Hierarchical Bayesian measurement models for continuous reproduction of visual features from working memory

Klaus Oberauer University of Zurich, Zurich, Switzerland

Colin Stoneking Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

Dominik Wabersich University of Tübingen, Tübingen, Germany

Hsuan-Yu Lin University of Zurich, Zurich, Switzerland

The article presents Bayesian hierarchical modeling frameworks for two measurement models for visual working memory. The models can be applied to the distributions of responses on a circular feature dimension, as obtained with the continuous reproduction (a.k.a. delayed estimation) task. The first measurement model is a mixture model that describes the response distributions as a mixture of one (Zhang & Luck, 2008) or several (Bays, Catalao, & Husain, 2009) von-Mises distribution(s) and a uniform distribution. The second model is a novel, interference-based measurement model. We present parameter recovery simulations for both models, demonstrating that the hierarchical framework enables precise parameter estimates when a small number of trials are compensated by a large number of subjects. Simulations with the mixture model show that the Bayesian hierarchical framework minimizes the previously observed estimation bias for memory precision in conditions of low performance. Unbiased and reasonably precise parameter estimates can also be obtained from the interference measurement model, though some parameters of this model demand a relatively large amount of data for precise measurement. Both models are applied to two experimental data sets. Experiment 1 measures the effect of memory set size on the model parameters. Experiment 2 provides evidence for the assumption in the interference model that the target feature tends to be confused with features of those nontargets that are close to the target on the dimension used as retrieval cue.

Introduction

The continuous-reproduction (a.k.a. delayed estimation) task (Prinzmetal, Amiri, Allen, & Edwards, 1998; Wilken & Ma, 2004) is a popular procedure for measuring and investigating visual working memory (VWM). Its popularity is due in large part to the availability of a simple measurement model initially proposed by (Zhang & Luck, 2008) and later extended by (Bays et al., 2009). This measurement model is known as the mixture model because it describes the error distribution as a mixture of two (Zhang & Luck, 2008) or three (Bays et al., 2009) components. Here we develop a hierarchical Bayesian framework for the mixture model, together with a method for applying it efficiently. In addition, we introduce an alternative measurement model, the interference measurement model (IMM), and present a hierarchical Bayesian framework for the IMM.

In the continuous-reproduction task, participants encode a set of visual objects into VWM, and are tested on one randomly selected object, the target. They are asked to reproduce the target's memory feature on a continuous circular response scale. For instance, the memory set often consists of an array of simultaneously presented color patches, of which one is identified as the target by its spatial location, and participants reproduce its color by clicking on that color on a color wheel (Wilken & Ma, 2004). Other applications involve the reproduction of the target's orientation by adjustment of a dial (Bays, Gorgoraptis, Wee, Marshall, & Husain, 2011). Because responses in this task vary on a continuous—usually circular—dimension, the precision of memory for an object's feature can be measured as a

Citation: Oberauer, K., Stoneking, C., Wabersich, D., & Lin, H.-Y. (2017). Hierarchical Bayesian measurement models for continuous reproduction of visual features from working memory. *Journal of Vision*, *17*(5):11, 1–27, doi:10.1167/17.5.11.

doi: 10.1167/17.5.11 Received October 29, 2016; published May 24, 2017 ISSN 1534-7362

distribution of the deviation of responses from the true target feature.

Measurement models

This article is concerned with psychological measurement models for VWM. A measurement model is a model that enables researchers to go beyond merely descriptive summary statistics of the data—such as the variance of responses, or the absolute mean error—and measure memory performance in terms of theoretically interpretable model parameters. This is possible because the model incorporates theoretical assumptions about memory, and thereby, the model parameters are latent variables of a memory model. In this regard psychological measurement models differ from statistical models, which usually have no interpretation in terms of hypothetical psychological mechanisms or processes. Measurement models also differ from explanatory models, which aim to explain the effects of experimental manipulations, such as manipulations of memory set size, encoding duration, or retention interval. Measurement models are applied to each experimental condition separately with the aim to measure one or several latent variables in each condition, and to evaluate the effect of the manipulation on the latent variable. In contrast, explanatory models aim to reproduce the experimental effect with a common set of parameters across all conditions. Here we are concerned only with measurement models, and therefore will not discuss explanatory models of VWM (for an overview of explanatory models applied to the continuous-reproduction task, see van den Berg, Awh, & Ma, 2014), but the modelling techniques we use could also be used to apply explanatory models in a hierarchical Bayesian fashion.

We next introduce two measurement models for continuous reproduction, the mixture model, and a new interference measurement model.

The mixture model

The two-component mixture model (Zhang & Luck, 2008) models the distribution of responses as a probabilistic mixture of two components. The first component is a von-Mises distribution (i.e., an approximation of a circular normal distribution) centered on the target's memory feature in each trial; the second component is a uniform distribution on the circle. The mixture model incorporates the assumption that an attempt to access VWM results in one of two discrete states: Either the target can be accessed in VWM, in which case it can be reproduced with a degree of precision given by the precision parameter of the von-Mises distribution, or it cannot be accessed at all, in which case the person can only guess at random, modeled as a uniform distribution. Hence, the model has two free parameters, the probability of a response coming from memory, *Pmem*, and the precision of the von-Mises, κ (higher precision implying a smaller standard deviation).

The three-component mixture model (Bays et al., 2009) adds a third component, capturing responses that reflect representations of nontarget objects in the memory set. The probability of a response reflecting a nontarget is given by parameter *Pnt*. For a memory set size of n, this component is the sum of $n-1$ von-Mises distributions centered on the $n-1$ nontarget features, respectively. For simplicity, the precision of these distributions is assumed to be the same as that of the target-related distribution, so that the model has three parameters: *Pmem*, *Pnt*, and κ.

The likelihood function of the three-parameter mixture model is given by:

$$P(\hat{x}|\mathbf{X}, Pmem, Pnt, \kappa) = Pmem \cdot VM(\hat{x}; x_1, \kappa) + Pnt\frac{\sum_{i=2}^{n} VM(\hat{x}; x_i, \kappa)}{n-1} + (1 - Pmem - Pnt)\frac{1}{2\pi}.$$

Here, $\hat{x}$ is the response, $\mathbf{X}$ is the vector of memory features in the array, with x_1 as the target and x_2 to x_n as the nontargets, and VM is the von-Mises distribution:

$$VM(x; x_i, \kappa) = \frac{\exp(\kappa \cdot \cos(x - x_i))}{2\pi \cdot I_0(\kappa)}$$

where $I_0(\kappa)$ is the modified Bessel function of order 0. The two-parameter model is a special case of the three-parameter model with *Pnt* fixed to 0.

