, 28,* 3049–3067.

MacLeod, C., & Donnellan, A. M. (1993). Individual differences in anxiety and the restriction of working memory capacity. *Personality and Individual Differences, 15*, 163–173.

Matzke, D., Dolan, C. V, Batchelder, W. H., & Wagenmakers, E.-J. (in press). Bayesian estimation of multinomial processing tree models with heterogeneity in participants and items. *Psychometrika: Application Reviews & Case Studies.*

McDaniel, M. A., & Einstein, G. O. (2000). Strategic and automatic processes in prospective memory retrieval: A multiprocess framework. *Applied Cognitive Psychology, 14,* 127–144.

McDaniel, M. A., & Einstein, G. O. (2007). *Prospective memory: An overview and synthesis of an emerging field.* Thousand Oaks, CA: Sage.

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods, 42,* 42–54.

Navarro, D., Griffiths, T., Steyvers, M., & Lee, M. D. (2006). Modeling individual differences using Dirichlet processes. *Journal of Mathematical Psychology, 50,* 101–122.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling [Computer software manual]. Retrieved from http://citeseer.ist.psu.edu/plummer03jags.html

Purdy, B., & Batchelder, W. H. (2009). A context–free language for binary multinomial processing tree models. *Journal of Mathematical Psychology, 53,* 547–561.

Ramponi, C., Murphy, F. C., Calder, A. J., & Barnard, P. J. (2010). Recognition memory for pictorial material in subclinical depression. *Acta Psychologica, 135,* 293–301.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108.

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

Reese, C. M., & Cherry, K. E. (2002). The effects of age, ability, and memory monitoring on prospective memory task performance. *Aging, Neuropsychology, and Cognition, 9,* 98–113.

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

Riefer, D. M., Hu, X., & Batchelder, W.H. (1994). Response strategies in source monitoring. *Journal of Experimental Psychology Learning, Memory, and Cognition, 20,* 680–693.

Rothkegel, R. (1999). AppleTree: A multinomial processing tree modeling program for Macintosh computers. *Behavior Research Methods, Instruments, & Computers, 31,* 696–700.

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

Rouder, J., Lu, J., Morey, R., Sun, D., & Speckman, P. (2008). A hierarchical process–dissociation model. *Journal of Experimental Psychology: General, 137,* 370–389.

Rouder, J., Lu, J., Sun, D., Speckman, P., Morey, R., & Naveh-Benjamin, M. (2007). Signal detection models with random participant and item effects. *Psychometrika, 72,* 621–642.

Shewchuk, J. R. (1994).An introduction to the conjugate gradient method without the agonizing pain, Technical Report CS-94-125, Carnegie Mellon University, Pittsburgh, PA, USA.

Singmann, H., & Kellen, D. (2013). MPTinR: Analysis of Multinomial Processing Tree Models in R. *Behavior Research Methods, 45,* 560–575.

Smith, J. B., & Batchelder, W.H. (2008). Assessing individual differences in categorical data. *Psychonomic Bulletin & Review, 15,* 713–731.

Smith, J. B., & Batchelder, W. H. (2010). Beta-MPT: Multinomial processing tree models for addressing individual differences. *Journal of Mathematical Psychology, 54*, 167–183.

Smith, R. E. (2003). 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, 29*, 347–361.

Smith, R. E., & Bayen, U. J. (2004). A multinomial model of event-based prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 756–777.

Smith, R. E., & Bayen, U. J. (2005). The effects of working memory resource availability on prospective memory: A formal modeling approach. *Experimental Psychology, 52,* 243–256.

Smith, R. E., Persyn, D., & Butler, P. (2011). Prospective memory, personality, and working memory: A formal modeling approach. *Zeitschrift für Psychologie / Journal of Psychology, 219,* 108–116.

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

Spaniol, J., & Bayen, U. J. (2002). When is schematic knowledge used in source monitoring? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 631–651.

Spiegelhalter, D. J., Best, N. G., Carlin, B. R., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B, Statistical Methodology, 64,* 583–616.

Spiegelhalter, D., Thomas, A., Best, N., Gilks, W., & Lunn, D. (2003). BUGS: Bayesian inference using Gibbs sampling [Computer software manual]. Retrieved from http://www.mrc-bsu.cam.ac.uk/bugs/

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *Manual for the State-Trait Anxiety Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Stahl, C., & Klauer, K. (2007). HMMTree: A computer program for latent–class hierarchical multinomial processing tree models. *Behavior Research Methods, 39,* 267–273.

Stout, D. M., & Rokke, P. D. (2010). Components of working memory predict symptoms of distress. *Cognition and Emotion, 24,* 1293–1303.

Sturtz, S., Ligges, U., & Gelman, A. (2005). R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software, 12*, 1–16.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review, 14,* 779–804.

Watts, F. N., Morris, L., & MacLeod, A. K. (1987). Recognition memory in depression. *Journal of Abnormal Psychology, 96*, 273–275.

West, R., & Craik, F. I. M. (2001). Influences on the efficiency of prospective memory in younger and older adults. *Psychology and Aging, 16,* 682-696.

Eidesstattliche Versicherung

Ich versichere an Eides Statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist.

Düsseldorf, den 13.3.2014 Unterschrift: ______________________________

Appendix

Study 1 was published in

Arnold, N. R., Bayen, U. J., Kuhlmann, B. G., & Vaterrodt, B. (2013). Hierarchical modeling of contingency-based source monitoring: A test of the probability-matching account. *Psychonomic Bulletin & Review, 20,* 326–333.

I planned the study based on the design by Beatrice G. Kuhlmann, Bianca Vaterrodt, & Ute J. Bayen (2012). I supervised the data collection. I developed routines for the beta-MPT analysis with the 2HTSM and performed the analyses. I prepared, submitted and revised the manuscript with the support of the co-authors.

Study 2 is submitted as:

Arnold, N. R., Bayen, U. J., & Böhm, M. F. (2014). Is prospective memory related to depression and anxiety? *Manuscript submitted for publication.*

I planned the study and adapted the experimental design. I supervised the data collection. Data were collected by Master student Pia Ewerdwalbesloh and Bachelor student Mateja F. Böhm. I developed routines for the beta-MPT analysis with the MPT model of event-based PM and performed the analyses. I prepared and submitted the manuscript with the support of the co-authors.

Study 3 is submitted as:

Arnold, N. R., Bayen, U. J., & Smith, R. E. (2014). Hierarchical multinomial modeling approaches: An application to prospective memory and working memory. *Manuscript submitted for publication.*

The study contains reanalyses of experiments designed and published by Rebekah E. Smith and Ute J. Bayen (2005) as well as by Rebekah E. Smith, Deborah Persyn, and Patrick Butler (2011). I adapted the routine for the beta-MPT analysis with the MPT model of event-based PM and developed the routine for the latent-trait analysis with the MPT model of event-based PM and performed the analyses. I prepared and submitted the manuscript with the support of the co-authors. The manuscript is currently under revision for resubmission to *Experimental Psychology*.

Psychon Bull Rev (2013) 20:326–333
DOI 10.3758/s13423-012-0342-7

BRIEF REPORT

Hierarchical modeling of contingency-based source monitoring: A test of the probability-matching account

Nina R. Arnold · Ute J. Bayen · Beatrice G. Kuhlmann · Bianca Vaterrodt

Published online: 8 December 2012

Abstract According to the probability-matching account of source guessing (Spaniol & Bayen, *Journal of Experimental Psychology: Learning, Memory, and Cognition 28*:631–651, 2002), when people do not remember the source of an item in a source-monitoring task, they match the source-guessing probabilities to the perceived contingencies between sources and item types. In a source-monitoring experiment, half of the items presented by each of two sources were consistent with schematic expectations about this source, whereas the other half of the items were consistent with schematic expectations about the other source. Participants' source schemas were activated either at the time of encoding or just before the source-monitoring test. After test, the participants judged the contingency of the item type and source. Individual parameter estimates of source guessing were obtained via beta-multinomial processing tree modeling (beta-MPT; Smith & Batchelder, *Journal of Mathematical Psychology 54*:167–183, 2010). We found a significant correlation between the perceived contingency and source guessing, as well as a correlation between the deviation of the guessing bias from the true contingency and source memory when participants did not receive the schema information until retrieval. These findings support the probability-matching account.

Keywords Source monitoring · Schemas · Mathematical modeling · Multinomial processing tree models

Source monitoring involves judgments regarding the origin of information (Johnson, Hashtroudi, & Lindsay, 1993). In typical source-monitoring tasks, participants are presented with items from two or more sources and are later required to judge whether the items were presented by one of the sources, and if so, which one.

How we interpret and use information is influenced by the source that we believe gave the information. For example, you trust doctors more than your hairdressers for advice on medicine but not on haircuts. According to Johnson's source-monitoring framework (Johnson et al., 1993), two types of information are used to attribute memories to sources, namely (1) episodic memory for features of the source and (2) general knowledge, plausibility, and beliefs. Either you may remember being in a hair salon when you heard the advice, or you may rely on your general knowledge. Specifically, you know that the probability of talking to your hairdresser about your hair style is much greater than the probability of talking to your doctor about this. Thus, there are certain expected contingencies of types of information and their sources.

Contingency knowledge may stem from actual experiences with the sources (e.g., Bridget has always been helpful) or from general schematic knowledge about them (e.g., Bridget is a girl scout, and one thus infers that she must be helpful). According to the *probability-matching account of source guessing* (Spaniol & Bayen, 2002), participants match learned contingencies about particular sources whenever possible,

Electronic supplementary material The online version of this article (doi:10.3758/s13423-012-0342-7) contains supplementary material, which is available to authorized users.

N. R. Arnold · U. J. Bayen (✉)
Mathematical and Cognitive Psychology, Institute of Experimental Psychology, Heinrich-Heine-Universität Düsseldorf, Universitätsstr. 1, Geb. 23.02, 40225 Düsseldorf, Germany
e-mail: ubayen@hhu.de

N. R. Arnold
e-mail: nina.arnold@hhu.de

B. G. Kuhlmann
Department of Psychology, University of North Carolina at Greensboro, Greensboro, NC, USA

B. Vaterrodt
Department of Psychology, Humboldt-Universität, Berlin, Germany

relying on more general schematic expectations only if the participants do not have a contingency representation.

Probability matching has been observed in source monitoring (e.g., Bayen & Kuhlmann, 2011; Erdfelder & Bredenkamp, 1998) and in other tasks, such as old–new recognition (e.g., Buchner, Erdfelder, & Vaterrodt-Plünnecke, 1995; Ehrenberg & Klauer, 2005) and human choice behavior (e.g., Estes & Straughan, 1954). Furthermore, studies have shown that prior knowledge influences source guessing (Bayen, Nakamura, Dupuis, & Yang, 2000; Ehrenberg & Klauer, 2005; Spaniol & Bayen, 2002). Thus, there is support for the idea that source guessing can be based on learned contingencies about specific sources or on more general prior knowledge about sources, as suggested by the probability-matching account.

Hicks and Cockmann (2003) found that the time when the schema-relevant information was given to participants affected the source-guessing bias. Participants who received the information after encoding showed schema-consistent bias in their source attributions, whereas participants who had already received the information before encoding showed no such bias. However, Bayen and Kuhlmann (2011) found that source guessing in this schema-before-encoding condition was only unbiased in a full-attention condition (in contrast to a divided-attention condition) at encoding. With divided attention at encoding, schema bias occurred. This suggests that when participants can process the true contingency between sources and items, they will rely on this information rather than on prior schematic knowledge, and this supports the probability-matching account. In line with this idea, the same authors manipulated the actual contingencies between item types and sources and found that guessing matched the experimental contingencies (under full attention at encoding and when participants knew about the schema-relevant information at the time of encoding).

Importantly, the probability-matching account predicts a positive correlation between contingency perception and source-guessing bias, such that people differing in contingency perception within the same experimental setting should also differ in source-guessing bias. In other words, individual differences in contingency perception should be related to variations in source-guessing bias. For example, Spaniol and Bayen (2002) found that in the same source-monitoring task, some participants relied on schematic knowledge in source guessing, while others did not. To reconcile these differential source-guessing patterns, the authors suggested that these participants differed in their perceived contingencies. However, to date, the relationship between individual contingency perception and source-guessing bias has not been investigated.

The purpose of the present study was to demonstrate that individual differences in contingency perception relate to individual differences in source-guessing bias. We used a new methodological approach, beta-MPT modeling (Smith & Batchelder, 2010), which allowed us to estimate individual participants' source-guessing probabilities. Bayen and Kuhlmann (2011; see also Bayen et al., 2000) used a multinomial processing tree (MPT) model, the two-high-threshold model of source monitoring (2HTSM; Bayen, Murnane, & Erdfelder, 1996), to disentangle memory and guessing in the source-monitoring paradigm. First we will describe the 2HTSM, then the basics of the beta-MPT approach. We then report an experiment to test the probability-matching account on an individual-differences level by investigating the relationship between perceived contingencies and guessing in source monitoring, using the beta-MPT approach.

The 2HTSM (Bayen et al., 1996) is a stochastic model that separates memory and guessing in source monitoring. We used Submodel 4 (see Bayen et al., 1996, for details), which assumes that the levels of item memory as well as source memory are equal for both sources, and which had fit the data from previous studies using the same standard source-monitoring paradigm we used in the present study (Bayen & Kuhlmann, 2011; Bayen et al., 2000; Kuhlmann, Vaterrodt, & Bayen, 2012). The model (see Fig. 1) assumes a source-monitoring task with two sources. Statements are presented either by the schematically expected source (e.g., the doctor presenting an expected-doctor statement) or by the schematically unexpected source (e.g., the lawyer presenting an expected-doctor statement). The first and second trees represent the cognitive processes involved in responses for items that originated from the schematically expected and the schematically unexpected sources, respectively. The third tree represents processes for unstudied distractor items (i.e., new items).

With probability D, participants correctly recognize an item as old or new. With probability d, they remember the source of the item. If they cannot remember the source (with probability $1 - d$), they must guess. With probability g, they guess that the item is from the source that is consistent with the schematic expectation; with probability $1 - g$, they guess that the item is from the schematically unexpected source. If participants do not remember whether an item is old or new (probability $1 - D$), they guess, with probability b, that the item is old or, with probability $1 - b$, that it is new. If they have guessed that an item is old, they must guess the source of the item. With probability g, the guess is the schematically expected source, and with probability $1 - g$, the schematically unexpected source.

Traditionally, data are aggregated over items and participants for MPT analysis, so that there is one set of parameters for all participants. Thereby, homogeneity is assumed for items and participants; that is, the data from different items and participants are assumed to be independent and identically distributed. However, this assumption is often violated and may lead to biased parameter estimates (Klauer, 2006, 2010; Smith & Batchelder, 2008, 2010). Furthermore, this approach only yields group-level estimates for parameters, not individual estimates.