The theoretical assumption underlying the mixture model is that of discrete memory states at test: The person either retrieves the target from memory, or retrieves one of the nontargets, or retrieves nothing at all and therefore is forced to guess at random. The parameters *Pmem* and *Pnt* jointly determine the probabilities of each of these three states occurring for a person in a given experimental condition. The precision parameter κ reflects the precision of a feature represented in VWM, given that the feature is retrievable at all.[1] Proponents of discrete-capacity models of VWM (Cowan, 2005; Fukuda, Awh, & Vogel, 2010; Zhang & Luck, 2008) can use *Pmem* to estimate the number of "slots", K, that characterizes a person's VWM capacity: $K = N \times Pmem$ (for $N > K$).

The price to pay for the measurement of theoretically interpretable latent variables is that there can be

disagreement about the theoretical assumptions underlying a measurement model. The assumption of discrete states of representations in VWM, or of discrete states at access, is not universally accepted (Ma, Husain, & Bays, 2014; Suchow, Fougnie, Brady, & Alvarez, 2014; van den Berg et al., 2014). For researchers who don't endorse a discrete-state view of VWM, the mixture-model parameters do not represent anything of interest. Therefore, it is important to also develop alternative measurement models that do not rely on the discrete-state assumption. We next introduce an interference-based measurement model, the IMM, which is based on the assumption that the information retrieved from memory varies on a continuous dimension of strength. The IMM shares the assumption of continuously varying memory strength with signal-detection models of VWM (Wilken & Ma, 2004) and of memory in general (Wixted, 2007). The IMM is a measurement model derived from our interference-based explanatory model of VWM (Oberauer & Lin, 2017).

The interference measurement model

The IMM incorporates core assumptions of cue-based retrieval theories of memory (Surprenant & Neath, 2009): Information is reactivated at retrieval to the extent that it is associated to the retrieval cues available, and attended to, at test. The probability of retrieving a given piece of information depends on its relative activation at test. In the continuous-reproduction paradigm, the main retrieval cue is the stimulus that identifies the target at test—in most experiments that is the target's spatial location in the array. Therefore, the key to performance is the binding between each object's memory feature (i.e., the feature to be reproduced) and its cue feature (i.e., the feature that serves as retrieval cue). These bindings have a limited precision on the memory-feature dimension and on the cue-feature dimension (see Figure 1). The limited precision on the cue-feature dimension implies that the target's cue feature (e.g., its location) acts as a retrieval cue not only for the target but also for nontargets to the extent that they are close to the target on the cue-feature dimension (e.g., objects spatially close to the target in an array). Thus, for example, when probed to recall an object based on its location in the array, participants would tend to recall the color of the object in that location, or the color of other objects that had been presented in nearby locations. In the IMM, the precision of bindings along the cue-feature dimension is modeled as an exponential generalization gradient along the cue-feature dimension (Brown, Neath, & Chater, 2007; Nosofsky, 1984).

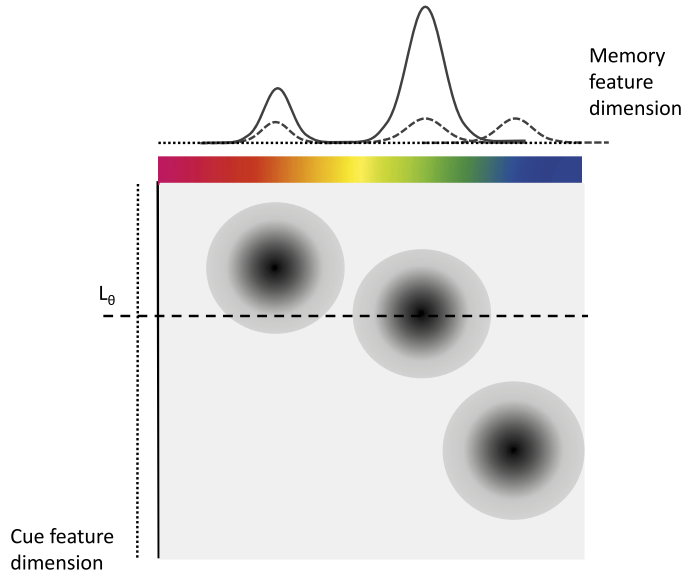

Figure 1. Schematic of the bindings between memory features (e.g., colors) and cue-features (e.g., spatial location) according to the interference measurement model. The Figure shows the state of the memory system at test. Each memory feature is bound to its location on the cue-feature dimension through a binding with limited precision in both dimensions, shown here as a two-dimensional distribution in binding space. The target location $L\theta$ acts as a retrieval cue, reactivating all features in proportion to the strength of their binding to $L\theta$, depicted by the darkness level in binding space. The resulting distribution of reactivation is added to the cue-independent activation of the three features, and uniform background noise, to form the distribution of activation $A(x)$ over the memory-feature dimension. Adapted from Oberauer, K., & Lin, H.-Y. (2017). An interference model of visual working memory. *Psychological Review, 124,* 21–59.

Cue-based retrieval results in a distribution of activation over all possible retrieval candidates. The degree of activation reflects the strength of evidence from memory for selecting a candidate for retrieval. On a continuous response scale the activation is distributed over the entire range of the response scale.

The activation distribution over response options x generated by a retrieval cue at the target location L_θ in cue-feature space is given by:

$$A_c(x|L_\theta) = \sum_{i=1}^{n} \exp[-s \cdot D(L_i, L_\theta)] \cdot VM(x, x_i, \kappa),$$

with $D(L_i, L_\theta)$ for the distance between the cued target location L_θ and the location L_i of an array object i, and s, a free parameter for the generalization gradient on the cue-feature dimension reflecting the precision of bindings on that dimension.

The IMM includes two further components to the activation distribution at test: Component A_a reflects cue-independent activation of the memory features of

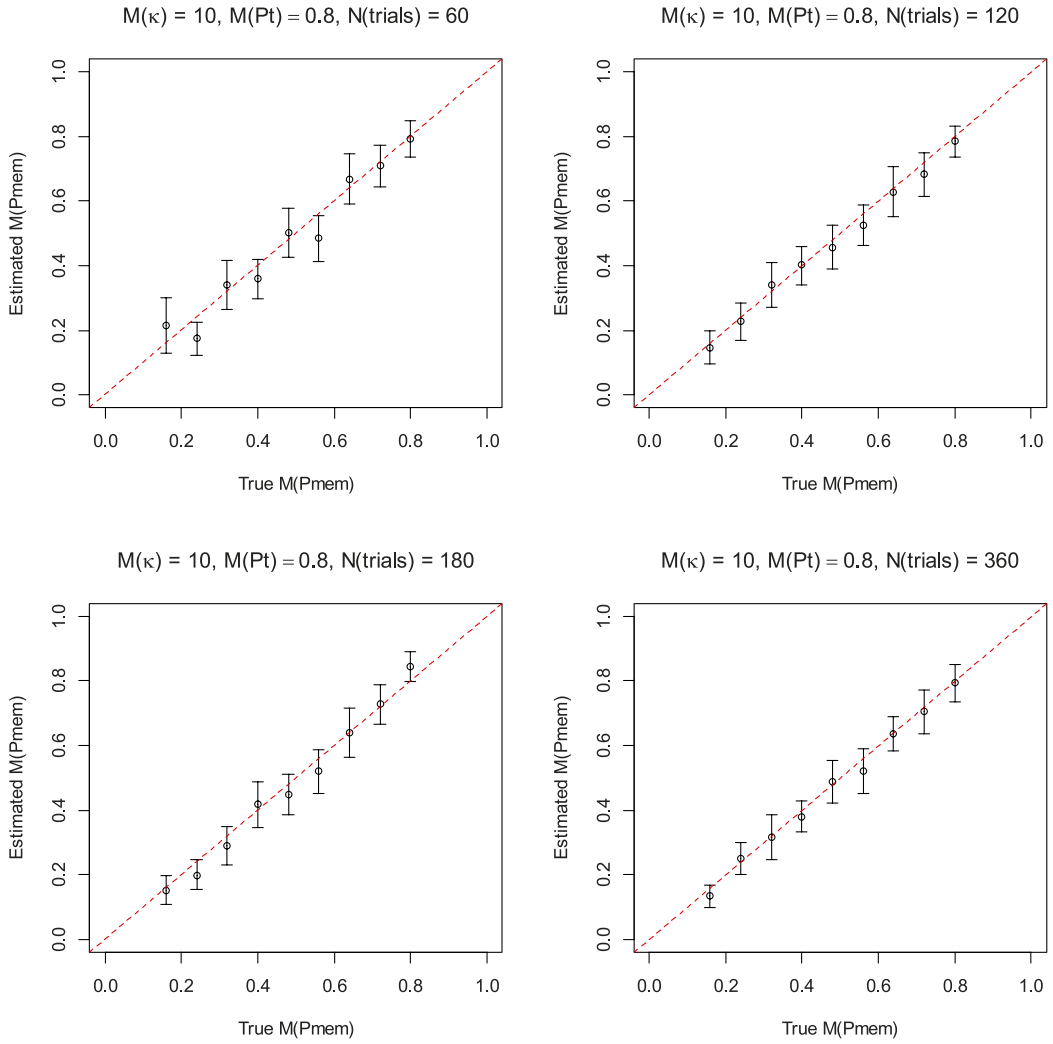

Figure 2. Recovery of the *M*(*Pmem*) parameter of the three-parameter mixture model, with typical values for nonfocal parameters: *M*(*Pt*) = 0.8, and *M*(κ) = 10, corresponding to *M*(*SD*) = 18.6. The Figure shows the means and the 95% highest-density intervals of the posterior of the group-mean parameter. The broken red diagonal line represents perfect recovery.

all objects in the memory set, and component A_b reflects stimulus-independent background noise in the memory system. The background noise is uncorrelated with the features of all objects in VWM, and can therefore be modelled as a uniform distribution on the circle, although the model is not committed to the assumption that background noise is actually uniform.

$$A_a(x) = \sum_{i=1}^{n} VM(x; x_i, \kappa)$$

$$A_b(x) = \frac{1}{2\pi}$$

The overall activation distribution is a weighted sum of the three components:

$$A(x|L_\theta) = bA_b(x) + aA_a(x) + cA_c(x|L_\theta)$$

The probability of choosing each response $\hat{x}$ out of N response options is obtained by Luce's choice rule:

$$P(\hat{x}|L_\theta) = \frac{A(\hat{x}|L_\theta)}{\sum_{j=1}^{N} A(j|L_\theta)}$$

The IMM has five parameters: a, b, c, s, and κ. Parameter estimation requires that one of the three weight parameters (a, b, and c) is fixed to set the scale of the other two. For mathematical convenience we fixed $c = 1$. Thus, the IMM has four free parameters.

The two-parameter and three-parameter mixture models discussed above can be regarded as special cases of the IMM. When the cue-generalization gradient parameter s is fixed to a very high value (e.g., $s = 20$ in our applications, measuring distance on the cue-feature dimension in radians), the gradient becomes so steep that only the target item receives any activation from the retrieval cue. As a consequence, nontarget items don't contribute to component A_c; they are included only in