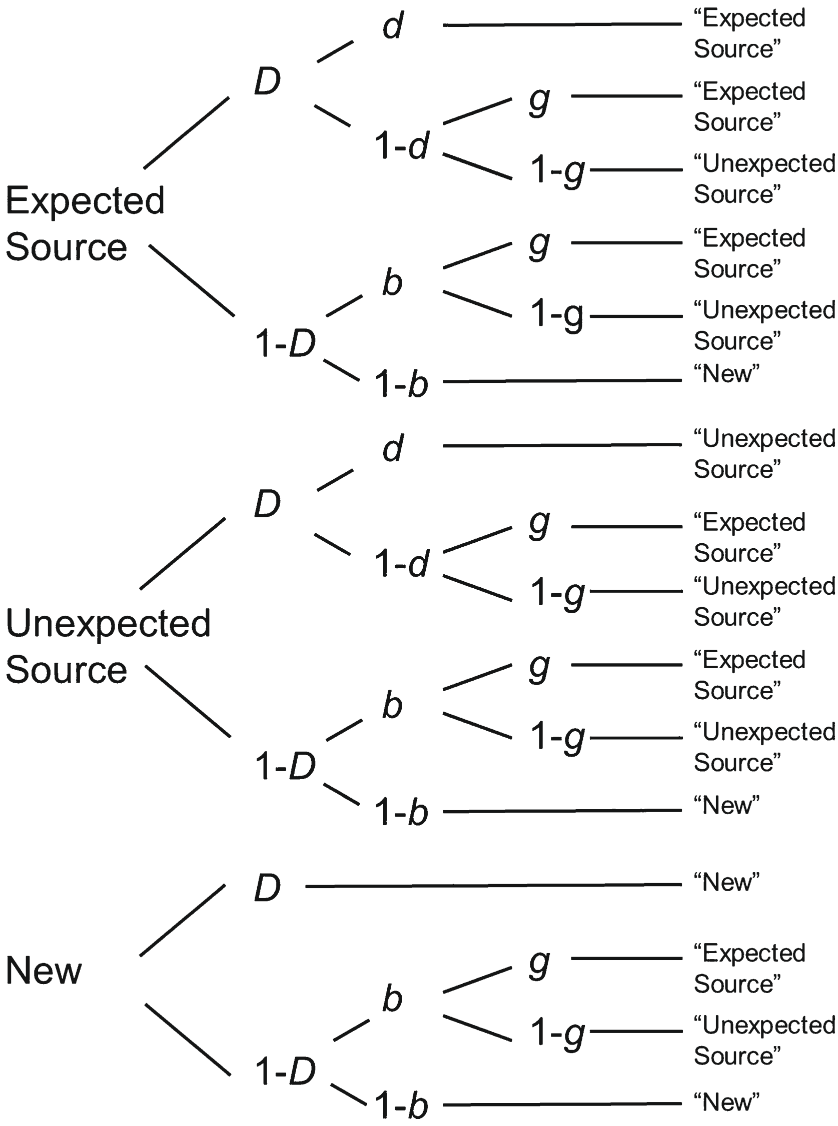


Fig. 1 Submodel 4 of the two-high-threshold model of source monitoring. *D* = probability of detecting that an item is old/new; *d* = probability of correctly remembering the source of an item; *g* = probability of guessing that an item is from the expected source; *b* = probability of guessing that an item is old. Adapted from "Source Discrimination, Item Detection, and Multinomial Models of Source Monitoring," by U. J. Bayen, K. Murnane, and E. Erdfelder, 1996, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, p. 202. Copyright 1996 by the American Psychological Association

Recently, hierarchical models have been developed to deal with heterogeneity (Klauer, 2006, 2010; Smith & Batchelder, 2010). We used the beta-MPT approach (Smith & Batchelder, 2010). The advantage of this method is that it uses a hierarchical distribution for each parameter that lies within the interval (0, 1), and thus has the same scale as the MPT model parameters, which indicate probabilities. The method assumes that participants' parameters are drawn independently from beta distributions for each model parameter. The beta distribution is a very flexible distribution that can also approximate the normal distribution (between 0 and 1).

The main purpose of our experiment was to test the core assumption of the probability-matching account: namely, that people guess according to the perceived contingency if they do not remember the source. Therefore, we should find a positive relationship between participants' perceived contingencies and the source-guessing parameter *g*. Bayen and Kuhlmann (2011) only demonstrated this relationship at a group level. However, within the same experimental setting, individual variations in source-guessing bias are possible (cf. Spaniol & Bayen, 2002), requiring an individual-differences approach. With the beta-MPT approach, it is possible to link the guessing parameter directly to participants' perceived contingencies. Along with the replication of previous results (i.e., that the guessing parameter *g* was larger in the retrieval than in the encoding condition), this is a crucial test of the probability-matching account, since it associates the guessing parameter directly with the perceived contingency. If guessing bias were unrelated to perceived contingency, the probability-matching account would be falsified. We additionally tested the hypothesis that the source-guessing biases of participants with good source memory would be closer to the actual contingency than would the source-guessing biases of participants with poor source memory, because participants with good source memory would be more likely to realize the actual contingency (cf. Spaniol & Bayen, 2002). Thus, there should be a negative correlation between the source memory parameter d and the difference between the perceived and the real contingencies.

In our experiment, participants in the *encoding* condition were told about the professions of the two sources before encoding. In contrast, participants in the *retrieval* condition did not know about the professions until the test phase. In both conditions, schematically expected statements were presented with equal probabilities by the expected and by the unexpected source. Previous studies with traditional MPT analyses on aggregated data had shown that the guessing bias is near the true contingency if the schematically relevant information is available during encoding (e.g., Kuhlmann et al., 2012). In this case, participants notice the contingencies during the encoding phase and later adjust their guessing accordingly. If, however, participants have difficulties accessing the true contingency (because the schema information was not available during encoding), source guessing is biased toward the schematically expected source (Kuhlmann et al., 2012). Thus, we wanted to replicate previous findings, namely that the guessing parameter g would equal .5 (i.e., reflect the true source–item contingency) in the encoding condition, and be larger than .5 (i.e., biased toward the schematically expected source) in the retrieval condition. Our main objective, however, was to test the probability-matching account more stringently with the new beta-MPT method. We hypothesized a positive correlation between guessing parameter g and the perceived contingency. Also, we expected to find a negative correlation between source memory parameter d and the deviation of the perceived contingency from the true contingency of .5.

Method

Participants

The participants were 48 native German speakers (41 students, 7 employed). The mean age was 22.6 years (range 18 to 32). All participants received €5.

Design

We used a 3 × 2 × 2 mixed factorial design, with Expectancy of Statements (expected-doctor statements, expected-lawyer statements, and equally expected filler statements) and Source of Statement (doctor vs. lawyer) as within-subjects factors, and Time of Schema Activation (before encoding vs. before retrieval) as a between-subjects factor. The participants were randomly and equally assigned to the two conditions.

Materials

The design relied on well-established doctor and lawyer schemas. We used the German version of Bayen et al.'s (2000) doctor–lawyer materials, as developed by Kuhlmann et al. (2012), who normed the sentences with expectancy ratings from 60 native German speakers. The materials consisted of 96 statement pairs, of which 32 were expected for a doctor, 32 were expected for a lawyer, and 32 were filler statement pairs that were equally expected for both. The two members of each pair differed in one word or phrase that changes the meaning, and they were randomly assigned as the target and distractor in our memory test.

Procedure

The participants were tested in groups of up to four in individual computer booths. Computerized instructions informed them that they would see the faces of Ralf and Uwe (German male names) accompanied by statements. In the encoding condition, they were also told that, for example, Ralf was a doctor and Uwe was a lawyer. The assignment of names to sources was counterbalanced. Participants were informed that they would have to recognize the statements later. There was no mention of the upcoming source memory test. The 96 statements were presented for 6 s each, in random order, above the face of the source "speaking." Participants also saw the name (e.g., "RALF") and, in the encoding condition, the profession of the source (e.g., "RALF = DOCTOR"). Four equally expected statements served as a primacy buffer. The statements were randomly assigned to the sources, with equal numbers of expected and unexpected statements (i.e., statements that were expected for the other source) being assigned to each source. Thus, there was a zero contingency between the source and the expectedness of the statement.

The instructions for the self-paced source-monitoring test started immediately after study. The participants had to judge whether each test statement had been said by Ralf, by Uwe, or by neither. At this time, participants in the retrieval condition were given the sources' professions. At test, the pictures of the sources were shown side by side on the screen, along with the names and professions (e.g., "RALF = DOCTOR"). The third option, "NEITHER," appeared centered without a picture. The 96 (32 expected-doctor, 32 expected-lawyer, and 32 equally expected filler) sentences were presented in a random order centered at the top of the screen, preceded by "Who said:". For each source, a random half of the statements of each type were tested in their study version, whereas for the remaining statements, the distractor version was used.

Assignment of the "D" and "K" keys to the sources (doctor, lawyer) was counterbalanced. The participants pressed the space bar when they believed that a statement was new, and error feedback was not provided.

After the memory test, the participants gave contingency judgments by answering two questions in counterbalanced order: namely (in translation), "How many of the 32 expected-doctor statements were said by RALF = doctor?" and "How many of the 32 expected-lawyer statements were said by UWE = lawyer?" Finally, they completed a demographic questionnaire and were debriefed and paid.

Results

For ease of presentation, we grouped the statements into two types, namely *schematically expected* (those presented by their expected source) and *schematically unexpected* (those presented by the unexpected source) statements. Separate analyses for the expected-doctor and expected-lawyer statements revealed the same pattern of results (see Online Supplement 1 for the raw data and Supplement 2 for the parameter estimates). We conducted traditional MPT analyses (with the data aggregated over items and participants) with the multiTree program (Moshagen, 2010) and hierarchical modeling with beta-MPT (Smith & Batchelder, 2010). We used an alpha level of .05 for all significance tests.

Traditional MPT analysis on aggregated data

We estimated separate models for the encoding and retrieval conditions on the basis of the aggregated data presented in Table 1. The parameter estimates and confidence intervals are in Table 2. We tested goodness of fit with the log-likelihood statistic G^2, which is asymptotically chi-square distributed. The four-parameter version of the model fit the data in both conditions, $G^2_{enc}(2) = 2.37$, $p = .31$, and $G^2_{ret}(2) = 2.00$, $p = .37$. As expected, in the retrieval condition, the guessing parameter g was significantly larger than .5, $G^2(1) = 223.55$, $p < .01$. In the encoding condition, g was also significantly larger than .5, $G^2(1) = 18.81$, $p < .01$. Thus, contrary to expectations for this condition, we did not find that participants guessed according to the true contingency, but instead were biased toward the schematically expected source. However, participants in the retrieval condition showed significantly larger guessing biases than did participants in the encoding condition, $G^2(1) = 97.05$, $p < .01$. Thus, as predicted by the probability-matching account, participants in the retrieval condition were more likely to guess according to the schematically expected source, whereas participants in the encoding condition appeared to have adjusted their guessing bias toward the true zero contingency. The raw data for the equally expected statements are in Online Supplement 1, and the parameter estimates are in Supplement 2. For these control statements, guessing parameter g did not differ significantly from .5 in either condition, as expected: $G^2_{enc}(1) = 0.75$, $G^2_{ret}(1) = 0.03$.

Table 1 Response category frequencies for schematically expected statements in the two experimental conditions of the experiment

	Encoding			Retrieval		
Source	"E"	"U"	"N"	"E"	"U"	"N"
Expected	168	67	149	210	47	127
Unexpected	90	124	170	170	83	131
Neither	154	104	510	193	67	508

Encoding = schema information was given at encoding. Retrieval = schema information was not given until retrieval. "E" = "expected source" response, "U" = "unexpected source" response, "N" = "new" response

Analysis with beta-MPT

We used the basic version of the Markov chain Monte Carlo (MCMC) algorithm provided by Smith and Batchelder (2010) for the pair-clustering model and adjusted it to the 2HTSM. We used WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000) to run the MCMC method. At convergence, the potential scale reduction factor $Rhat = 1$. Again, we estimated separate models for the encoding and retrieval conditions. Each algorithm was run with 100,000 iterations, with the first half removed as a burn-in period. For all parameter estimates, $Rhat = 1$, except for α, β, and the variance of the item memory parameter D in the encoding condition, where $Rhat = 1.5$. Table 3 shows the posterior distributions of the parameters of the hierarchical beta distributions. Credible intervals of the standard deviations for the parameters did not include zero for any of the parameters. This means that parameter homogeneity (i.e., $SD = 0$) was very unlikely. Thus, there is strong evidence for heterogeneity, especially for guessing parameter g. As is shown in Table 2, the general patterns of the results are similar for the standard aggregated analysis with multiTree and the group parameters from the beta-2HTSM analyses. The 95 % confidence intervals and the credible intervals (Bayesian confidence intervals) overlapped for all parameter estimates.

We transformed participants' absolute contingency judgments to relative contingency judgments. The contingency judgments for expected-doctor ($M_{enc} = .57$, $SD_{enc} = .10$; $M_{ret} = .60$, $SD_{ret} = .19$) and expected-lawyer ($M_{enc} = .58$, $SD_{enc} = .10$; $M_{ret} = .62$, $SD_{ret} = .19$) statements did not differ significantly for either experimental group, both ps > .40. In the encoding condition, the mean contingency judgment was $M = .57$, $SD = .08$. In the retrieval condition, the mean contingency judgment was $M = .61$, $SD = .17$. The correlations between perceived contingency and the source-guessing bias g were significant, with $r = .45$, $p = .02$ (see Fig. 2a, all correlations one-tailed), in the encoding condition and $r = .55$, $p < .01$ (see Fig. 2b), in the retrieval condition. This means that the higher that the contingency of items and their expected sources was perceived, the higher was the probability that the

Table 2 Group parameter estimates for schematically expected items with both traditional MPT and beta-MPT

	Encoding Condition				Retrieval Condition			
	Traditional MPT		Beta-MPT		Traditional MPT		Beta-MPT	
	M (*SD*)	95 % CI	*M* (*SD*)	95 % BCI	*M* (*SD*)	95 % CI	*M* (*SD*)	95 % BCI
D	.24 (.02)	[.20 – .30]	.25 (.03)	[.20 – .30]	.33 (.02)	[.28 – .38]	.33 (.04)	[.25 – .41]
d	.70 (.12)	[.46 – .92]	.67 (.11)	[.46 – .89]	.28 (.08)	[.13 – .43]	.22 (.09)	[.07 – .40]
b	.45 (.02)	[.41 – .48]	.45 (.04)	[.37 – .53]	.53 (.02)	[.49 – .56]	.52 (.04)	[.45 – .60]
g	.60 (.02)	[.55 – .64]	.58 (.05)	[.49 – .67]	.78 (.02)	[.75 – .81]	.73 (.04)	[.64 – .81]

The parameters represent probability estimates that can range from 0 to 1. *D* = probability of item recognition; *d* = probability of remembering the source; *b* = probability of guessing that an item is old (chance level is .5); *g* = probability of guessing that an item was presented by the schematically expected source (estimates higher than the chance level of .5 indicate a guessing bias toward the schematically expected source; estimates lower than .5 indicate a guessing bias toward the schematically unexpected source); CI = confidence interval; BCI = Bayesian confidence interval

participants guessed that an item was from the expected source. The main hypothesis derived from the probability-matching account was hence confirmed.

The correlation between source memory parameter *d* and the absolute deviation of the contingency judgments from the true contingency of .5 was $r = .05$, $p = .81$, in the encoding condition (see Fig. 2c), but in the retrieval condition we found a significant negative correlation, $r = -.42$, $p = .04$ (see Fig. 2d), as expected. We found the same pattern for the correlations between source memory and the absolute deviation of the guessing bias *g* from the true contingency—that is, $r = .02$, $p = .46$, in the encoding condition and $r = -.62$, $p < .01$, in the retrieval condition. This means that the source-guessing bias was independent of source memory in the encoding condition, but in the retrieval condition there was a significant negative correlation. That is, participants with poor source memory showed a larger bias than did participants with good source memory.

Discussion

The main purpose of this study was to test the assumption of the probability-matching account that people guess according to individually perceived source–item contingencies if they do not remember the source in a source-monitoring task. Using the beta-MPT approach, we found medium to large correlations between perceived contingencies and guessing probabilities, both in a condition in which schematic information about the sources was known at encoding and in a condition in which that information was not known until retrieval.