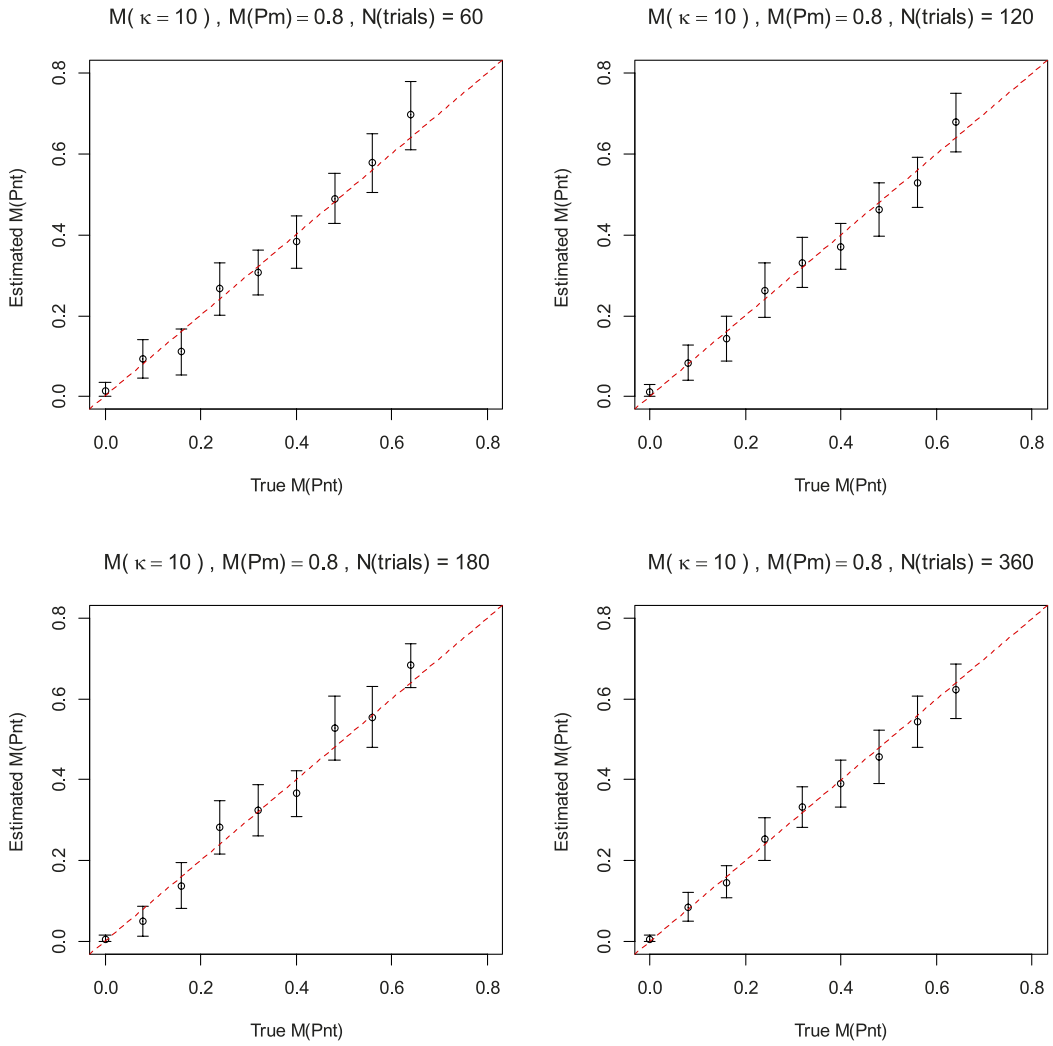

Figure 3. Recovery of the *M*(*Pnt*) parameter of the three-parameter mixture model, with typical values for the nonfocal parameters: *M*(*Pm*) = 0.8, and *M*(κ) = 10, corresponding to *M*(*SD*) = 18.6. The Figure shows the means and the 95% highest-density intervals of the posterior of the group-mean parameter. The broken red diagonal line represents perfect recovery.

A_a, which sums activation across all items independent of the cue. In this model version (IMM-*abc*), nontarget items can be recalled with above-chance probability, but their recall probability is independent of their proximity to the target on the cue-feature dimension. As a consequence, the IMM with *s* fixed to a high value is equivalent to the three-parameter mixture model of Bays et al. (2009). Constraining the IMM further by setting *a* to zero eliminates the A_a component reflecting information from the nontargets. This further constrained model (IMM-*bc*) is equivalent to the two-parameter mixture model of Zhang and Luck (2008). Appendix 1 explains this equivalence in more detail. Finally, we will also consider a model version in which we constrain only *a* to zero while leaving *s* free, implying that nontarget intrusions are entirely governed by the cue-generalization gradient parameter *s* (version IMM-*bsc*). This model can be used to investigate whether the assumption of cue-independent memory information (i.e., component A_a) is necessary.

Hierarchical Bayesian modeling

Hierarchical Bayesian models combine the strengths of hierarchical models with Bayesian parameter estimation techniques (Lee, 2011; Lee & Wagenmakers, 2014). In a hierarchical model a basic model describes the data of one observational unit with a set of parameters for that unit. Because in psychological studies the observational unit is usually a person, we will refer to these parameters as the person-specific parameters. On a higher level, the person-specific parameters are described as dependent on a more encompassing process. In the most common application of hierarchical models, person-specific parameters are modeled as drawn from a distribution of parameter values across individuals, characterized by a

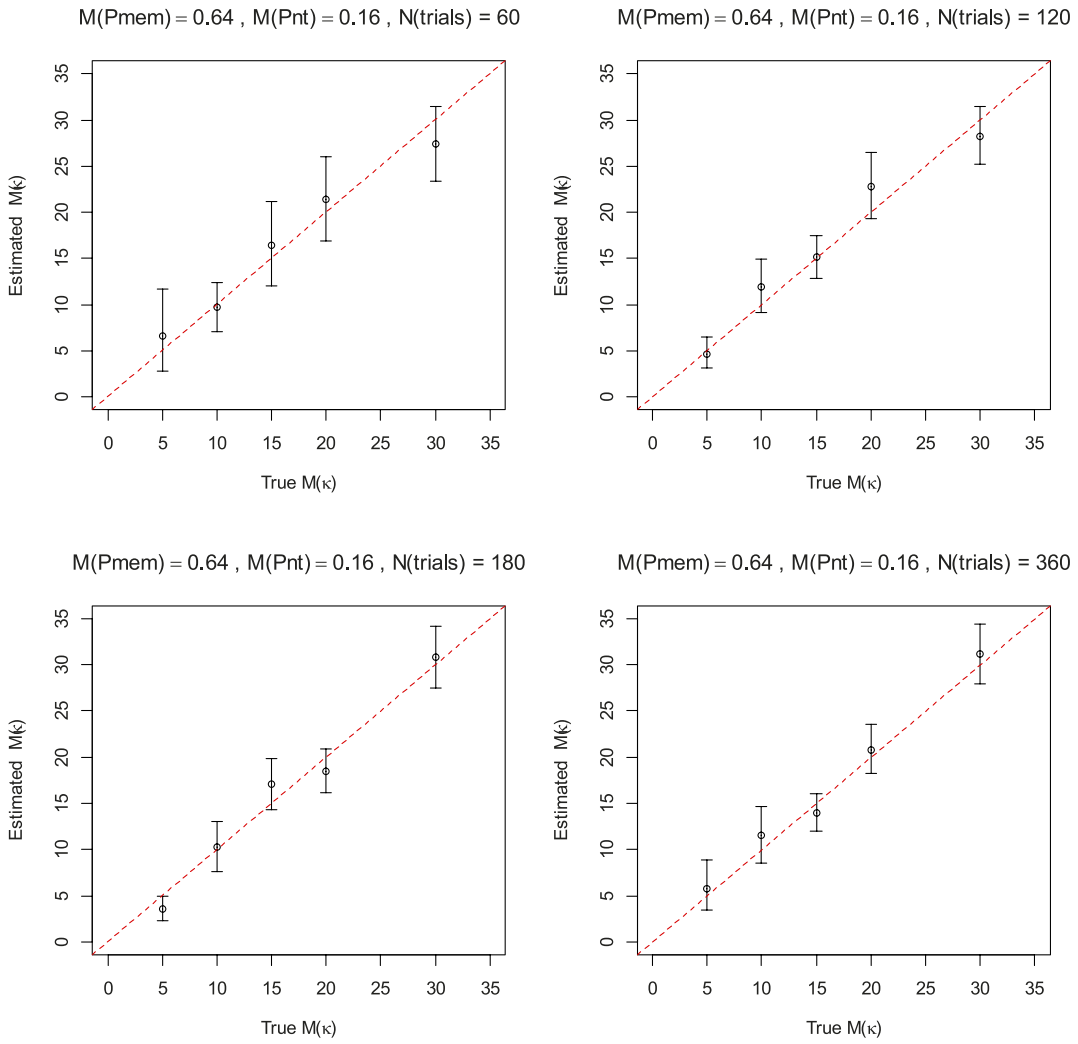

Figure 4. Recovery of parameter $M(\kappa)$ of the three-parameter mixture model with $M(Pm) = 0.8$ and $M(Pt) = 0.8$. The Figure shows the means and the 95% highest-density intervals of the posterior of the group-mean parameter. The broken red diagonal line represents perfect recovery.

mean and a dispersion parameter. In this way, hierarchical models can measure parameters on the individual level (e.g., one person's probability of recalling a memory item in an experiment) and the population level (e.g., the mean recall probability of individuals in a population, and the standard deviation characterizing individual differences). One major advantage of hierarchical modeling is that we can pool the data from all participants in a study—thereby gaining much more robust parameter estimates and statistical power—without having to average the data across participants (for the artifacts potentially arising from averaging, see Estes, 1956; Heathcote, Brown, & Mewhort, 2000).

Hierarchical models have been developed within the classical statistical framework, most notably in the form of statistical (i.e., mixed-effects) models (Pinheiro & Bates, 2000). The advantages of using a Bayesian framework are threefold. First, whereas classical models, using maximum-likelihood fitting algorithms, provide point estimates of the best-fitting parameters, applying Bayesian models yields a posterior probability density over possible parameter values, thereby providing information not only about the best point estimate but also the range of credible parameter estimates, and the uncertainty of parameter estimation. Second, whereas classical hierarchical models require a special statistical apparatus that adds an extra layer of complexity over and above the basic model, Bayesian hierarchical models are natural extension of basic Bayesian models, and are therefore very easy to build (Lee & Wagenmakers, 2014). Moreover, they are very flexible, allowing the user to specify any distribution of data or parameters on each level of the hierarchy. This is particularly pertinent for modeling data from the continuous-reproduction paradigm, which are distributed over a continuous circular response scale. Modeling these data requires distributions from circular statistics, pri-

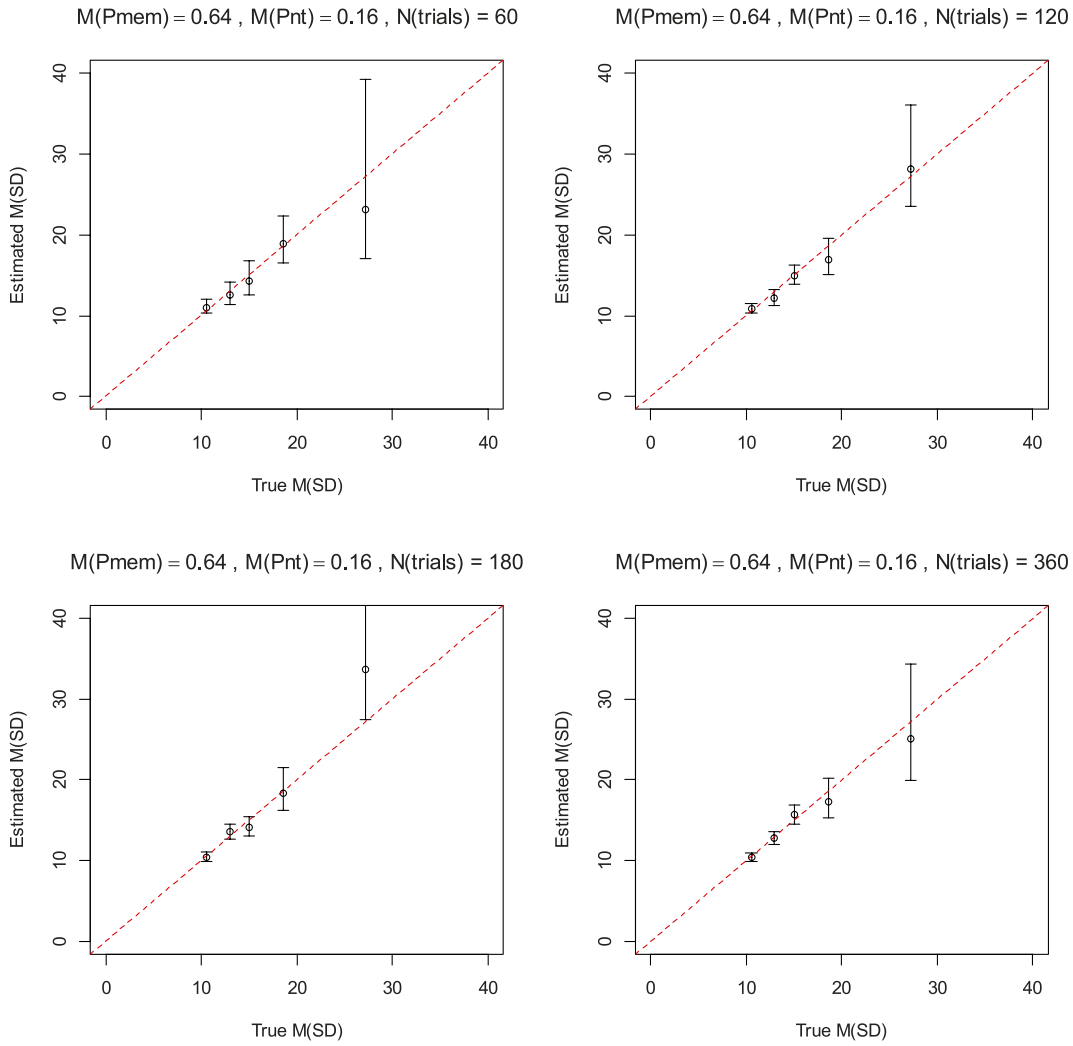

Figure 5. Recovery of the $M(SD)$ parameter implied by $M(\kappa)$ of the three-parameter mixture model with $M(Pm) = 0.8$ and $M(Pt) = 0.8$. The Figure shows the means and the 95% highest-density intervals of the posterior of the group-mean parameter. The broken red diagonal line represents perfect recovery.

marily the circular normal or von Mises distribution, for which, to the best of our knowledge, no hierarchical model has been developed yet. Third, the Bayesian framework draws on Markov-Chain Monte-Carlo (MCMC) algorithms for estimating the posterior densities of parameter values. MCMC algorithms provide powerful tools for efficiently searching high-dimensional parameter spaces, enabling the user to build more complex models with a larger number of parameters.

We implemented the two measurement models in JAGS (Plummer, 2016), a program for running Bayesian models, in combination with the *rjags* package for R that interfaces between R and JAGS. One obstacle was that JAGS has no built-in von Mises distribution. Therefore we built a JAGS extension, following the tutorial of Wabersich and Vandekerckhove (2014). The von Mises extension for JAGS can be downloaded from GitHub [https://github.com/yeagle/jags-vonmises/releases]. The modelling code and the data of the two experiments reanalyzed below can be found on the Open Science Framework (osf.io/wjg7y/).