Our hypothesis that the source-guessing parameter *g* should not differ from .5 in the encoding condition but should in the retrieval condition was not confirmed in the traditional analysis with aggregated data. The provision of schematic information about the sources at encoding should have improved contingency detection (Kuhlmann et al., 2012); however, individual differences in contingency detection had consequences for the source-guessing bias. In our sample, several participants misperceived the source–item contingency as somewhat conforming with schematic knowledge, and hence the overall source-guessing bias was above .5. This finding underscores the value of our individual-differences approach with the beta-MPT analysis. According to Smith and Batchelder (2010), one of the disadvantages of the traditional analysis is that it can result in confidence intervals that are too narrow, and therefore, goodness-of-fit tests can become significant too frequently. Thus, we can place more trust into the beta-MPT analysis. Because of individual differences, experimental manipulations do not have equal effects on all

Table 3 Posterior distributions of the parameters of the hierarchical beta distributions

	Encoding Condition				Retrieval Condition			
	M	*SD* [95 % BCI]	α	β	*M*	*SD* [95 % BCI]	α	β
D	.25	.04 [.01 – .11]	230.88	678.19	.33	.15 [.09 – .22]	3.10	6.38
d	.67	.10 [.01 – .25]	55.50	28.05	.22	.14 [.02 – .24]	5.59	34.64
b	.45	.18 [.13 – .24]	3.17	3.90	.52	.15 [.10 – .21]	5.73	5.20
g	.58	.20 [.15 – .26]	3.01	2.15	.73	.19 [.15 – .24]	3.17	1.11

D = probability of item recognition; *d* = probability of remembering the source; *b* = probability of guessing that an item is old; *g* = probability of guessing that an item was presented by the schematically expected source; BCI = Bayesian confidence interval of the standard deviation

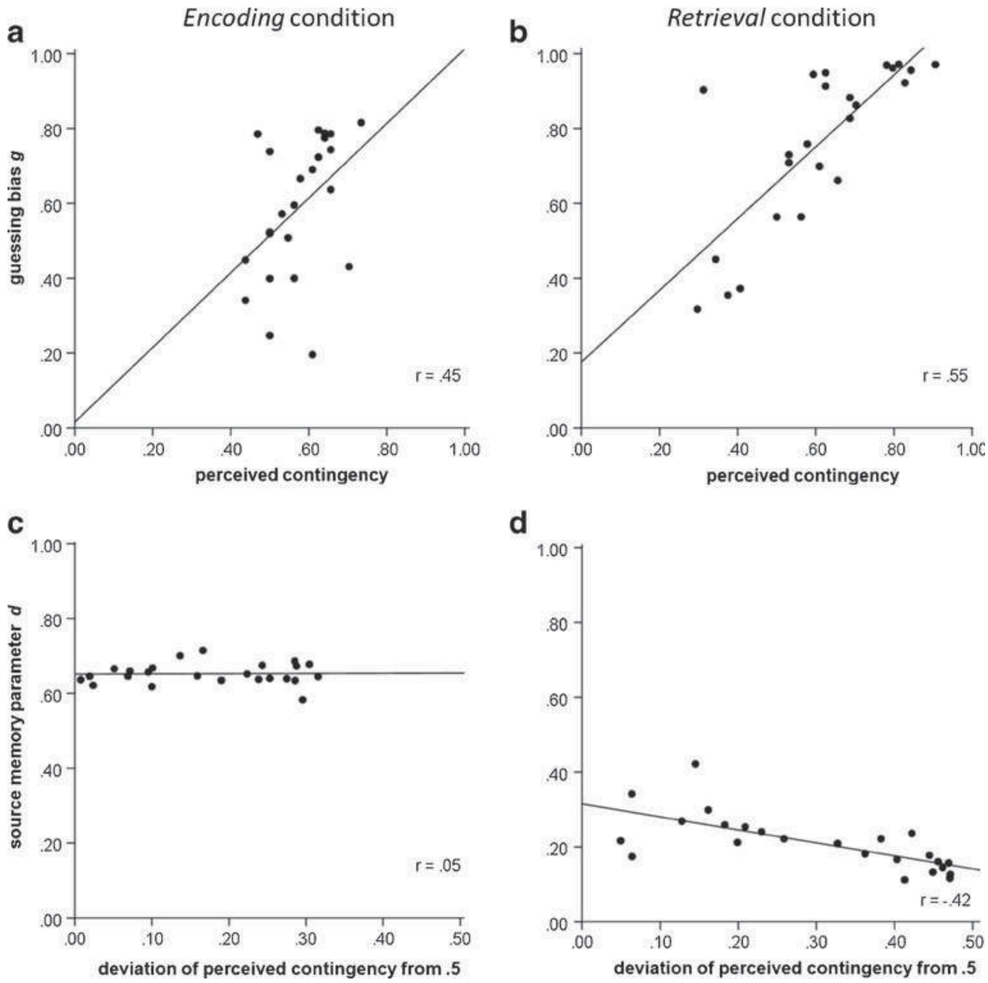


Fig. 2 **a** Correlation between contingency judgments (transformed to relative frequencies) and individual guessing parameters in the encoding condition. **b** Correlation between contingency judgments (transformed to relative frequencies) and individual guessing parameters in the retrieval condition. **c** Correlation between the deviations of participants' contingency judgments from the true contingency of .5 (transformed to relative frequencies) and their individual source memory parameters in the encoding condition. **d** Correlation between the deviations of participants' contingency judgments from the true contingency of .5 (transformed to relative frequencies) and their individual source memory parameters in the retrieval condition

participants. These individual differences are captured by the correlations made possible by the beta-MPT approach. However, even with traditional analyses, the source-guessing parameter was significantly higher in the retrieval condition, supporting the probability-matching account. In the encoding condition, we found no correlation between source memory and the deviation of the guessing bias from the true contingency. In the retrieval condition, however, we did find a significant negative correlation. Thus, the source-guessing bias was independent of source memory if participants learned about the professions of the sources before encoding. Possibly, in the encoding condition, even participants with poor source memory were able to recognize the true contingency during encoding. Thus, participants in this condition did not need good source memory to adjust their source guessing to the true contingency. For participants in the retrieval condition, on the other hand, it was more difficult to recognize the contingency during encoding; they may have recognized the contingency if they had good source memory, or else they adjusted their source guessing according to schematic knowledge.

Overall, our findings strongly support the probability-matching account of source guessing. We found a relationship between perceived source–item contingencies and source guessing, which is a core assumption of the probability-matching account. The results thus confirm a crucial prediction of this account. Contrary results would have meant falsification of the probability-matching account, which claims that participants match their response biases to the perceived ratio of different item types at test (Spaniol & Bayen, 2002). The findings concur with previous MPT analyses of aggregated data and, importantly, lend additional support through individual parameter estimates. Thus, for the first time, we have shown at an individual level that people match their source-guessing biases to perceived source–item contingencies.

Author note The research reported in this article was supported by Grant No. BA 3539/1-1 from the Deutsche Forschungsgemeinschaft. We thank Maike Lex and André Haese for assistance with the data collection.

References

Bayen, U. J., & Kuhlmann, B. G. (2011). Influences of source–item contingency and schematic knowledge on source monitoring: Tests of the probability-matching account. *Journal of Memory and Language, 64,* 1–17.

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 197–215. doi:10.1037/0278-7393.22.1.197

Bayen, U. J., Nakamura, G. V., Dupuis, S. E., & Yang, C.-L. (2000). The use of schematic knowledge about sources in source monitoring. *Memory & Cognition, 28,* 480–500. doi:10.3758/BF03198562

Buchner, A., Erdfelder, E., & Vaterrodt-Plünnecke, B. (1995). Toward unbiased measurement of conscious and unconscious memory processes within the process dissociation framework. *Journal of Experimental Psychology: General, 124,* 137–160. doi:10.1037/0096-3445.124.2.137

Ehrenberg, K., & Klauer, K. C. (2005). Flexible use of source information: Processing components of the inconsistency effect in person memory. *Journal of Experimental Social Psychology, 41,* 369–387.

Erdfelder, E., & Bredenkamp, J. (1998). Recognition of script-typical versus script-atypical information: Effects of cognitive elaboration. *Memory & Cognition, 26,* 922–938.

Estes, W. K., & Straughan, J. H. (1954). Analysis of a verbal conditioning situation in terms of statistical learning theory. *Journal of Experimental Psychology, 47,* 225–234.

Hicks, J. L., & Cockman, D. W. (2003). The effect of general knowledge on source memory and decision processes. *Journal of Memory and Language, 48,* 489–501. doi:10.1016/S0749-596X(02)00537-5

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3–28. doi:10.1037/0033-2909.114.1.3

Klauer, K. C. (2006). Hierarchical multinomial processing tree models: A latent-class approach. *Psychometrika, 71,* 7–31.

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika, 75,* 70–98.

Kuhlmann, B. G., Vaterrodt, B., & Bayen, U. J. (2012). Schema-bias in source monitoring varies with encoding conditions: Support for a probability-matching account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 1365–1376.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS—A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing, 10,* 325–337. doi:10.1023/A:1008929526011

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods, 42,* 42–54. doi:10.3758/BRM.42.1.42

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

Smith, J. B., & Batchelder, W. H. (2010). Beta-MPT: Multinomial processing tree models for addressing individual differences. *Journal of Mathematical Psychology, 54,* 167–183.

Spaniol, J., & Bayen, U. J. (2002). When is schematic knowledge used in source monitoring? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 631–651.

Is Prospective Memory Related to Depression and Anxiety? A Hierarchical MPT Modelling Approach

Nina R. Arnold, Ute J. Bayen, and Mateja F. Böhm

Heinrich-Heine-Universität Düsseldorf, Germany

Author Note

Nina Rebecca Arnold, Institute for Experimental Psychology, Heinrich-Heine-Universität Düsseldorf, Universitätsstr. 1, D-40225 Düsseldorf, phone: +49 621 181-2102, e-mail: niarnold@mail.uni-mannheim.de; Ute Johanna Bayen, Institute for Experimental Psychology, Heinrich-Heine-Universität Düsseldorf, Universitätsstr. 1, D-40225 Düsseldorf , phone: +49 211 81-12144, e-mail: ubayen@hhu.de; Mateja F. Böhm, Institute for Experimental Psychology, Heinrich-Heine-Universität Düsseldorf, Universitätsstr. 1, D-40225 Düsseldorf, phone: +49 211 81-12144; e-mail: Mateja.Boehm@hhu.de;

Nina R. Arnold is now in the Department of Psychology at the University of Mannheim.

Correspondence concerning this article should be addressed to Nina R. Arnold, Experimental Psychology, Schloss, Ehrenhof Ost, 68131 Mannheim, Germany. E-mail: niarnold@mail.uni-mannheim.de

Abstract

Prospective memory (PM) refers to remembering to perform an action in the future. One hundred thirty students completed a laboratory event-based PM task as well as depression and anxiety questionnaires. The data were analysed with the beta-MPT version (J. B. Smith & Batchelder, 2010) of the multinomial processing tree model of event-based PM (R. E. Smith & Bayen, 2004). Thereby, the prospective and retrospective components of PM were estimated for each participant and were then correlated with depression and anxiety. State anxiety was negatively correlated with the prospective component of PM. Neither depression nor trait anxiety were related to either component of PM.

Keywords: prospective memory, multinomial modelling, depression, anxiety, hierarchical modelling

Is Prospective Memory Related to Depression and Anxiety?
A Hierarchical MPT Modelling Approach

In a prospective memory (PM) task, we must remember to perform an action at an appropriate time in the future (e.g., Einstein & McDaniel, 1990). For event-based PM (e.g., McDaniel & Einstein, 2007), the appropriate time is defined by the occurrence of a specific event (e.g., taking medicine when chest pain occurs), whereas for time-based PM, the appropriate time is defined by a certain point in time (e.g., switching off the oven after 20 minutes). PM is very important in everyday life as PM failures may have serious consequences.

In daily life, performing a PM action often interrupts an ongoing activity. For example, one may have to stop watching a movie to take medicine. The interruption of an ongoing activity has been incorporated into many laboratory studies of PM; that is, the PM task is usually embedded in an ongoing activity. For event-based PM tasks, the appropriate action must be carried out in response to specific target events that may appear at any time during an ongoing task. For example, participants may be asked to press a certain key on a computer keyboard when a PM target event appears while they are busily engaged in a short-term memory task (e.g., Einstein & McDaniel, 1990). For time-based PM tasks, participants must remember to initiate the appropriate action at a certain point in time while engaged in a different ongoing activity.

It is important to distinguish between two different components of PM, namely, the prospective component and the retrospective component (Einstein & McDaniel, 1990). The prospective component refers to remembering *that* one must do something. In a laboratory paradigm, this means remembering that there is an additional task. The retrospective component refers to remembering *what* action to perform and *when* to perform it. In a laboratory paradigm, the retrospective component may be remembering which of the events occurring during the ongoing task are PM target events.

Several researchers have pointed out the role of mental health in PM (e.g., Harris & Cumming, 2003; Kliegel & Jäger, 2006; Rude, Hertel, Jarrold, Covich, & Hedlund, 1999). As PM is very important for everyday functioning, it is important to know how emotions and affective disorders influence PM. The results are mixed but mostly suggest that negative emotions and affective disorders negatively affect PM performance (for an overview, see Kliegel & Jäger, 2006). In particular, the influence of anxiety and depression on PM has been investigated as will be reviewed below.

The purpose of our study was to investigate the relationship between event-based PM and mental health using an innovative data analytical technique that allowed us to disentangle the prospective and retrospective components of PM and to correlate them with measures of depression and anxiety. We will first review the literature on the relationship between mental health and PM, separately for depression and anxiety, and discuss the importance of disentangling the retrospective and prospective components of PM to elucidate the relationship between PM and mental health. We will then explain our formal modelling approach before presenting our empirical study.

Prospective Memory and Depression

Depression and Event-Based Prospective-Memory Performance

Performance on event-based PM tasks is usually measured as the proportion of target events that participants respond to with the required PM action. So far, only a few studies have examined the influence of depression on event-based PM performance.

According to the resource allocation model of depression (Ellis & Ashbrook, 1988), depression limits the amount of resources that can be allocated to a task and, therefore, negatively affects task performance. In support of this view, studies have shown that depression accompanies working-memory impairments (De Lissnyder et al., 2012; Joormann, Levens, & Gotlib, 2011; Rose & Ebmeier, 2006). Impairments in cognitive resources should particularly affect performance in resource-demanding self-initiated processes that are often required to perform prospective memory tasks.

Livner, Berger, Karlsson, and Bäckman (2008) examined a sample of older nonclinically depressed adults with a naturalistic PM task. Participants had to remind the experimenter to do something at the end of the experiment. They found that performance was not influenced by depressive symptoms. On a similar task, Cuttler and Graf (2008) also did not detect a relationship between depression and PM for patients with obsessive-compulsive disorder. Lee et al. (2010) conducted a similar task with patients with bipolar disorder and healthy controls and found no difference in PM performance. Harris and Menzies (1999) found no influence of (nonclinical) depression on event-based PM performance.

However, the PM tasks used in these studies had only a single PM target. Event-based PM tasks with multiple different targets require more cognitive resources than tasks with single targets (Einstein & McDaniel, 2010). Only two studies that examined the relationship between depression and event-based PM have used multiple targets. Altgassen, Kliegel, and Martin (2009) used tasks with four PM trials, each with a different PM target. They found that event-based PM performance was impaired in a depressed group as compared with a control group for nonfocal PM tasks but not for focal PM tasks. Focality concerns whether or not the PM task requires the same type of item processing as the ongoing task. For instance, if the ongoing task involves colour discrimination, then a specific colour would be a focal PM target. By contrast, a specific word would be a nonfocal task on such an ongoing colour-discrimination task. Nonfocal tasks have been shown to require more cognitive resources than focal tasks (e.g., Brewer, Knight, Marsh, & Unsworth, 2010; McDaniel & Einstein, 2000). Accordingly, a relationship between working memory and PM performance has been shown for nonfocal tasks only (e.g., Brewer et al., 2010; Rose, Rendell, McDaniel, Aberle, &

Kliegel, 2010). Thus, Altgassen et al.'s (2009) results suggest that depression-related PM impairments are more likely when the task requires relatively high amounts of cognitive resources as is the case with nonfocal tasks and multiple targets. However, Kliegel and Jäger (2006) also used a nonfocal PM task with multiple targets. They did not find a relationship between sad mood that was induced in a nonclinical group of participants and event-based PM performance.

Depression and the Prospective Component of PM

It has been suggested that a lack of cognitive resources affects the prospective component of PM in particular, especially on nonfocal tasks (e.g., R. E. Smith, 2003; R. E. Smith & Bayen, 2004, 2005). In fact, studies in which the prospective component was measured separately have shown that this component was positively related to working-memory span (Arnold, Bayen, & Smith, 2014; R.E. Smith & Bayen, 2005; R.E. Smith, Persyn, & Butler, 2011). Hence, if the PM task requires high levels of self-initiated processing (as is expected on working-memory-demanding nonfocal tasks), we should find a negative relationship between the prospective component and depression.

Other tasks that demand high levels of self-initiation are time-based PM tasks, as they always require self-initiated time monitoring (e.g., through clock checking). In accordance with the resource allocation model (Ellis & Ashbrook, 1988), depression and negative mood have been found to negatively affect time-based PM tasks (Lee et al., 2010; Kliegel et al., 2005; Kliegel & Jäger, 2006; Rude et al., 1999). In fact, depression-related deficits are more evident on time-based PM tasks, which always require self-initiated processes, than for event-based PM tasks, which require self-initiated processes to varying degrees depending on task characteristics. This finding supports the idea that difficulties with self-initiation might be the reason for PM impairments in depression (Einstein, McDaniel, Richardson, Guynn, & Cunfer, 1995).

Depression and Retrospective Memory

In addition to the expected relationship between depression and the prospective component of PM on resource-demanding PM tasks, the retrospective task component may also contribute to depression-related differences in task performance. There is a long tradition of examining relationships between emotions and retrospective memory. A main result of this research is that negative emotions have been found to negatively affect retrospective memory (e.g., Ellis & Ashbrook, 1988). This was found in studies using mood induction (e.g., Ellis, Thomas, & Rodriguez, 1984) as well as in studies on depression (e.g., Potts, Camp, & Sturcke, 1989). Participants with depression have frequently been found to show deficits in recognition memory (Brand, Jolles, & Gispen-de Wied, 1992; Hertel & Milan, 1994; Ramponi, Murphy, Calder, & Barnard, 2010; Watts, Morris, & MacLeod, 1987).

There are two studies that examined both PM performance and retrospective memory with separate tests. Livner et al. (2008) found a negative relationship between depressive symptoms and both free and cued recall. Harris and Menzies (1999) found no relationship with recall. As most studies have shown a negative relationship between depression and retrospective memory including recognition, we assume that depression affects the retrospective component of PM that involves recognizing target events.