The hierarchical Bayesian mixture model (HMM)

We reparameterized the three-component mixture model to circumvent the problem of constraining the sum of three probability parameters to one. Parameter *Pm* is the probability that the response comes from any of the objects in the memory set, including the target and all nontargets, and *Pt* is the conditional probability that the response reflects the feature of the target, given that it reflects the feature of any memory object. Hence,

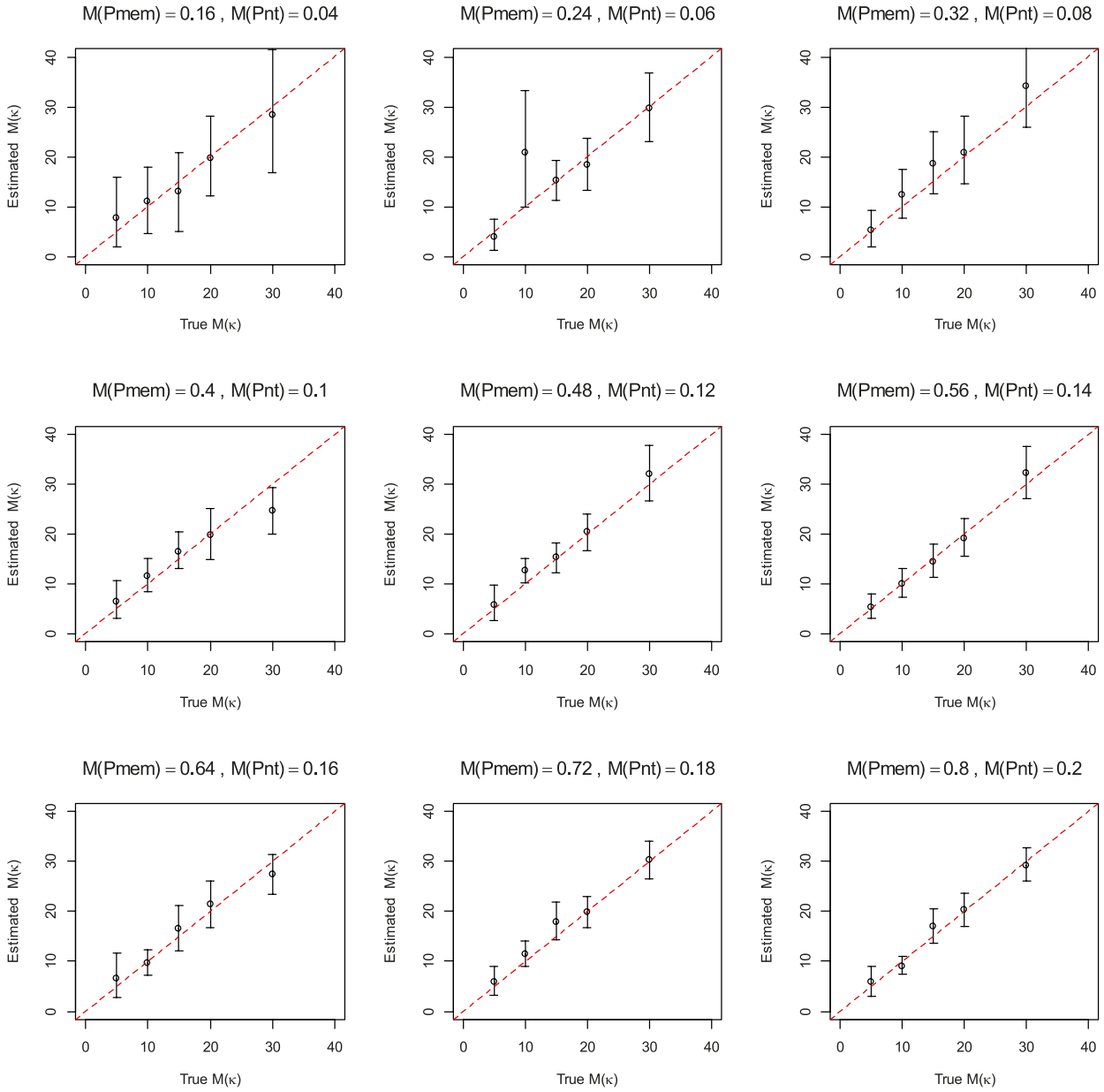

Figure 6. Recovery of the precision parameter $M(\kappa)$ of the three-parameter mixture model as a function of $M(Pm)$, with $M(Pt) = 0.8$ and $N(\text{trials}) = 60$. The Figure shows the means and the 95% highest-density intervals of the posterior of the group-mean parameter. The broken red diagonal line represents perfect recovery.

$$Pmem = Pm \cdot Pt$$
$$Pnt = Pm \cdot (1 - Pt)$$

The hierarchical mixture model (HMM) predicts each response $\hat{x}_{i,j}$ of participant j in trial i as distributed according to a von Mises distribution:

$$\hat{x}_{i,j} \sim VM\left(m_{i,j}; z_{i,j} \cdot \kappa_j\right)$$

The mean of the von Mises, $m_{i,j}$, is the true feature of the array object from which the response derives. The object determining the response is sampled from the array objects according to a categorical distribution with probabilities $\mathbf{P}_{i,j}$

$$m_{i,j} = \mathbf{M}_{i,j}(object_{i,j})$$
$$object_{i,j} \sim Cat\left(\mathbf{P}_{i,j}\right)$$
$$\mathbf{P}_{i,j}(1) = Pt_j$$
$$\mathbf{P}_{i,j}(2:n) = \frac{(1 - Pt_j)}{n - 1}$$

$\mathbf{M}_{i,j}$ is the vector of memory features in array i presented to person j, the first element being the target,

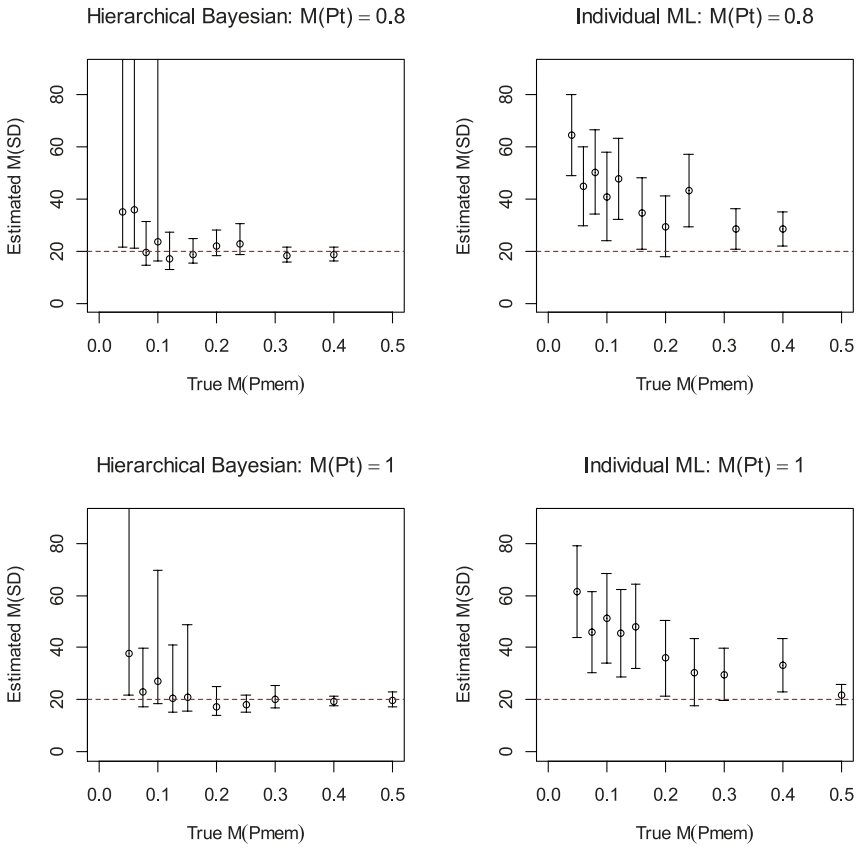

Figure 7. Recovery of the *M*(*SD*) parameter of the three-parameter mixture model as a function of *M*(*Pmem*), Simulation 2. The left panels show the means and the 95% highest-density intervals of the posterior of the group-mean parameter from the hierarchical Bayesian mixture model; the right panels show the means and 95% confidence intervals of individual parameter estimates obtained from fitting the mixture model separately to data from each subject with maximum likelihood (ML). The broken red line represents the true value of *M*(*SD*).

and $\mathbf{P}_{i,j}$ is the vector of probabilities of a response reflecting each of the array features.

The response is related to the sampled memory feature $m_{i,j}$ with probability Pm_j, whereas with probability $1 - Pm_j$ the response is drawn from a uniform distribution. Conveniently, the von Mises distribution with precision $\kappa = 0$ is a uniform distribution on the circle. Therefore, we can multiply the precision parameter κ_j, with a binary variable $z_{i,j}$ drawn from a Bernoulli distribution with parameter Pm_j:

$$z_{i,j} \sim Bern\left(Pm_j\right)$$

When $z_{i,j} = 1$ (with probability Pm_j), the response is modeled as drawn from a von Mises centered on $m_{i,j}$ whereas when $z_{i,j} = 0$ (with probability $1 - Pm_j$) the response is drawn from a uniform distribution.

The parameters for individual subjects, Pm_j, Pt_j, and κ_j, are drawn from distributions characterizing their variability across subjects:

$$Pm_j \sim Beta(A_{Pm}, B_{Pm})$$
$$Pt_j \sim Beta(A_{Pt}, B_{Pt})$$
$$\kappa_j \sim Gamma(S_\kappa, R_\kappa)$$

We placed moderately informative priors on the group level parameters (a.k.a., hyper-parameters) for these distributions:

$$A_{Pm} \sim Gamma(1, 0.1)$$
$$B_{Pm} \sim Gamma(1, 0.1)$$
$$A_{Pt} \sim Gamma(1, 0.1)$$
$$B_{Pt} \sim Gamma(1, 0.1)$$
$$M(\kappa) \sim Gamma(1, 0.1)$$
$$\sigma(\kappa) \sim Gamma(1, 0.1)$$
$$S_\kappa = \frac{M(\kappa)^2}{\sigma(\kappa)^2}$$
$$R_\kappa = \frac{M(\kappa)}{\sigma(\kappa)^2}$$

The Gamma distribution is parameterized by shape and rate; we found it convenient to place

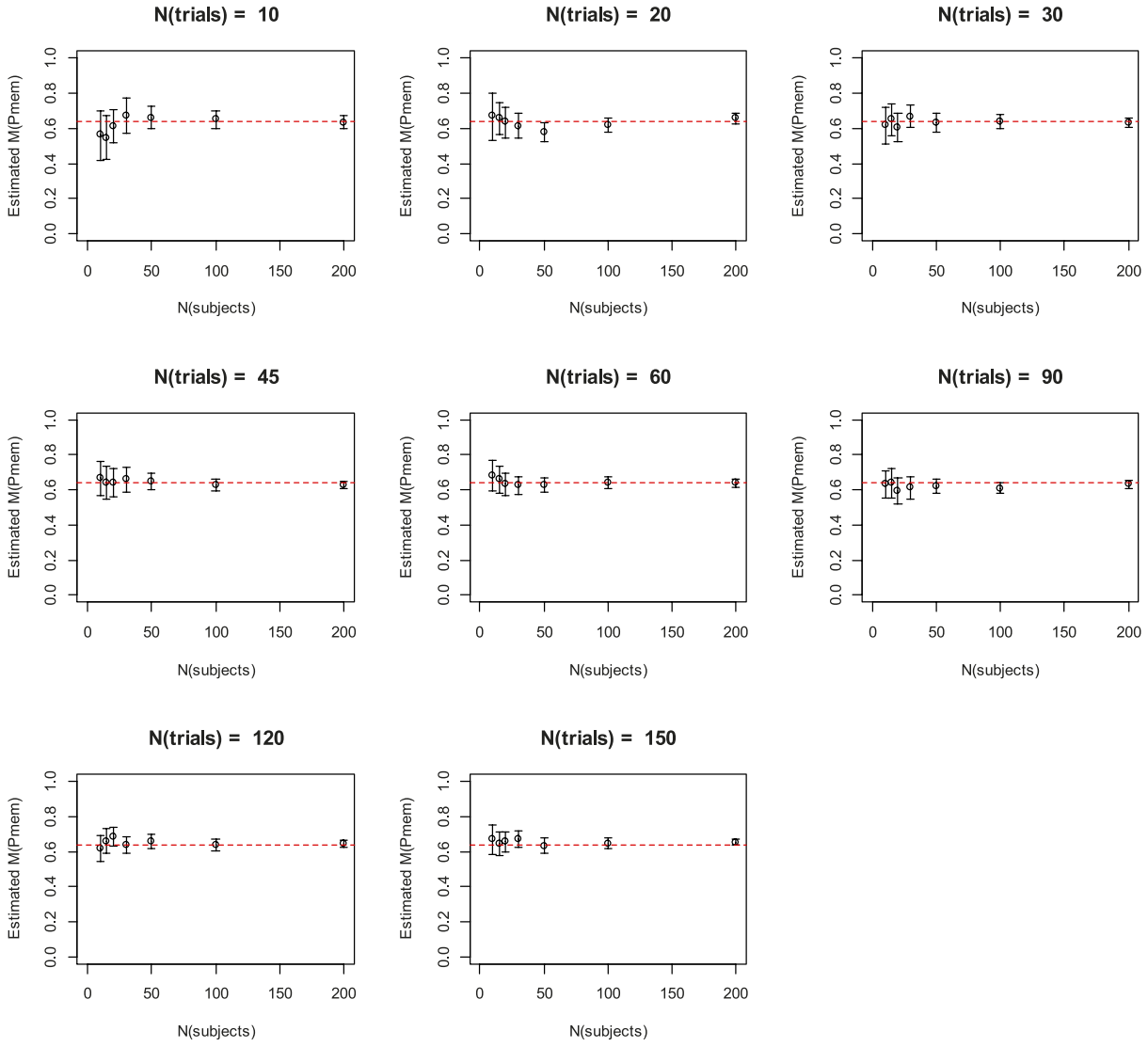

Figure 8. Recovery of the M(Pmem) parameter of the three-parameter mixture model as a function of number of trials and of subjects (Simulation 3), with $M(Pt) = 0.8$, and $M(\kappa) = 10$. The Figure shows the means and the 95% highest-density intervals of the posterior of the group-mean parameter. The broken red line represents the true value of $M(Pmem)$.