If depression negatively affects retrospective memory, then it is possible that the effects of depression on PM performance might be due to differences in the retrospective component alone. Therefore, to test the hypothesis that depression affects a resource-demanding prospective component, it would be essential to measure the two components separately. None of the prior studies on depression and PM performance have disentangled the two underlying components. Thus, the results of these studies cannot be ascribed to either component.

Prospective Memory and Anxiety

Anxiety is a construct that is closely related to depression (Last, Strauss, & Francis, 1987; Overall, 1962), and this is why both constructs are often included in one study. There are two kinds of anxiety. Trait anxiety refers to a person's basic level of anxiety, whereas state anxiety usually refers to the anxiety that a person is currently experiencing.

Anxiety and Event-Based Prospective-Memory Performance

Trait anxiety. The results regarding the relationships between event-based PM and both trait and state anxiety have been mixed. Harris and Menzies (1999) as well as Kliegel and Jäger (2006) found a significant negative relationship between trait anxiety and performance on nonfocal event-based PM tasks with nonclinical samples. However, Cuttler and Graf (2008) did not find a relationship between trait anxiety and event-based PM with a naturalistic nonfocal PM task involving one target only. Harris and Cumming (2003) did not find a relationship between trait anxiety and a focal-event-based PM task.

State anxiety. Harris and Cumming (2003) found that participants with higher levels of state anxiety showed poorer performance than participants with lower levels of state anxiety. By contrast, Cuttler and Graf (2008) did not find a relationship between state anxiety and event-based (nonfocal) PM. Cockburn and Smith (1994) stated that the relationship between state anxiety and a naturalistic event-based nonfocal PM task resembled a complex curvilinear function.

Anxiety and the Prospective Component of PM

As stated above, the prospective component of PM requires working-memory capacity when performing nonfocal event-based tasks. Similar to depression, trait anxiety impairs working memory (Eysenck, 1985; MacLeod & Donnellan, 1993). Stout and Rokke (2010) also found a link between working memory and anxiety. Highly anxious participants were found to have lower working-memory capacity than participants with low levels of anxiety (Darke, 1988). Therefore, it would be expected that higher levels of trait and state anxiety would accompany a lower prospective component.

Anxiety and Retrospective Memory

As discussed for depression, it is conceivable that the effects of anxiety on PM performance are due to differences in the retrospective component alone. Therefore, we next review the findings on anxiety and retrospective memory performance.

Trait anxiety and retrospective memory. De Pascalis and Morelli (1990) found that trait anxiety did not affect recognition memory. Beato, Pulido, Pinho, and Gozalo (2013) found that trait anxiety did not affect correctly and falsely recognized words in the Deese-Roediger-McDermott (DRM) paradigm in both a correlational analysis and an extreme-groups analysis. Mathews and MacLeod (1985) compared patients who were treated for an anxiety disorder and a matched control group. The groups differed significantly in trait anxiety but not in recognition memory for both threat-related and neutral words.

State anxiety and retrospective memory. Eysenck, Derakshan, Santos, and Calvo (2007) proposed that state anxiety may not impair recognition memory because it may lead to the use of compensatory strategies. For state anxiety, Beato et al. (2013) found no effect on correctly and falsely recognized words in the DRM paradigm in both a correlational analysis and an extreme-groups analysis. Mathews and MacLeod´s (1985) groups also differed significantly in state anxiety but not in recognition memory for threat-related and neutral words.

As in studies on depression, it is important to separately measure the prospective and the retrospective components of PM in studies on anxiety and PM. However, the two components were not disentangled in any of the cited studies. In some of the studies, both a PM task and a retrospective memory task were administered, but the PM task was still a conglomerate of both components. Separating the two components is not only important for theoretical reasons but is also crucial for clinical interventions. If retrospective memory is impaired, patients might write lists or might learn mnemonic techniques or use other tools to remember target events. If, however, the prospective component of PM is impaired, patients need different types of interventions such as reminders or special routines to remember to implement their everyday PM tasks.

The MPT Model of Event-Based PM

R. E. Smith and Bayen (2004) developed a multinomial processing tree (MPT) model of event-based PM. This is a stochastic model that allows the prospective and retrospective components of PM to be measured separately. Thus, this model can be used to determine how different variables affect each of the two components and how this contributes to PM performance. In our study, we applied the beta-MPT version of this model (J. B. Smith & Batchelder, 2010) to yield individual parameter estimates that we then compared to measures of anxiety and depression. Thus, we examined the effects of mental health on the retrospective and prospective components of PM separately. We will now describe the MPT model and the beta-MPT approach. After that, we will state our hypotheses with respect to the model parameters.

MPT models are useful tools when attempting to unravel latent cognitive processes (Batchelder & Riefer, 1999) and have been applied in various domains of cognitive psychology (see Erdfelder et al., 2009, for a review). R. E. Smith and Bayen (2004) developed an MPT model that disentangles the retrospective and prospective components of PM. It has been applied frequently (e.g., Pavawalla, Schmitter-Edgecombe, & Smith, 2012; Schnitzspahn, Horn, Bayen, & Kliegel, 2012; R. E. Smith & Bayen, 2005, 2006; R. E. Smith, Bayen, & Martin, 2010; R. E. Smith et al., 2011) and validated thoroughly (Horn, Bayen, Smith, & Boywitt, 2011; R. E. Smith & Bayen, 2004).

The model was developed for ongoing tasks with two response options and an embedded PM task. We will explain the model using the example of the ongoing colour-matching task that was used in a number of prior studies (e.g., Horn, Bayen, Smith, & Boywitt, 2011; R. E. Smith & Bayen, 2004, 2006; R. E. Smith et al., 2010) as well as in the current study. Participants are shown four coloured rectangles followed by a coloured word. The ongoing task is to decide whether or not the colour of the word matches the colour of one of the preceding rectangles. The embedded PM task is to press a special key when specific words appear. There are four possible trial types in this task: (a) a PM target appears, and the colours match (target, match); (b) no PM target appears, but the colours match (nontarget, match); (c) a PM target appears, but the colours do not match (target, nonmatch); and (d) no PM target occurs, and the colours do not match (nontarget, nonmatch).

The model (Figure 1) includes one tree for each trial type in which any of the three responses (i.e., "match", "nonmatch", and "PM target") can occur. The first tree refers to PM targets occurring in trials in which the colours match. Parameter C_1 denotes the probability that a colour match is detected. The complementary probability of 1 - C_1 is the probability that a colour match goes undetected. P is the probability that the participant remembers *that* there was an additional PM task (the prospective component). Thus, 1 - P is the probability that the participant does not remember the PM task. Detecting a colour match but not remembering the PM task leads to a "Match" response. If the participant detects the colour match and remembers the PM task, M_1 is the probability that the PM target is recognized, leading to a "PM" response. If the target is not recognized (with a probability of 1 - M_1), the participant must guess whether a target is present or not. With probability g, the participant guesses that there is a target, leading to a "PM" response, whereas with probability 1 - g, the participant guesses that there is no target, leading to a "Match" response.

The lower part of the first tree illustrates trials in which the colour match was not detected (with a probability of 1 - C_1). Nevertheless, the participant may remember the PM task (with probability P) or may not remember the PM task (with probability 1 - P). If the PM task is remembered, a target is recognized with probability M_1. If the PM target is not recognized (1 - M_1), guessing leads to a "PM" response with probability g. 1 - g is the probability of guessing that no PM target is present. Because a colour match was initially not detected (with probability 1 - C_1), the participant must guess whether the colours in the trial matched (guessing probability c) or not (1 - c), leading to a "Match" or a "Nonmatch" response, respectively. Similarly, if the PM task was not remembered (with probability 1 - P), participants must guess whether there was a colour match (c) or not (1 - c).

The second tree refers to PM targets occurring in trials in which the colours do not match. It is similar to the first tree. The main difference is that C_1 is replaced by C_2, which denotes the probability that a participant detects that the colours do not match. As a result, the upper half of the tree is the same with the only difference being that remembering the PM task but not recognizing the PM target can lead to a "Nonmatch" response. Additionally, not remembering the PM task results in a "Nonmatch" response. The lower half of the second tree is exactly the same as the lower half of the first tree.

The third tree, for trials in which the colours match but without a PM target, is also similar to the first tree with the exception that if participants remember the PM task, M_2 is the probability of recognizing that a word is not a PM target, resulting in a "Match" response. If participants fail to recognize that the word is not a PM target (1 - M_2), they must guess. Guessing, again, can lead to a "PM" response, with probability g, or to a "Match" response, with probability 1 - g.

The fourth tree refers to trials in which the colours do not match but without a PM target. It is almost identical to the third tree. The only difference is that C_1 is replaced by C_2, that is, the probability of detecting that the colours do not match. Therefore, the responses in the upper half of the fourth tree are different from the responses in the upper half of the third tree. That is, instead of leading to "Match" responses, the upper half of the fourth tree leads to

"Nonmatch" responses. Again, the lower half of the fourth tree is exactly the same as the lower half of the third tree.

As described in detail in R. E. Smith and Bayen (2004), the model is identifiable only after posing some theoretically motivated constraints on the model parameters. The guessing parameter *c* is set to the probability of a colour match occurring during the task. Similarly, the guessing parameter *g* is set to the probability of a PM target occurring during the ongoing task. $M_1 = M_2$ reflects the assumption that recognizing a PM target and recognizing that an item is not a PM target are equally likely. All in all, these constraints limit the number of free parameters in the MPT model to four: P, M, C_1, and C_2.

Beta-MPT Models

In traditional MPT modelling, all participants and all items are assumed to be homogeneous, that is, to have the same parameter values (J. B. Smith & Batchelder, 2008). If this assumption does not hold, parameter estimates may be biased (Klauer, 2006, 2010; J. B. Smith & Batchelder, 2008, 2010). In addition, the traditional approach yields only group-wise parameter estimates. J. B. Smith and Batchelder (2010) introduced beta-MPT models, which can be used to yield individual parameter estimates and thus to investigate individual differences. These models assume that each parameter is independently beta distributed. Beta distributions are parameterized by two different variables, namely α and β, which can range between 0 and ∞. The resulting range of parameter values lies between 0 and 1, thus meeting the requirements for probabilities. This provides us with individual model parameter estimates and thus allows us to run correlations with other variables. J. B. Smith and Batchelder (2010) described a method for using beta distributions with the pair-clustering MPT model (Batchelder & Riefer, 1980, 1986). We customized this method for the MPT model of event-based PM. The method uses Markov Chain Monte Carlo (MCMC) methods that can be applied using the WinBUGS software (Lunn, Thomas, Best, & Spiegelhalter, 2000). Thereby, we obtained individual parameter estimates as well as estimates for the hyperparameters α and β, including uncertainty estimates.

Aim of the Study and Hypotheses

None of the studies on PM and mental health cited above used an MPT model. Therefore, it would be difficult to ascribe the results to the prospective or retrospective components of PM or to determine how anxiety or depression influences the contributions of the underlying processes to PM performance. In the current study, we investigated whether depression, state anxiety, and trait anxiety influence PM performance. If depression or anxiety influences PM performance, it would be of great interest to determine whether this effect is due to differences in the prospective component or the retrospective component of PM. To compute correlations between the participants' model parameters and their scores on different tests of depression and anxiety, we applied the beta-MPT version of the MPT model by R. E. Smith and Bayen (2004).

We used an ongoing colour-matching task with a nonfocal PM task. The ongoing task focused on the colours of rectangles and words, whereas the PM task focused on the words themselves and was thus nonfocal with respect to the ongoing task. We used this nonfocal PM task with multiple targets because such a task is very resource demanding and thus allowed us to test the hypothesis that depression-related resource deficits would negatively affect the prospective component as measured by model parameter *P*. The *M* parameter of the MPT model denotes the probability with which participants will recognize PM targets. As reviewed above, depression has been shown to affect recognition memory. Thus, we expected estimates of the *M* parameter to decrease as levels of depression increased. As reviewed earlier, most studies have shown a negative relationship between PM performance and anxiety. As we explained above, the prospective component *P* is related to working-memory capacity on nonfocal tasks with multiple targets. Similar to depression, trait anxiety

(Eysenck, 1985; MacLeod & Donnellan, 1993) and state anxiety (Darke, 1988) have been found to impair working memory. Therefore, both state anxiety and trait anxiety were predicted to negatively influence the prospective component *P*. As outlined above, neither state (Eysenck et al., 2007) nor trait anxiety (De Pascalis & Morelli, 1990) affects recognition memory. Therefore, we did not predict a correlation between the retrospective component *M* and anxiety. As anxiety is assumed to have a negative impact on the estimate of parameter *P*, it was also expected to negatively influence PM performance.

Method

Participants

Participants were 129 students at the University of Düsseldorf. Seventy-nine of them were female; their ages ranged from 18 to 52 years ($M = 22.31$, $SD = 5.48$). All of them were native German speakers, and none of them suffered from achromatopsia. They received either course credit or € 10 for their participation. One additional participant never gave a “PM” response and, thus, presumably did not understand the instructions. Therefore, we excluded this participant from all analyses. The resulting statistical power to detect a medium effect of $r = .30$ with alpha = .05 was .97.

Measures and Materials

The Beck Depression Inventory II. We used the Beck Depression Inventory II (BDI-II; Beck, Steer, & Brown, 1996; translated into German by Hautzinger, Keller, & Kühner, 2006) to measure depression. The BDI-II has a high retest reliability, ranging from $r_\alpha = .74$ to $r_\alpha = .96$ and is strongly correlated with other depression scales, $r = .68$ to $r = .89$ (Hautzinger et al., 2006). For 21 items, participants select the statement that applies most to them. Each of the statements is designated a value between 0 and 3. The BDI-II is scored by simply adding the values of the statements. Thus, the total scores awarded on the BDI-II can range from 0 to 63, with higher scores indicating stronger depression.

Hospital Anxiety and Depression Scale. The German version of the Hospital Anxiety Depression Scale (HADS-D; Herrmann, Buss, & Snaith, 1995) was adapted from the original version developed by Zigmond and Snaith (1983). This self-report scale is composed of two subscales, one measuring depression, and the other one measuring anxiety. Each subscale consists of seven items with four different response alternatives. Every response alternative is assigned a value ranging from 0 to 3. The values are added separately for each of the subscales, thus leading to total scores on the subscales that range between 0 and 21, with higher scores indicating more severe depression or anxiety. The HADS-D has a satisfactory validity and its retest reliability is $r = .71$ (Herrmann et al., 1995).

State-Trait Anxiety Inventory. The State-Trait Anxiety Inventory (STAI) was developed by Spielberger, Gorsuch, and Lushene (1970) and translated into German by Laux, Glanzmann, Schaffner, and Spielberger (1981). The STAI assesses levels of state anxiety and trait anxiety. It contains two subscales with 20 items each, which are rated on a 4-point Likert scale. The state anxiety subscale asks participants to indicate their momentary feelings by rating the intensity of the anxiety they are experiencing. The trait anxiety subscale is aimed at determining a person's level of general anxiety by collecting information about how the person feels in general. The STAI is evaluated by adding the values of each response alternative ranging from 1 to 4 separately for each subscale. Higher scores indicate higher levels of anxiety. The STAI has satisfactory reliability and validity (Laux et al., 1981).

Word items for the PM task. We selected 168 words from the CELEX database (Baayen, Piepenbrock, & Gulikers, 1995). The words were between five and eight letters long with two or three syllables and a frequency between 16 and 40 per million. Emotional valence varied between -1 and 3; arousal varied between 1.3 and 4.1. From these 168 words, we chose 15 to serve as PM targets, leaving 153 words to serve as distractor items. The target and distractor items were matched in length, emotional valence, and frequency. The targets and distractor words were the same for each participant.

Procedure

The ongoing task with the embedded PM task was computer based. Participants were tested in groups of up to six in separate computer booths. Computerized instructions informed them that the study was aimed at investigating colour-matching abilities, followed by the instructions for the colour-matching task. The assignments of the "M" and "V" keys to the answers *match* and *nonmatch* were counterbalanced. Both speed and accuracy were emphasized. After reading the instructions, participants were given 10 trials of the colour-matching task for practice. For this task, we selected five different colours: yellow, red, blue, green, and white. The colours were shown as rectangles that were 166 x 120 pixels in size. One trial consisted of four coloured rectangles presented in the middle of a black screen for 500 ms each, followed by a blank screen for 250 ms before the next rectangle was presented. After the fourth rectangle, a coloured word appeared in the middle of the screen in a 24-point font size. The participants judged whether the colour of the word matched one of the colours of the previously presented rectangles or not (colour-match vs. colour-nonmatch). The word stayed on the screen until the participants pressed a key. The first rectangle in the next trial appeared after a blank screen that lasted for 250 ms.

After the practice trials, the PM task was introduced. Participants were told that they would see five words that they should memorize. Participants were instructed to press the space bar instead of the *match* or *nonmatch* keys if one of these words appeared during the colour-matching task. The five PM targets were then consecutively presented on the screen for 5 s each. After that, a 2-min retention interval followed during which participants filled out a demographics questionnaire.

The ongoing task with the embedded PM task consisted of three blocks with 112 trials each. We divided the 168 words into three sets comparable in frequency, emotional valence and arousal with every set appearing in one of three blocks of the PM task. The order of the sets was counterbalanced across participants. Each set consisted of five PM targets and 51 distractor words. After the first half of each block, participants took a break of 60 s before the second half started. It contained the same words as the first half and was also followed by a 60-s break. Following this break, five new PM targets were presented, and the participants were asked to memorize them. Again, presentation of the PM targets was followed by a 2-min break, designated to filling out the demographics questionnaire. Then the second block of the PM task began. After completing the second half of this block, the new PM targets for the third block were presented. Again, the presentation of the PM targets was followed by a 2-min break during which participants answered questions on the demographics questionnaire. Following the last block, participants were asked some questions concerning the PM task; namely, whether they perceived speed or accuracy as more important during the task, whether they remembered which key they were supposed to press upon encountering a PM target, whether they remembered to press the space bar at all during the experiment, and whether they applied a certain strategy for memorizing the PM targets. Then the 15 PM targets and the same number of distractor words were presented one after the other in a random order, and participants were asked to indicate whether each of the presented words was a PM target or not. This recognition-memory test was self-paced.

After the PM task, participants completed several questionnaires that we administered in a paper-and-pencil format. These questionnaires included, in order, a questionnaire on caffeine consumption, the short version of a questionnaire investigating achievement motivation (Leistungsmotivationsinventar-kurz; Schuler & Prochaska, 2001), and German versions of the following tests: Fagerström-Test for Nicotine Dependence (Bleich, Havemann-Reinecke, & Kornhuber, 2002), the Epworth Sleepiness Scale (Bloch, Schoch, Zhang, & Russi, 1999), the Karolinska Sleepiness Scale (Åkerstedt & Gillberg, 1990), the Prospective and Retrospective Memory Questionnaire (Crawford, Henry, Ward, & Blake, 2006; German translation by Kaschel, 2002), the HADS-D (Herrmann et al., 1995), the BDI-II (Hautzinger et al., 2006), the STAI (Laux et al., 1981), and a translation of the passionate love scale (Hatfield & Sprecher, 1986). For the present study, we will report the results of the BDI-II, HADS-D, and STAI only. Finally, the participants were debriefed and paid.

Results

Prospective Memory Task and Ongoing Task

On average, participants pressed the PM key on *M* = 20.91 (*SD* = 6.38) PM trials; hence, 69.69% of the PM targets were correctly responded to (*SD* = 0.21). For the ongoing task, participants answered *M* = 274.45 trials (*SD* = 23.44) correctly; that is, 89.69% of the trials (*SD* = 0.08). If participants corrected themselves (i.e., they pressed a key even though they had already submitted an answer), we used the second corrected answer instead of the first one. However, using only the first answer did not change the pattern of the results.

We calculated MPT model parameters by adapting J. B. Smith and Batchelder's (2010) version of the Markov Chain Monte Carlo (MCMC) algorithm to the MPT model of event-based PM.[1] The algorithm was run with the computer program WinBUGS 14 (Lunn et al., 2000) with 1,000,000 iterations. The first half of these iterations was removed as a burn-in period. Convergence was assessed using the potential scale reduction factor $\hat{R}$ provided by the R2WinBUGS package (Sturtz, Ligges, & Gelman, 2005). For all parameters, $\hat{R} < 1.05$, indicating good convergence. The group parameter estimates are presented in Table 1 (Table 1 about here).

Prospective Memory and Depression

Descriptive statistics for the BDI-II, HADS-D, and STAI are listed in Table 2. For the BDI-II, one participant's data were not included in the analysis because he failed to fill out the second page. The mean of 9.64 indicates that the sample was minimally depressed (Hautzinger et al., 2006). Only one participant scored higher than 29 on the BDI-II, indicating depression. However, nine participants produced scores between 20 and 28, indicating moderate depression. On the depression subscale of the HADS-D, five participants produced scores equal to or above the cut-off value of 9 for depression. (Table 2 about here)

We calculated correlations between individual MPT model parameter estimates and depression scores (see Table 3). For all analyses, we set α to .05. The BDI-II scores were not significantly correlated with the prospective component *P* or the retrospective component *M*. Similarly, the depression scale of the HADS-D did not produce significant results when correlated with Parameter *P* and Parameter *M*. Additionally, neither depression score was significantly correlated with the number of PM hits (BDI: $r = -.03$, $p = .358$; HADS: $r = -.03$, $p = .367$). (Table 3 about here)

Prospective Memory and Anxiety

Means and standard deviations for the anxiety measures are listed in Table 2; coefficients for correlations between anxiety and PM are presented in Table 3. On the anxiety subscale of the HADS-D, the scores of 19 participants equaled or exceeded the cut-off value of 11 for anxiety. The STAI does not provide cut-off values. Neither scale was significantly related to the number of PM hits, the prospective component *P*, or the retrospective component *M* (all *p*s > .05).

The state-anxiety subscale of the STAI was significantly correlated with PM performance as measured by the number of PM hits. That is, participants with lower state anxiety showed better PM performance. The beta-MPT parameter estimates showed that this difference was due to the negative correlation between state anxiety and the prospective component *P*. There was no significant correlation with the retrospective component *M*. Finally, the trait subscale of the STAI was not significantly correlated with any of the parameter estimates or with the number of PM hits (all *p*s > .05).

Discussion

The main purpose of our study was to examine the effects of anxiety and depression for the different components of PM separately, namely, the prospective and retrospective components. We found a negative correlation between state anxiety and the prospective component of PM. State anxiety was not correlated with the retrospective component of PM. Neither depression nor trait anxiety were significantly correlated with either component of PM.

Most of the previous studies examining the role of mental health in PM used the traditional measure of PM performance (i.e., PM hits). This is the first study to investigate how the different underlying components of PM are influenced by depression and anxiety. By applying the new beta-MPT approach combined with the MPT model of event-based PM, we were able to look at individual differences in depression, anxiety, and different components of PM simultaneously.

The results concerning the relationship between depression and prospective memory were not consistent with our hypotheses. Like most of the previous studies, we did not find a relationship between depression and PM performance. Depression was not related to the Parameters *P* and *M* or to PM hits. This is in line with the results obtained by Cuttler and Graf (2008) and Harris and Menzies (1999), although we included more observations of PM performance. Our task was an event-based PM task. Depression has been shown to influence time-based PM (e.g., Kliegel & Jäger, 2006) but event-based PM only with a nonfocal task and multiple targets (Altgassen et al., 2009). Because the negative relationship between depression and time-based PM is assumed to be due to a lack of self-initiation (Kliegel et al., 2005; Rude et al., 1999), performance on nonfocal event-based PM tasks (which also require self-initiated processes) may be more vulnerable to effects of depression than focal tasks (cf.

McDaniel & Einstein, 2000). Furthermore, because working memory is related to the *P* parameter on nonfocal tasks, and the *M* parameter measures recognition memory, we predicted that deficits in working memory and recognition memory, as found in depressed individuals, might lead to a decrease in the corresponding parameter estimates on our nonfocal event-based task.

As these hypotheses were not confirmed, we may draw several conclusions. First, it is possible that depression-related deficits occur only in time-based PM as opposed to event-based PM. As Einstein and McDaniel (1990) argued, time-based PM requires more self-initiation than any type of event-based PM, and this difference may account for the absence of a depression-related impairment in Parameter *P* and in PM performance in the present study. Second, sample characteristics should be considered. As we used a normal student sample, it is possible that levels of depression were not high enough to produce PM impairments. Event-based prospective memory may be impaired in clinically depressed individuals but not in nonclinical samples. Supporting this idea, only Altgassen et al. (2009) who used a clinically depressed sample found group differences in nonfocal event-based PM compared with normal controls.

In interpreting the results of our study, we should also consider that the retrospective component *M* was quite high at .89. Some participants reached the ceiling, and thus a restriction of range might account for the lack of significant correlations between this parameter and the depression scores. Replication of our study with samples including a larger range of both depression scores and cognitive abilities would be desirable but difficult to achieve due to the large number of participants with depression that would be needed to achieve sufficient statistical power for the tests of the correlations.

Our hypothesis regarding a negative influence of anxiety on PM was confirmed in part. State anxiety as measured with the STAI was negatively correlated with the prospective component *P* and with PM performance. This result was as predicted because the prospective component *P* is related to working memory capacity (Arnold et al., 2014; R. E. Smith & Bayen, 2005; R. E. Smith et al., 2011), and state anxiety has been found to impair working memory (Darke, 1988). However, neither trait anxiety as measured with the STAI nor the anxiety subscale of the HADS-D showed a relationship with PM performance or the prospective component *P*. As hypothesized, none of the anxiety measures were significantly correlated with the retrospective component *M*.

The results obtained for anxiety emphasize the importance of differentiating between state anxiety and trait anxiety. Only state anxiety was correlated with the prospective component and with PM performance. State anxiety scores are closely linked to the situation in which a test is taken and thus, situational influences on PM seem to be quite important. Trait anxiety, on the other hand, which is not specific to the situation but is more robust and long-lasting, left PM unaffected. The same holds for measures of depression; the HADS-D and BDI-II are both questionnaires that refer to time periods of several weeks and may therefore be less related to PM.

Working-memory deficits that are related to state anxiety (Stout & Rokke, 2010) may provide an explanation for the negative correlation between state anxiety and Parameter *P*. However, trait anxiety did not impair PM, although working memory deficits are also assumed to exist in trait anxiety (Eysenck, 1985; MacLeod & Donnellan, 1993). The present study is the first to apply the MPT model of event-based PM to examine depression and anxiety in relation to PM. We found that state anxiety was correlated only with the prospective component but not the retrospective-memory component, thus demonstrating the usefulness of distinguishing between these components when measuring PM.

Acknowledgements

We thank Pia Ewerdwalbesloh for assisting with the data collection and Jane Zagorski for providing comments on a prior version of the manuscript. Part of this research was presented at the Conference "Tagung experimentell arbeitender Psychologen" (TeaP) in Vienna (2013).

References

Åkerstedt, T., & Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience, 52,* 29–37.

Altgassen, M., Kliegel, M., & Martin, M. (2009). Event-based prospective memory in depression: The impact of cue focality. *Cognition and Emotion, 23*, 1041–1055.

Arnold, N. R., Bayen, U. J., & Smith, R. E. (2014). *A hierarchical MPT modeling approach to investigating the relationship between prospective memory and working memory.* Manuscript submitted for publication.

Baayen, R.H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database (CDROM)*. University of Pennsylvania, Philadelphia, PA: Linguistic Data Consortium.

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

Batchelder, W. H., & Riefer, D. M. (1986). The statistical analysis of a model for storage and retrieval processes in human memory. *British Journal of Mathematical and Statistical Psychology, 39,* 129–149.

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

Beato, M. S., Pulido, R. F., Pinho, M. S., & Gozalo, M. (2013). Reconocimiento falso y ansiedad estado/rasgo [False recognition and state/trait anxiety]. *Psicológica*, *34*, 299–311.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II.* San Antonio, TX: Psychological Corporation.

Bleich, S., Havemann-Reinecke, U., & Kornhuber, J. (2002). *Fagerström-Test für Nikotinabhängigkeit (FTNA) [Fagerström-Test for Nicotine Dependence]*. Göttingen: Hogrefe Verlag.

Bloch, K. E., Schoch, O. D., Zhang, J. N., & Russi, E. W. (1999). German Version of the Epworth Sleepiness Scale. *Respiration, 66,* 440–447.

Brand, A. N., Jolles, J., & Gispen-de Wied, C. (1992). Recall and recognition memory deficits in depression. *Journal of Affective Disorders, 25,* 77–86.

Brewer, G. A., Knight, J. B., Marsh, R. L., & Unsworth, N. (2010). Individual differences in event-based prospective memory: Evidence for multiple processes supporting cue detection. *Memory & Cognition, 38*, 304–311.

Cockburn, J., & Smith, P. T. (1994). Anxiety and errors of prospective memory among elderly people. *British Journal of Psychology, 85,* 273–282.

Crawford, J. R., Henry, J. D., Ward, A. L., & Blake, J. (2006). The Prospective and Retrospective Memory Questionnaire (PRMQ): Latent structure, normative data and discrepancy analysis for proxy-ratings. *British Journal of Clinical Psychology, 45,* 83–104.

Cuttler, C., & Graf, P. (2008). Sub-clinical checking compulsions are related to impaired prospective memory independently of depression, anxiety and distractibility. *Journal of Anxiety Disorders, 22,* 642–654.

Darke, S. (1988). Anxiety and working memory capacity. *Cognition & Emotion, 2,* 145–154.

De Lissnyder, E., Koster, E. H. W., Everaert, J., Schacht, R., Van den Abeele, D., & De Raedt, R. (2012). Internal cognitive control in clinical depression: General but no emotion-specific impairments. *Psychiatric Research, 199,* 124–130.

De Pascalis, V., & Morelli, A. (1990). Anxiety and individual differences in event-related potentials during the recognition of sense and nonsense words. *Personality and Individual Differences, 11*, 741–749.

Einstein, G. O., & McDaniel, M. A. (1990). Normal aging and prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 717–726.

Einstein, G. O., & McDaniel, M. A. (2010). Prospective memory and what costs do not reveal about retrieval processes: A commentary on Smith, Hunt, McVay, and McConnell (2007). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1082–1088.

Einstein, G. O., McDaniel, M. A., Richardson, S. L., Guynn, M. J., & Cunfer, A. R. (1995). Aging and prospective memory: Examining the influences of self-initiated retrieval processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 996–1007.

Ellis, H. C., & Ashbrook, P. W. (1988). Resource allocation model of the effects of depressed mood states on memory. In K. Fiedler & J. Forgas (Eds.), *Affect, cognition and social behavior* (pp. 25–42). Toronto: Hogrefe.

Ellis, H. C., Thomas, R. L., & Rodriguez, I. A. (1984). Emotional mood states and memory: Elaborative encoding, semantic processing, and cognitive effort. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 470–482.

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Journal of Psychology, 217*, 108–124.

Eysenck, M. W. (1985). Anxiety and cognitive-task performance. *Personality and Individual Differences, 6*, 579–586.

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion, 7,* 336–353.

Harris, L. M., & Cumming, S. R. (2003). An examination of the relationship between anxiety and performance on prospective and retrospective memory tasks. *Australian Journal of Psychology, 55,* 51–55.

Harris, L. M., & Menzies, R. G. (1999). Mood and prospective memory. *Memory, 7*, 117–

127.

Hatfield, E., & Sprecher, S. (1986). Measuring passionate love in intimate relationships. *Journal of Adolescence, 9,* 383–410.

Hautzinger, M., Keller, F., & Kühner, C. (2006). *BDI II: Beck Depressions-Inventar - Manual* [BDI II: Beck Depression-Inventory – Manual]. Frankfurt am Main: Harcourt Test Services.

Herrmann, C., Buss, U., & Snaith, R. P. (1995). *HADS-D: Hospital Anxiety and Depression Scale – Deutsche Version – Ein Fragebogen zur Erfassung von Angst und Depressivität in der somatischen Medizin* [HADS-D: Hospital Anxiety and Depression Scale – German Version – A Questionnaire Measuring Anxiety and Depression in Psychosomatic Medicine]. Bern: Verlag Hans Huber.

Hertel, P. T., & Milan, S. (1994). Depressive deficits in recognition: Dissociation of recollection and familiarity. *Journal of Abnormal Psychology, 103*, 736–742.

Horn, S. S., Bayen, U. J., Smith, R. E., & Boywitt, C. D. (2011). The multinomial model of prospective memory: Validity of ongoing-task parameters. *Experimental Psychology, 58*, 247–255.

Joormann, J., Levens, S. M., & Gotlib, I. H. (2011). Sticky thoughts: Depression and rumination are associated with difficulties manipulating emotional material in working memory. *Psychological Science, 22,* 979–983.

Kaschel, R. (2002). *Erinnern von Erledigungen* [Remembering of Plans]. Retrieved from http://www.psy.ed.ac.uk/psy_research/documents/prmq/grprmq-1.pdf

Klauer, K. C. (2006). Hierarchical multinomial processing tree models: A latent-class approach. *Psychometrika, 71*, 7–31.

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika. 75,* 70–98.

Kliegel, M., & Jäger, T. (2006). The influence of negative emotions on prospective memory: A review and new data. *International Journal of Computational Cognition, 4,* 1–17.

Kliegel, M., Jäger, T., Phillips, L. H., Federspiel, E., Imfeld, A., Keller, M., & Zimprich, D. (2005). Effects of sad mood on time-based prospective memory. *Cognition and Emotion, 19,* 1199–1213.

Last, C. G., Strauss, C. C., & Francis, G. (1987). Comorbidity among childhood anxiety disorders. *Journal of Nervous and Mental Disease, 175,* 726–730.

Laux, L., Glanzmann, P., Schaffner, P., & Spielberger, C. D. (1981). *STAI: Das State-Trait-Angstinventar* [STAI: The State-Trait Anxiety Inventory]. Weinheim: Beltz Testgesellschaft.

Lee, E., Xiang, Y.-T., Man, D., Au, R. W. C., Shum, D., Tang, W.-K., … Ungvari, G. S. (2010). Prospective memory deficits in patients with bipolar disorder: A preliminary study. *Archives of Clinical Neuropsychology, 25,* 640–647.

Livner, A., Berger, A.-K., Karlsson, S., & Bäckman, L. (2008). Differential effects of depressive symptoms on prospective and retrospective memory in old age. *Journal of Clinical and Experimental Neuropsychology, 30*, 272–279.

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

MacLeod, C., & Donnellan, A. M. (1993). Individual differences in anxiety and the restriction of working memory capacity. *Personality and Individual Differences, 15,* 163–173.

Mathews, A. & MacLeod, C. (1985). Selective processing of threat cues in anxiety states. *Behaviour Research and Therapy, 23,* 563–569.

McDaniel, M. A., & Einstein, G. O. (2000). Strategic and automatic processes in prospective memory retrieval: A multiprocess framework. *Applied Cognitive Psychology, 14,* S127–S144.

McDaniel, M. A., & Einstein, G. O. (2007). *Prospective memory: An overview and synthesis of an emerging field.* Thousand Oaks, CA: Sage Publications.

Overall, J. E. (1962). Dimensions of manifest depression. *Journal of Psychiatric Research, 1,* 239–245.

Pavawalla, S. P., Schmitter-Edgecombe, M., & Smith, R. E. (2012). Prospective memory after moderate-to-severe traumatic brain injury: A multinomial modeling approach. *Neuropsychology, 26*, 91–101.

Potts, R., Camp, C., & Coyne, C. (1989). The relationship between naturally occurring dysphoric moods, elaborative encoding, and recall performance. *Cognition and Emotion, 3,* 197–205.

Ramponi, C., Murphy, F. C., Calder, A. J., & Barnard, P. J. (2010). Recognition memory for pictorial material in subclinical depression. *Acta Psychologica, 135,* 293–301.

Rose, E. J., & Ebmeier, K. P. (2006). Pattern of impaired working memory during major depression. *Journal of Affective Disorders, 90,* 149–161.

Rose, N. S., Rendell, P. G., McDaniel, M. A., Aberle, I., & Kliegel, M. (2010). Age and individual differences in prospective memory during a "virtual week": The roles of working memory, vigilance, task regularity, and cue focality. *Psychology and Aging, 25*, 595–605.

Rude, S. S., Hertel, P. T., Jarrold, W., Covich, J., & Hedlund, S. (1999). Depression-related impairments in prospective memory. *Cognition and Emotion, 13*, 267–276.

Schnitzspahn, K. M., Horn, S. S., Bayen, U. J., & Kliegel, M. (2012). Age effects in emotional prospective memory: Cue valence differentially affects the prospective and retrospective component. *Psychology and Aging, 27*, 498–509.

Schuler, H., & Prochaska, M. (2001). *LMI: Leistungsmotivationsinventar* [AMI: Achievement Motivation Inventory]. Göttingen: Hogrefe Verlag.

Smith, J. B., & Batchelder, W. H. (2008). Assessing individual differences in categorical data. *Psychonomic Bulletin & Review, 15,* 713–731.

Smith, J. B., & Batchelder, W. H. (2010). Beta-MPT: Multinomial processing tree models for addressing individual differences. *Journal of Mathematical Psychology, 54*, 167–183.

Smith, R. E. (2003). 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, 29,* 347–361.

Smith, R. E., & Bayen, U. J. (2004). A multinomial model of event-based prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 756–777.

Smith, R. E., & Bayen, U. J. (2005). The effects of working memory resource availability on prospective memory: A formal modeling approach. *Experimental Psychology, 52*, 243–256.

Smith, R. E., & Bayen, U. J. (2006). The source of adult age differences in event-based prospective memory: A multinomial modeling approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 623–635.

Smith, R. E., Bayen, U. J., & Martin, C. (2010). The cognitive processes underlying event-based prospective memory in school-age children and young adults: A formal model-based study. *Developmental Psychology, 46*, 230–244.

Smith, R. E., Persyn, D., & Butler, P. (2011). Prospective memory, personality, and working memory: A formal modeling approach. *Journal of Psychology, 219*, 108–116.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *Manual for the State-Trait Anxiety Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Stout, D. M., & Rokke, P. D. (2010). Components of working memory predict symptoms of distress. *Cognition and Emotion, 24,* 1293–1303.

Sturtz, S., Ligges, U., & Gelman, A. (2005). R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software, 12*, 1–16.

Watts, F. N., Morris, L., & MacLeod, A. K. (1987). Recognition memory in depression. *Journal of Abnormal Psychology, 96*, 273–275.

Zigmond, A. S., & Snaith, R. P. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica, 67,* 361–370.

Table 1

Parameters of the Hierarchical Beta Distributions

	Mean [95% BCI]	*SD* [95% BCI]	α	β
C_1	.71 [.67 - .74]	.20 [.18 - .23]	2.82	1.17
C_2	.89 [.88 - .91]	.07 [.06 - .08]	16.55	1.96
P	.76 [.73 - .79]	.18 [.16 - .20]	3.58	1.10
M	.89 [.86 - .91]	.10 [.08 - .12]	8.24	1.10

Note. C_1 = probability of detecting a colour match; C_2 = probability of detecting a colour nonmatch; P = prospective component of PM; M = retrospective component of PM; BCI = Bayesian confidence interval; α and β are the parameters that describe the beta distribution.

Table 2

Scores on the BDI-II, HADS-D, and STAI

Test	*N*	*Mean*	*SD*	Range	
				Actual	Potential
BDI-II	128	9.64	7.07	0 - 33	0 - 63
STAI state	129	38.82	9.54	21 - 69	20 - 80
STAI trait	129	41.52	10.45	22 - 72	20 - 80
HADS-D anxiety	129	6.42	3.55	0 - 17	0 - 21
HADS-D depression	129	3.43	2.76	0 - 13	0 - 21

Note. BDI-II = Beck Depression Inventory II; STAI = State-Trait Anxiety Inventory; HADS-D = Hospital Anxiety and Depression Scale – German version.

Table 3

Correlations between Parameters M *and* P *of the Beta-MPT Model, Prospective-Memory Performance and Depression and Anxiety Scores*

Test	*N*		C_1	C_2	*P*	*M*	PM Hits
BDI-II	128	*r*	-.06	.04	-.08	.10	-.03
		p	.23	.32	.183	.143	.358
HADS-D	129	*r*	-.07	.04	-.04	.03	-.03
depression		*p*	.23	.32	.336	.37	.367
HADS-D	129	*r*	-.15*	-.05	-.01	-.01	-.01
anxiety		*p*	.05	.27	.469	.48	.458
STAI state	129	*r*	-.09	-.06	-.18*	-.07	-.19*
		p	.17	.25	.024	.213	.018
STAI trait	129	*r*	-.03	-.08	-.08	.05	-.06
		p	.36	.18	.177	.306	.265

Note. PM = prospective memory; C_1 = probability of detecting a colour-match; C_2 = probability of detecting a colour-nonmatch; *P* = prospective component of PM; *M* = retrospective component of PM; BDI-II = Beck Depression Inventory; HADS-D = Hospital Anxiety and Depression Scale – German version; STAI = State-Trait Anxiety Inventory.

* $p < .05$.

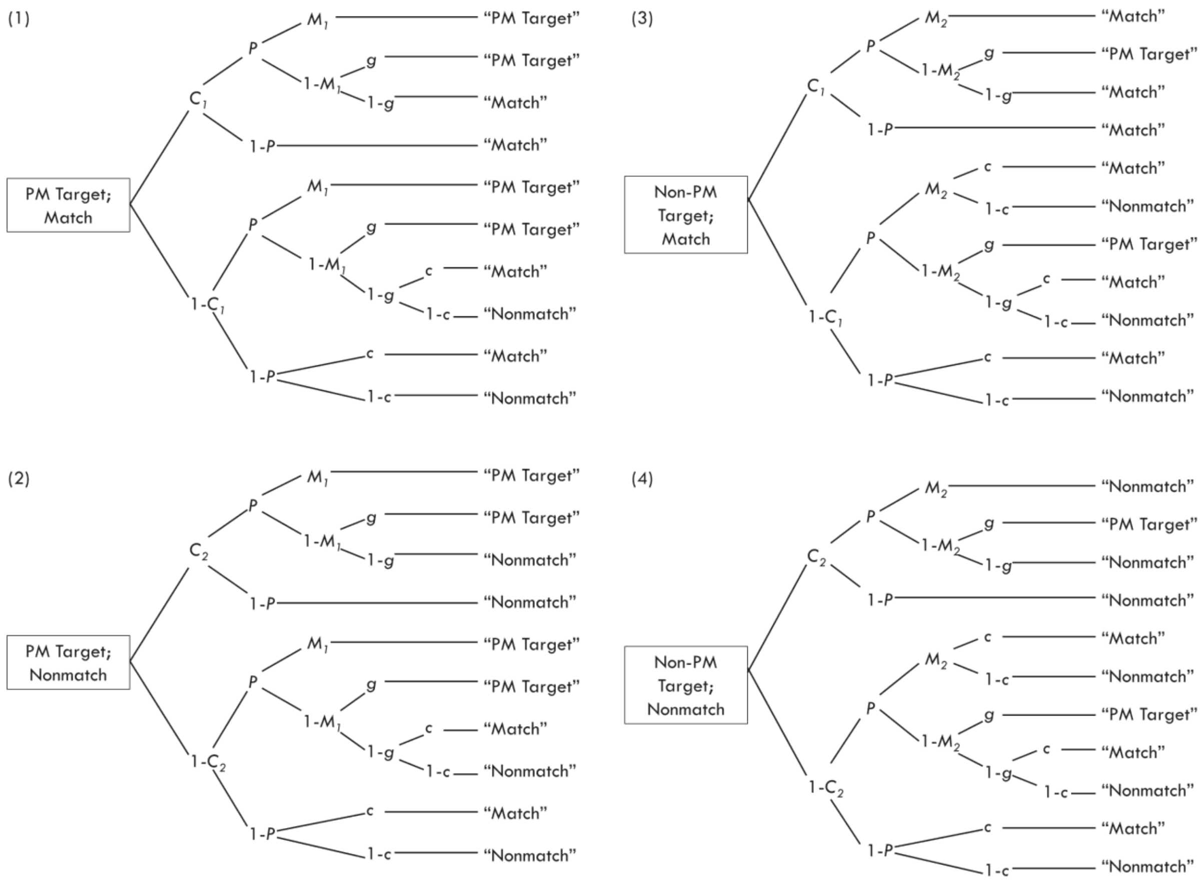
Figure Captions

Figure 1. Smith and Bayen's (2004) multinomial processing tree model of event-based prospective memory (PM). C_1 = probability of detecting a colour match; C_2 = probability of detecting that a colour does not match; P = prospective component of PM. M_1 = probability of recognizing a PM target; M_2 = probability of recognizing that a word is not a PM target; g = probability of guessing that a word is a PM target; c = probability of guessing that a colour matches. 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.

Footnote

[1] We report medians of the MCMC chains for the parameter estimations because they are less sensitive to outliers than means.

Figure 1


(1)
PM Target; Match
C1
1-C1
P
1-P
M1
1-M1
g
1-g
c
1-c
"PM Target"
"Match"
"Nonmatch"
(2)
PM Target; Nonmatch
C2
1-C2
(3)
Non-PM Target; Match
M2
1-M2
(4)
Non-PM Target; Nonmatch

Hierarchical multinomial modeling approaches:

An application to prospective memory and working memory

Nina R. Arnold

Heinrich-Heine-Universität Düsseldorf, Germany

Ute J. Bayen

Heinrich-Heine-Universität Düsseldorf, Germany

Rebekah E. Smith

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

Author Note

Nina R. Arnold and Ute Johanna Bayen, 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.

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). The research was supported in part by Grant AG034965 from the National Institute on Aging of the US National Institutes of Health. Nina R. Arnold is now at the Department of Psychology of the University of Mannheim.

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

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 three experiments investigating the relationship between PM and working memory (R. E. Smith & Bayen, 2005, Experiments 1 and 2; R. E. Smith, Persyn, & Butler, 2011) were reanalyzed using the two hierarchical modeling approaches, both of which allowed for the estimation of individual parameters for the components of PM. The results generally 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

Hierarchical multinomial modeling approaches:

An application to prospective memory and working memory

Multinomial processing tree (MPT) models are nonlinear statistical models that have been applied to many paradigms in cognitive psychology (for a review, see Erdfelder et al., 2009). They were designed to estimate probability parameters that measure latent cognitive processes from observed categorical data (Batchelder & Riefer, 1999). A MPT model is usually tailored to a specific research paradigm. In traditional MPT modeling, data are aggregated over participants and items, and observations are assumed to be independent and identically distributed (i.i.d.), ignoring differences between participants and items (Batchelder & Riefer, 1999; Hu & Batchelder, 1994; Knapp & Batchelder, 2004; Riefer & Batchelder, 1988). However, J. B. Smith and Batchelder (2008) showed that this assumption of parameter homogeneity is often violated - even for a carefully constructed item pool and a relatively homogeneous group of participants. This can result in biased parameter estimates (e.g., Klauer, 2006; J.B. Smith & Batchelder, 2008; J.B. Smith & Batchelder 2010). In most cases, the alternative of calculating a separate model for each participant is not a good option either, because some of the categories contain very few observations per participant.

There have been several recent attempts to deal with parameter heterogeneity in MPT modeling using either 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. This approach 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, on the other hand, provide individual parameter estimates for each participant by defining an individual MPT model for each participant. However, the individual parameter estimates for these individual MPT models are assumed to arise from a common distribution. Hence, there is a hyperdistribution for each model parameter that allows researchers to calculate group mean and variance for each parameter, as can be done for the parameter estimates obtained in traditional MPT modeling.

In addition to this benefit, hierarchical modeling techniques have another advantage that, as of yet, has been rarely focused on. Because hierarchical modeling allows for the estimation of individual parameters for each participant, these individual parameter estimates can then be correlated with variables from other tests or questionnaires. Thus, the use of hierarchical MPT models facilitates the investigation of the role of individual differences in explaining performance on cognitive tasks. In this article, we take advantage of this use of hierarchical modeling with the MPT model of event-based prospective memory (PM; Smith & Bayen, 2004). Specifically, we assessed PM experimentally and estimated individual model-based parameters measuring different components of PM that we then compared to individual measures of WM span.

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). The type of PM task referred to as event-based PM task (Einstein & McDaniel, 1990) involves performing the intended action when a certain event occurs; for example, remembering to mail a letter when you see a mailbox.

In everyday life, performing the intended action often involves interrupting a current activity. For example, you might have to remember to stop at the post office on your way to work. Therefore, in standard laboratory tests of PM, participants are also engaged in an ongoing task, for example a lexical decision task. The PM task is an additional task that must be performed when a certain target event occurs during the ongoing task (Einstein & McDaniel, 1990). For example, participants must press a special key when certain target syllables appear during the ongoing lexical decision task. The PM targets occur only infrequently. Einstein and McDaniel (1990) proposed that there are two components of PM, namely the prospective and the retrospective component. The prospective component is remembering *that* you have to do something. The retrospective-memory component is remembering *what* you wanted to do and *when* you wanted to do it. For successful PM performance, both components are necessary (Einstein & McDaniel, 1990). 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.

The goal of this article is to show the usefulness of hierarchical MPT modeling for research on individual differences in PM. We reanalyzed data from R. E. Smith, Persyn, and Butler (2011) and R. E. Smith and Bayen (2005) to examine how individual differences in working memory (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.

The MPT model of Event-Based Prospective Memory

R. E. Smith and Bayen (2004) introduced a MPT model of event-based PM in order to separate prospective and retrospective components of PM tasks. The model was designed to analyze data from standard laboratory event-based PM tasks, in which the PM task is embedded in an ongoing task, as described above. For example, in one of the studies we reanalyzed, participants performed a lexical-decision task as the ongoing task with the additional PM task of remembering to press a particular key when certain target syllables appeared.

The model (see Figure 1) is designed for a binary ongoing task, such as a lexical-decision task with the response options “word” and “non-word.” PM targets can occur on both options of the ongoing task; for example, a PM target can occur on word trials and non-word trials. This produces four types of trials, each represented by a separate tree in Figure 1: (1) a PM target occurs on a word trial, (2) a PM target occurs on a non-word trial, (3) a word trial occurs without a PM target, and (4) a non-word trial occurs 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 (i.e., 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 (i.e., second and fourth tree), C_2 is the probability that the participant recognizes that the string is not a word. In all trees, P is the probability that the participant remembers that there is an additional task (i.e., the prospective component). In all trees, M is the probability that a participant successfully discriminates 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 processes. If the participant is unable to discriminate between PM targets and non-targets, he or she must guess whether the string is a PM target. Parameter g denotes 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 the participant does not remember that there is a PM task (with probability $1 - P$), this will also result in an ongoing-task response. If a participant responds to the ongoing task but does not detect that a string is a word (with probability $1 - C_1$) or does not correctly detect that a string is a non-word (with probability $1 - C_2$), he or she can guess with probability c that the letter string is a word, and with probability $1 - c$ that the letter string is not a word.

The original model (R. E. Smith & Bayen, 2004) includes seven free parameters and is not identifiable. The parameters are, therefore, restricted based on theoretical assumptions (for a detailed description see R. E. Smith & Bayen, 2004). The resulting model has four free parameters P, M, C_1, and C_2, and is identifiable (as shown by R.E. Smith & Bayen, 2004). It has been validated (Horn, Bayen, R. E. Smith, & Boywitt, 2011; R. E. Smith & Bayen, 2004), and has been successfully applied in a number of studies (e.g., Pavawalla, Schmitter-Edgecombe, & R. E. Smith, 2012; Schnitzspahn, Horn, Bayen, & Kliegel, 2012; R. E. Smith & Bayen, 2006; R. E. Smith, Bayen, & Martin, 2010; R. E. Smith, Horn, & Bayen, 2012; R. E. Smith & Hunt, in press; R. E. Smith et al., 2011). The four-parameter model adequately fits the aggregated response frequencies of participant groups as reported by R. E. Smith and Bayen (2005) and R.E. Smith et al. (2011).

Hierarchical MPT Modeling

Hierarchical MPT models define a multinomial model for each participant and/or item. It is assumed that these individual parameters arise from a common distribution that may be described by hyperparameters. In 2010, J.B. Smith and Batchelder as well as Klauer proposed hierarchical models to deal with participant heterogeneity in MPT models. The two approaches differ mainly in the assumed underlying parameter distribution. J.B. Smith and Batchelder (2010) used a beta distribution, whereas Klauer (2010) used a transformed normal distribution.

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 (2010) described the beta-MPT approach to data analysis for the MPT model of pair clustering (Batchelder & Riefer, 1986) and provided an implementation for the software package WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000). We have demonstrated the usefulness of the beta-MPT approach in an application to source-monitoring data (Arnold, Bayen, Kuhlmann, & Vaterrodt, 2013). Here, we adapted this approach for the MPT model of PM.

The latent-trait approach (Klauer, 2010), on the other hand, 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 that it also accounts for correlations between the parameters of a model. Matzke et al. (in press) provided a WinBUGS implementation for the latent-trait version of the pair clustering model. We adapted this implementation for the MPT model of PM.

Both approaches rely on Bayesian modeling techniques. In Bayesian statistics, initial beliefs are represented by treating the parameters as random variables. A prior distribution is specified before starting the analysis. This can either be very vague or more concrete depending on prior knowledge about the parameter distributions. Then, 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 the prior distribution diminishes. One nice property that comes with Bayesian statistics is that they offer credible intervals, often called Bayesian confidence intervals (BCI). Classical frequentist confidence intervals are interpreted as random intervals that include, with a certain probability, the unknown model parameter that generated the data. BCI, by contrast, treat the parameter of interest as a random variable, and the interval is fixed. The interval reflects the range of values in the posterior distribution where the researcher has a certain confidence (given the data and the prior) that the true parameter will be found. Thus, we will present the BCI for our parameter estimates.

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 these draws, one can obtain knowledge about the properties of the model parameters. This analysis is possible without having to solve the integrals analytically, which can be computationally very expensive. MCMC chains can be computed using Gibbs sampling. The Gibbs sampler is a method of producing (serially dependent) draws from the target posterior distribution, under certain conditions, and is perhaps the most popular MCMC technique. It is implemented in a number of software packages (e.g., WinBUGS, Lunn, et al., 2000; OpenBugs, Lunn, Spiegelhalter, Thomas, & Best, 2009; JAGS, Plummer, 2003).

To draw statistical inference from MCMC chains, convergence (i.e., the chain has reached a stationary distribution) must be evaluated. Convergence can be assessed using the $\hat{R}$ statistic (Gelman & Rubin, 1992). This statistic is provided by several programs such as the R2WinBUGS package (Sturtz, Ligges, & Gelman, 2005) which may also be used as an interface between R and WinBUGS. The $\hat{R}$ statistic compares the variances within and between the chains. Under convergence, $\hat{R}$ will be close to 1. Often, early draws have poor convergence. Therefore, there usually is a burn-in period which is discarded and is not used for parameter estimates or convergence estimation. Successful implementation of an MCMC chain results in a sample from the full posterior distribution. This means that we can calculate statistics about all basic and hyperparameters. We can also estimate the uncertainty of the estimates by using BCI or the MCMC error. We describe our hierarchical analyses following a summary of previous findings and of the method and results reported by R.E. Smith et al. (2011).

Previous Findings on the Relationship of Prospective Memory and Working Memory

People differ in their ability to successfully perform PM tasks and it is possible that individual differences in WM may contribute to variability in PM performance, especially in the case of non-focal tasks, which have been shown to require more cognitive resources than focal tasks (e.g., Brewer, Knight, Marsh, & Unsworth, 2010; Einstein et al., 2005). A non-focal PM task is one in which the processing needed for the ongoing task does not require processing of the relevant features of the PM target. For example, if the ongoing task is a lexical-decision task, a word would be a focal PM target, whereas a syllable would be a non-focal PM target (as in the third experiment reanalyzed here). However, even focal tasks might be likely to show a relationship with working memory when the focal task encourages the use of non-automatic strategic processes, for instance when multiple targets are used (Einstein & McDaniel, 2010).

In several studies, a positive relationship between PM performance and WM span has been found (e.g., Ball, Knight, Dewitt, & Brewer, 2013; Brewer et al., 2010; Cherry & LeCompte, 1999; Einstein, McDaniel, Manzi, Cochran, & Baker, 2000; Reese & Cherry, 2002; R. E. Smith, 2003; R. E. Smith & Bayen, 2005; R. E. Smith et al., 2011; West & Craik, 2001). R. E. Smith and Bayen (2005), and R. E. Smith et al. (2011) used the MPT model of event-based PM (R.E. Smith & Bayen, 2004) described above to assess the relationship between components of a non-focal PM task and WM. The MPT approach is the ideal method to address this question because it separates the prospective component of PM from the retrospective component. In both studies, participants with higher WM span had a higher probability of engagement in the prospective component of PM. Differences in the retrospective component only emerged when an additional memory load was imposed by the ongoing task (see Experiment 2 of R. E. Smith & Bayen, 2005).

Although MPT modeling has the advantage of separating the prospective and the retrospective components of PM, the approach taken by R. E. Smith and Bayen (2005) and R.E. Smith et al. (2011) to examine the relationship between WM and PM components has its limitations. First, they used traditional MPT modeling (e.g., Batchelder & Riefer, 1988) in which parameter estimates are based on data aggregated over participants which may lead to biased estimates (e. g., Klauer, 2010, J. B. Smith & Batchelder, 2008, J. B. Smith & Batchelder, 2010). 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% of the participants with the lowest WM span scores, thus excluding half of the data from their analyses. We addressed these limitations by performing beta-MPT (J. B. Smith & Batchelder, 2010) and latent-trait (Klauer, 2010) re-analyses of the original data from R. E. Smith and Bayen (2005) and R. E. Smith et al. (2011). As discussed above, using these alternative approaches allows us to perform correlational analyses to examine the extent to which individual differences in PM components are responsible for the relationship between working memory span and PM performance. Importantly, this new approach also allows us to include the data from all participants.

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

Experiment 1 by R.E. Smith and Bayen (2005)

The participants were 20 young adults. The ongoing task was a sentence verification task. The PM task was to press the F1 key when one of four target words appeared. Although this is a focal PM task, because the words are processed during the course of the ongoing task, the use of multiple targets would encourage reliance on non-automatic processes, making it likely that the PM task will rely on working memory. Participants completed a counting span test (Conway, 1998) as a measure of WM span. R. E. Smith and Bayen´s (2005) analysis was based on data that were aggregated within WM span groups formed via median split. Participants in the higher-WM group had a greater probability of remembering that they had to perform the PM task, that is the prospective component (Parameter *P*), than did participants in the lower-WM group. However, the two WM 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 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. As a matter of consistency, we report Bayesian statistics for the correlations as well. For a discussion of the advantages of Bayesian testing over the classical frequentist null hypothesis testing, see, for example, Wagenmakers (2007). The Bayes factor (*BF*; e.g., Jeffreys, 1961; Wetzels & Wagenmakers, 2012) denotes the probability of the data under the null hypothesis relative to the alternative hypothesis. For our analyses, 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). 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. If the *BF* is smaller than 1/3, this indicates substantial evidence for the absence of a correlation (H_0).

For interested readers, we report "traditional" *p* values in addition to the *BF*s. We conducted a power analysis with G*Power (Faul, Erdfelder, & Buchner, 2007). With 20 participants and a conventional (one-tailed) α-level of .05, the power to detect a medium effect of $r = .30$ (e.g., Cohen, 1992) was $1\text{-}\beta = .39$, and the power to detect a large effect of $r = .50$ was $1\text{-}\beta = .80$. We computed correlations with WM span for the two relevant PM parameters, namely the prospective component *P* and the retrospective component *M*. We also report the correlation of WM span with the traditional measure of PM performance *PM hits*, that is the proportion of PM targets that were correctly responded to in Table 1. In addition, we report the parameter estimates for aggregated data estimated with the MultiTree program (Moshagen, 2010) in Table 2. For all tests, we used $\alpha = .05$.

Reanalysis with the beta-MPT approach. For the beta-MPT of event-based PM, 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 421.5. 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, $r = .400$, $p = .040$ (one-tailed), and between WM span and the prospective component *P*, $r = .394$, $p = .043$ (one-tailed). The Bayes factors (*BF*) were 0.779 and 0.741 not indicating decisive evidence to support or reject a correlation. WM span and the retrospective-memory component *M* did not correlate significantly, and the $BF = 0.237$ indicated support for the absence of a correlation. Thus, at least with the non-Bayesian measures, 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*.

Reanalysis with 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., in press; 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 376.0. We also computed correlations with the resulting individual parameter estimates of the latent-trait approach, shown 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*, $r = .406$, $p = .038$. The *BF* is 0.819, indicating neither evidence for the alternative nor for the null hypothesis. WM span and the retrospective-memory component *M* did not correlate, $r = .168$, $p = .239$, $BF = 0.219$.

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 *BF*s < 2/3. No approach showed consistently larger correlations than the other approach (see Table 3). As shown in Table 3, the parameter estimates obtained with both methods showed very high correlations, all $r > .98$, all $p <.001$, all $BF > 100$. In terms of *DIC*, the latent-trait approach fit the model better than the beta-MPT approach.

Experiment 2 by R.E. Smith and Bayen (2005)

The participants were 21 young adults. WM span task and PM task were the same as in Experiment 1. As the ongoing task, participants again performed a sentence verification task. Additionally, however, after every fourth sentence, they had to report the last word of each of the four sentences. This was done to increase the memory load of the ongoing task. As in Experiment 1, R. E. Smith and Bayen´s (2005) analysis was based on data that were aggregated within WM span groups formed via median split. WM span affected all parameter estimates. Therefore, for the current analyses, we hypothesized that the *P* parameter and the *M* parameter correlated with WM span scores.

Reanalysis with Hierarchical MPT Models

All analyses for Experiment 2 were conducted exactly like those for Experiment 1.

Reanalysis with the beta-MPT approach. *DIC* for the beta-MPT approach was 474.8. The posterior distributions are shown in Table 2. The correlations along with means and standard deviations of the WM span measures are shown in Table 1. Contrary to predictions, we did not find a significant correlation of PM hits and the PM components with

WM span, all $p > .05$ (one-tailed), all BFs < 0.60. However, BFs were not small enough for evidence in favor of the null hypothesis.

Reanalysis with the latent-trait approach. *DIC* for the beta-MPT approach was 435.1. The posterior distributions are shown in Table 2. Concerning the correlations with WM span (shown in the lower half of Table 1), we found the same pattern as with the beta-MPT approach. That is, we did not find a significant correlation of the PM components with WM span, all $p > .05$ (one-tailed), all BFs < 0.60.

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. Again, estimates were 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 do not differ significantly, all $p > .05$, and all *BF*s < 2/3. No approach showed consistently larger correlations than the other approach (see Table 1). As shown in Table 4, the parameter estimates obtained with both methods were very highly correlated, all $r > .97$, all $p < .001$, all $BF > 100$. Again, in terms of *DIC*, 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 *BCI*s overlapped for all parameters. There was no advantage of either approach.

The results of our reanalyses concur with the findings reported by R. E. Smith and Bayen (2005) only for Experiment 1. For Experiment 2, we did not find correlations of model parameters with WM span. However, the sample size was very small, limiting the statistical power to detect a medium size effect to less than .50. Likewise, for the Bayesian approach, we did not have enough information to report evidence in favor for either the absence or the presence of a correlation (except for the correlation of the parameter estimates obtained with both methods). To yield higher power, we next reanalyzed the study by R. E. Smith et al. (2011) with a sample size of 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)

The Study by R.E. Smith et al. (2011)

Participants were 413 young adults. The ongoing task was a lexical decision task. The PM task was to press the F1 key when the syllables "low" or "per" appeared. This is a non-focal PM task, because the detection of certain syllables requires different processes than a lexical decision task. In addition, participants completed a symmetry span test (Unsworth, Redick, Heitz, Broadway, & Randall, 2009) as a measure of WM span. Of the participants, only those with the lowest 25% of the span scores and those with the highest 25% of the span scores were included in the extreme groups analysis.

R. E. Smith et al.'s (2011) analyses were based on data that were aggregated within these extreme WM span groups. 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 two WM groups did not differ in the retrospective-memory component (Parameter *M*). For our analyses, which included data from all 413 participants, we hypothesized that the *P* parameter would correlate 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 (Table 1). Again, we report *p* values as well as Bayes factors In Table 2, we report parameter estimates obtained with the traditional approach based on aggregated data.

Reanalysis with the beta-MPT approach. For the beta-MPT of event-based PM, we used a uniform distribution between 1 and 5000 as prior for each α_s and β_s of each parameter Θ_s which ensures that the beta distribution is bell shaped. Due to slower conversion, 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, $\hat{R} < 1.05$. The posterior distributions of the hierarchical beta distributions are shown in Table 2.

DIC was 8553.17. The correlations along with means and standard deviations of WM span are shown in Table 1. We found a significant correlation between WM span and PM hits, $r = .114$, $p = .010$. However, the $BF = 0.572$ did not indicate substantial support for either the presence or the absence of a correlation. We also found a significant correlation between WM span and the prospective component *P*, $r = .148$, $p = .003$. The *BF* was 3.672 indicating substantial support for a correlation. WM span and the retrospective-memory component *M* did not correlate significantly, and the *BF* indicated support for the absence of a correlation. Thus, our results concur with those by 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*.

Reanalysis with 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. 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, $\hat{R} < 1.05$. The posterior distributions are shown in Table 2.

DIC was 8590.94. We also computed correlations with the resulting individual parameter estimates of the latent-trait approach, shown 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*, $r = .143$, $p = .003$. The *BF* is 2.707, indicating only anecdotal evidence for the presence of a correlation. WM span and the retrospective-memory component *M* did not correlate, $r = .041$, $p = .404$, $BF = 0.055$.

Comparison of the approaches. We compared the parameter estimates obtained with both 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 $p > .05$, and all *BF*s < 1/8. Only the retrospective component *M* shows higher beta-MPT estimates than latent-trait estimates, $t(412) = 7.473$, $p < .001$, $BF > 1{,}000{,}000$. 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). As shown in Table 5, the correlations between the parameters obtained with both methods were very high, all $r > .91$, all $p <.001$, all $BF > 100$. 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.

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 the data by R. E. Smith and Bayen (2005). This time, however, the *BCI*s did not overlap. The results concur with the original findings (Smith et al., 2011) with the exception that the *BF* for the prospective component *P* estimated via the latent-trait approach did not show decisive evidence for a correlation.

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. Except for Experiment 2 of R. E. Smith and Bayen (2005), WM span correlated with the prospective component of PM, but not with the retrospective-memory component. However, sample sizes in the studies by R. E. Smith and Bayen (2005) were very small leading to limited power to detect a medium effect sized correlation.

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, thus, it is possible to put a strong prior on the correlations if there are reasons to expect a correlation of a specific magnitude. However, in our case, we had a uniform prior on the parameter correlations assuming that all values for the correlation were equally likely. 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. 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 mentioned by Matzke et al. (2013) is that two cognitive abilities that both reflect aspects of memory retrieval are likely to be related. The possibility to incorporate between-parameter correlations, thus, speaks in favor of the latent-trait approach.

PM tasks usually have only few targets, and error rates are often low. Therefore, a general problem of parameter estimation for the MPT model of event-based PM are low cell 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 as a goodness-of-fit statistic, 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.

References

Arnold, N. R., Bayen, U. J., Kuhlmann, B. G., & Vaterrodt, B. (2013). Hierarchical modeling of contingency-based source monitoring: A test of the probability-matching account, *Psychonomic Bulletin & Review, 20,* 326-333.

Ball, B. H., Knight, J. B., Dewitt, M. R., & Brewer, G. A. (in press). Individual differences in the delayed execution of prospective memories. *The Quarterly Journal of Experimental Psychology.*

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

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

Brewer, G. A., Knight, J. B., Marsh, R. L., & Unsworth, N. (2010). Individual differences in event-based prospective memory: Evidence for multiple processes supporting cue detection. *Memory & Cognition, 38*, 304–311.

Cherry, K. E., & LeCompte, D. C. (1999). Age and individual differences influence prospective memory. *Psychology and Aging, 14,* 60-76.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155-159.

Conway, A. R. A (1998). Counting Span Task [Computer program]. Chicago, IL: University of Illinois at Chicago.

Einstein, G. O., & McDaniel, M. A. (1990). Normal aging and prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 717–726.

Einstein, G. O., & McDaniel, M. A. (2010). 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, 36.* 1082-1088.

Einstein, G. O., McDaniel, M. A., Manzi, M., Cochran, B., & Baker, M. (2000). Prospective memory and aging: Forgetting intentions over short delays. *Psychology and Aging, 15,* 671-683.

Einstein, G. O., McDaniel, M. A., Thomas, R. A., Mayfield, S., Shank, H., Morrisette, N., & Breneiser, J. (2005). Multiple processes in prospective memory retrieval: Factors determining monitoring versus spontaneous retrieval. *Journal of Experimental Psychology: General, 134,* 327-342.

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Journal of Psychology, 217*, 108–124.

Faul, F., Erdfelder, E., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175-191.

Gelman, A., & Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7 ,* 457–472.

Hays, W. L. (1994). *Statistics* (5th ed.). Fort Worth, TX: Harcourt Brace College Publishers.

Horn, S. S., Bayen, U. J., Smith, R. E., & Boywitt, C. D. (2011). The multinomial model of prospective memory: Validity of ongoing-task parameters. *Experimental Psychology*, *58*, 247-255.

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

Jeffreys, H. (1961). *Theory of probability.* Oxford: UK Oxford University Press.

Klauer, K. C. (2006). Hierarchical multinomial processing tree models: A latent-class approach. *Psychometrika, 71*, 7–31.

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika. 75,* 70-98.

Knapp, B. R., & Batchelder, W. H. (2004). Representing parametric order constraints in multi-trial applications of multinomial processing tree models. *Journal of Mathematical Psychology*, *15*, 215–229.

Liang, F., Paulo, R., Molina, G., Clyde, M., & Berger, J. (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association, 103,* 410.

Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The BUGS project: Evolution, critique and future directions. *Statistics in Medicine, 28,* 3049–3067.

Lunn, D., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS – A Bayesian modeling framework: Concepts, structure, and extensibility. *Statistics and Computing, 10,* 325–337.Matzke, D., Dolan, C.V, Batchelder, W.H., & Wagenmakers, E.-J. (in press). Bayesian estimation of multinomial processing tree models with heterogeneity in participants and items. *Psychometrika: Application Reviews & Case Studies.*

McDaniel, M. A., & Einstein, G. O. (2007). *Prospective memory: An overview and synthesis of an emerging field.* Thousand Oaks, CA: Sage.

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods, 42,* 42-54.

Pavawalla, S. P., Schmitter-Edgecombe, M., & Smith, R. E. (2012). Prospective memory following moderate-to-severe traumatic brain injury: A multinomial modeling approach. *Neuropsychology, 26,* 91-101.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling [Computer software manual]. Retrieved from http://citeseer.ist.psu.edu/plummer03jags.html

Reese, C. M., & Cherry, K. E. (2002). The effects of age, ability, and memory monitoring on prospective memory task performance. *Aging, Neuropsychology, and Cognition, 9,* 98-113.

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

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

Schnitzspahn, K. M., Horn, S. , Bayen, U. J., & Kliegel, M. (2012). Age effects in emotional prospective memory: Cue valence differentially affects the prospective and retrospective component. *Psychology and Aging, 27,* 498-509.

Smith, J. B., & Batchelder, W.H. (2008). Assessing individual differences in categorical data. *Psychonomic Bulletin & Review, 15,* 713–731.

Smith, J. B., & Batchelder, W. H. (2010). Beta-MPT: Multinomial processing tree models for addressing individual differences. *Journal of Mathematical Psychology, 54*, 167–183.

Smith, R. E. (2003). 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, 29*, 347-361.

Smith, R. E., Bayen, U. J., & Martin, C. (2010). The cognitive processes underlying event-based prospective memory in school age children and young adults: A formal model-based study. *Developmental Psychology, 46,* 230-244.

Smith, R. E., & Bayen, U. J. (2004). A multinomial model of event-based prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 756-777.

Smith, R. E., & Bayen, U. J. (2005). The effects of working memory resource availability on prospective memory: A formal modeling approach. *Experimental Psychology, 52,* 243-256.

Smith, R. E., & Bayen, U. J. (2006). The source of adult age differences in prospective memory: A multinomial modeling approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 623-635.

Smith, R. E., Horn, S. S., & Bayen, U. J. (2012). Prospective memory in young and older adults: The effects of ongoing task load. *Aging, Neuropsychology, and Cognition, 19,* 495-514.

Smith, R. E., & Hunt, R. R. (in press). Prospective memory in young and older adults: The effects of task importance and ongoing task load. *Aging, Neuropsychology, and Cognition.*

Smith, R. E., Persyn, D., & Butler, P. (2011). Prospective memory, personality, and working memory: A formal modeling approach. *Zeitschrift für Psychologie / Journal of Psychology, 219,* 108-116.Sturtz, S., Ligges, U., & Gelman, A. (2005). R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software, 12*, 1-16.

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory, 17,* 635 - 654.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review, 14,* 779-804.

West, R., & Craik, F. I. M. (2001). Influences on the efficiency of prospective memory in younger and older adults. *Psychology and Aging, 16,* 682-696.

Wetzels, R., & Wagenmakers, E.-J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review, 19,* 1057–1064.

Footnote

[1] Note 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.

Table 1

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

		Smith & Bayen (2005) Exp1	Smith & Bayen (2005) Exp2	Smith, Persyn, & Butler et al. (2011)
Working Memory	*Mean*	22.15	20.43	28.67
Span	*SD*	10.82	11.69	7.39
PM Hits	*Mean*	.78	.62	.55
	SD	.21	.30	.33
	r	.40	.29	.11
	p	.04	.10	.02
	BF	0.78	0.38	0.57
Beta-MPT Prospective	*Mean*	.81	.64	.64
Component *P*	*SD*	.15	.26	.01
	r	.39	.28	.15
	p	.04	.11	< .01
	BF	0.74	0.34	3.67
Beta-MPT Retrospective	*Mean*	.96	.96	.75
Component *M*	*SD*	.02	.01	.01
	r	.19	.35	.01
	p	.21	.06	.92
	BF	0.24	0.56	0.04

Latent-Trait Prospective	*Mean*	.82	.65	.72
Component *P*	*SD*	.17	.29	.02
	r	.41	.28	.143
	p	.04	.11	< .01
	BF	0.82	0.34	2.71
Latent-Trait	*Mean*	.97	.96	.82
Retrospective Component	*SD*	.05	.03	.02
M	*r*	.17	.34	.041
	p	.24	.07	.40
	BF	0.22	0.52	0.06

Note. PM Hits = rate of PM targets correctly responded to; *P* = prospective component of PM; *M* = retrospective component of PM; 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.

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

		Beta-MPT		*Latent-Trait*		*MultiTree*	
		M	[95 % *BCI*]	*M*	[95 % *BCI*]	*M*	[95 % *CI*]
	P	.79	[.70 - .86]	.84	[.73 - .94]	.81	[.76 - .85]
2005	*M*	.96	[.92 - .98]	.98	[.94 - 1]	.97	[.94 - .99]
Exp 1	C_1	.90	[.85 - .93]	.92	[.87 - .95]	.90	[.88 - .93]
	C_2	.77	[.71 - .82]	.78	[.72 - .84]	.78	[.74 - .81]
	P	.61	[.50 - .71]	.70	[.46 - .88]	.65	[.59 - .70]
2005	*M*	.95	[.91 - .98]	.97	[.92 - 1]	.96	[.93 - .99]
Exp 2	C_1	.81	[.75 - .85]	.82	[.77 - .87]	.81	[.78 - .84]
	C_2	.77	[.70 - .83]	.76	[.72 - .86]	.78	[.74 - .81]
	P	.64	[.61 - .66]	.72	[.68 - .76]	.67	[.66 - .69]
	M	.75	[.72 - .77]	.82	[.78 - .86]	.80	[.79 - .82]
2011	C_1	.94	[.93 - .94]	.95	[.95 - .96]	.94	[.94 - .94]
	C_2	.95	[.94 - .95]	.96	[.96 - .97]	.95	[.94 - .95]

Note. 2005 Exp 1 = R. E. Smith and Bayen (2005), Experiment 1; 2005 Exp 2 = R. E. Smith and Bayen (2005), Experiment 2; 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, or that a sentence is true in sentence verification; C_2 = probability to detect that a letter string is a non-word, or that a sentence is false in sentence verification; BCI = Bayesian confidence interval; CI = (traditional) confidence interval.

Table 3

Correlation Coefficients of the Individual Level Parameters of the Beta-MPT and the Latent-Trait Approach, in the Reanalysis of Experiment 1 of Smith and Bayen (2005)

		Beta-MPT				*latent-trait approach*			
		P	M	C_1	C_2	P	M	C_1	C_2
Beta-MPT	P	-	.267	.512	.086	.991	.180	.588	.200
	M		-	-.122	.103	.178	.987	-.066	.077
	C_1			-	.592	.612	-.239	.982	.721
	C_2				-	.144	.080	.667	.985
Latent-trait approach	P					-	.081	.677	.269
	M						-	-.175	.031
	C_1							-	.783
	C_2								-

Note. P = prospective component of PM; M = retrospective-memory component of PM; C_1 = probability to detect that a sentence is true in sentence verification; C_2 = probability to detect that a sentence is false in sentence verification.

Table 4

Correlation Coefficients of the Individual Level Parameters of the Beta-MPT and the Latent-Trait Approach in the Reanalysis of Experiment 2 of Smith and Bayen (2005)

		Beta-MPT				*latent-trait approach*			
		P	M	C_1	C_2	P	M	C_1	C_2
Beta-MPT	P	-	.135	.302	.486	.998	.063	.420	.580
	M		-	-.050	.042	.095	.990	-.036	.057
	C_1			-	.544	.330	-102	.978	.642
	C_2				-	.527	-.003	.700	.985
Latent-trait approach	P					-	.023	.452	.618
	M						-	-.093	.001
	C_1							-	.786
	C_2								-

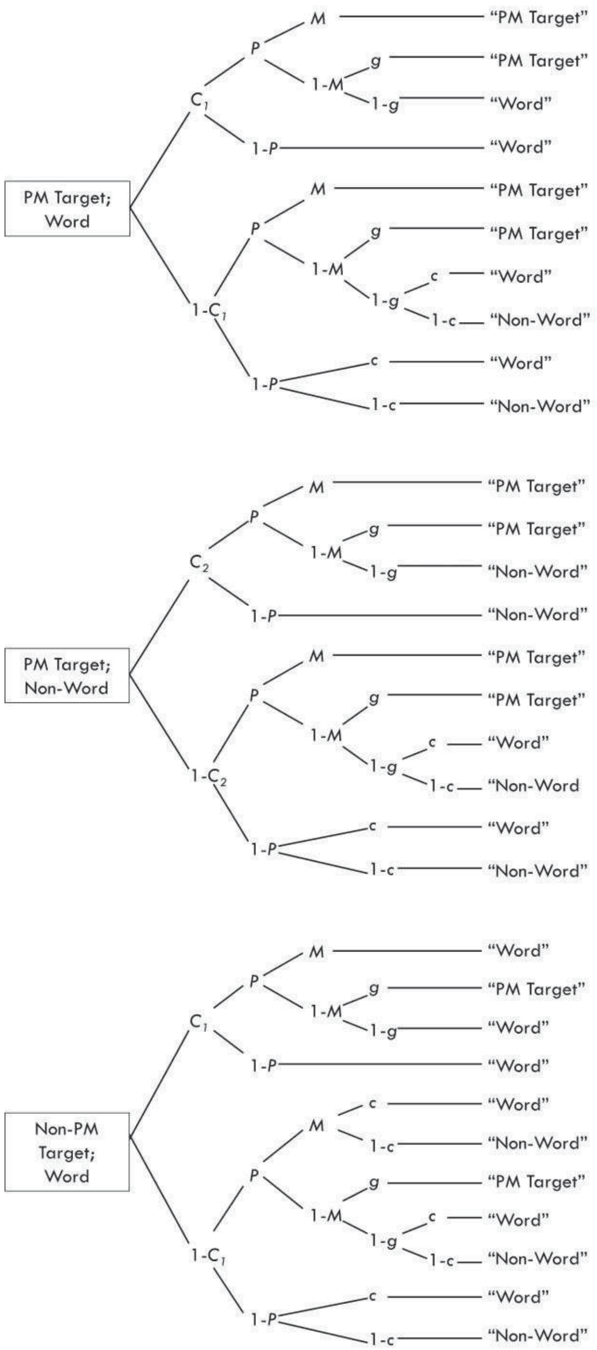
Note. P = prospective component of PM; M = retrospective-memory component of PM; C_1 = probability to detect that a sentence is true in sentence verification; C_2 = probability to detect that a sentence is false in sentence verification.

Table 5

Correlation Coefficients of the Individual Level Parameters of the Beta-MPT and the Latent-Trait Approach in the Reanalysis of Smith, Persyn, and Butler (2011)

		Beta-MPT				*latent-trait approach*			
		P	M	C_1	C_2	P	M	C_1	C_2
Beta-MPT	P	-	.476	.060	.275	.997	.740	.090	.362
	M		-	.150	.261	.531	.898	.200	.331
	C_1			-	.172	.063	.117	.992	.230
	C_2				-	.316	.301	.258	.991
Latent-trait approach	P					-	.783	.099	.402
	M						-	.169	.389
	C_1							-	.314
	C_2								-

Note. P = prospective component of PM; M = retrospective-memory component of PM; C_1 = probability to detect that a letter string is a word; C_2 = probability to detect that a letter string is a non-word.


PM Target; Word
C1
1-C1
P
1-P
M
1-M
g
1-g
c
1-c
"PM Target"
"Word"
"Non-Word"
PM Target; Non-Word
C2
1-C2
Non-PM Target; Word

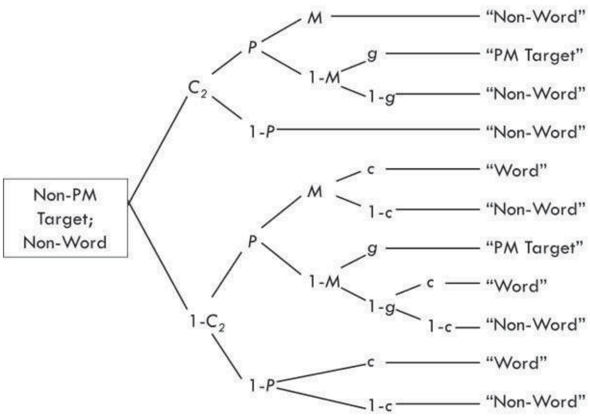


Figure 1. Multinomial model of event-based prospective memory. PM = prospective memory; C_1 = probability of detecting that a letter string is a word in lexical decision (or that a sentence is true in sentence verification); C_2 = probability of detecting that a letter string is a non-word in lexical 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).