priors on the mean and the standard deviation of the precision parameter, M_κ and σ_κ, respectively, and derive scale and rate from them. Conversely, the means and the variances of the beta distributions of the probability parameters can be calculated as:

$$M(Pm) = \frac{A_{Pm}}{A_{Pm} + B_{Pm}}$$

$$Var(Pm) = \frac{A_{Pm} \cdot B_{Pm}}{[A_{Pm} + B_{Pm}]^2 [A_{Pm} + B_{Pm} + 1]}$$

$$M(Pt) = \frac{A_{Pt}}{A_{Pt} + B_{Pt}}$$

$$Var(Pt) = \frac{A_{Pt} \cdot B_{Pt}}{[A_{Pt} + B_{Pt}]^2 [A_{Pt} + B_{Pt} + 1]}$$

Parameter recovery simulations with the HMM

To evaluate the HMM's accuracy and precision of measurement we ran three parameter-recovery simulations. We simulated data from the three-component mixture model with known parameter values and applied the HMM to them. The first simulation study varied the three model parameters over a broad range, and in addition varied the number of trials per subject. The second simulation study focused on a particularly challenging problem for the mixture model, estimating the precision parameter when *Pmem* is very small, such as when memory set size is large, or recall is from long-term memory (Brady, Konkle, Gill, Oliva, & Alvarez, 2013). The third simulation demonstrates the power of hierarchical modeling for pooling information from

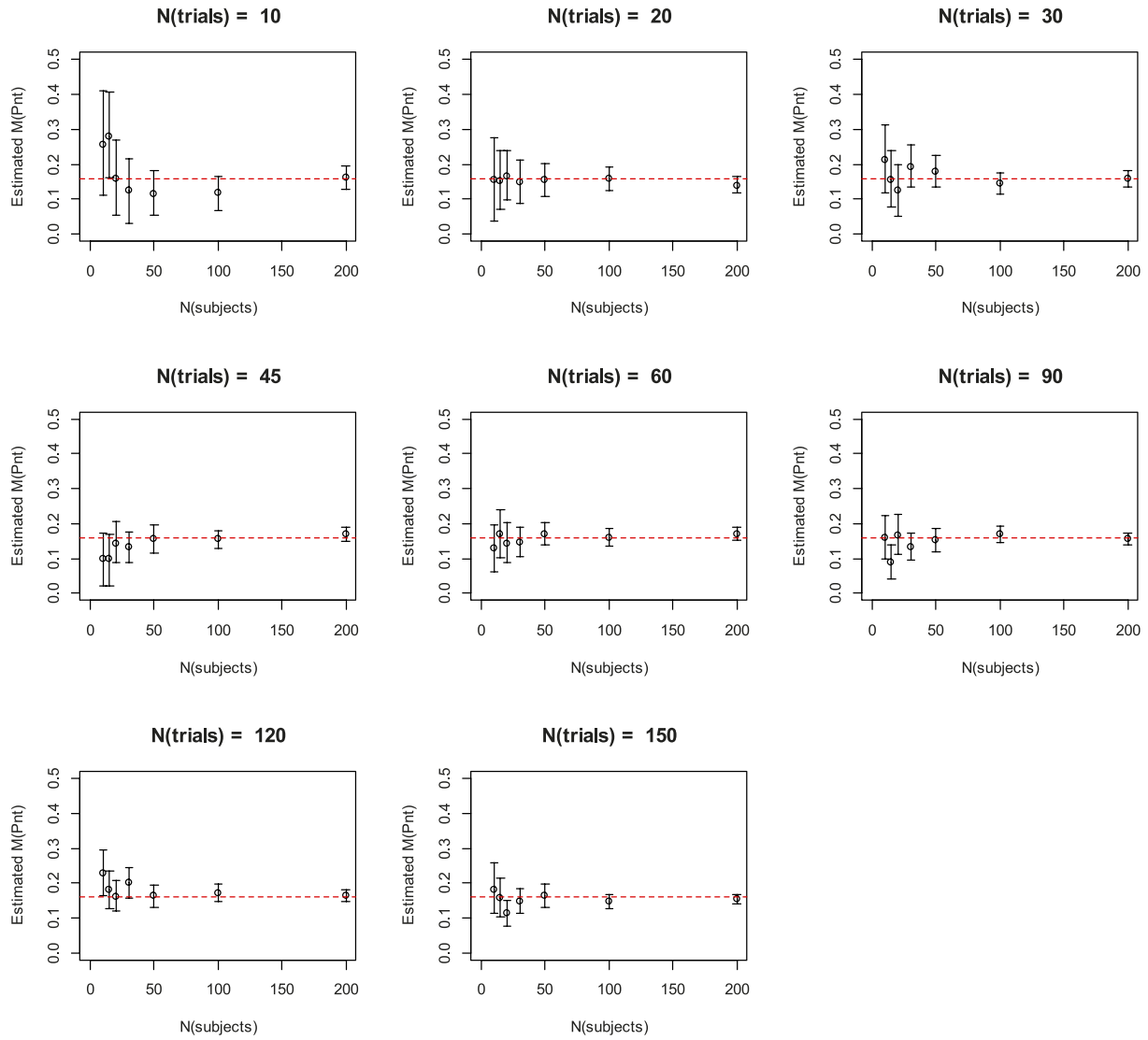

Figure 9. Recovery of the *M*(*Pnt*) parameter of the three-parameter mixture model as a function of number of trials and of subjects (Simulation 3), with $M(Pm) = 0.8$, and $M(\kappa) = 10$. The Figure shows the means and the 95% highest-density intervals of the posterior of the group-mean parameter. The broken red line represents the true value of *M*(*Pnt*).

many subjects, each contributing only few trials, for estimating group-level parameters.

Simulation 1

Each simulation generated data for 20 subjects for a variable number of trials, $N(\text{trials}) = 60$, 120, 180, or 360. Each simulation had true population mean parameters $M(Pm)$, $M(Pt)$, and $M(\kappa)$. We varied these mean parameters in a fully crossed design, with

$$M(Pm) = [0.5, 0.6, 0.7, 0.8, 0.9, 1]$$

$$M(Pt) = [0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]$$

$$M(\kappa) = [5, 10, 15, 20, 30].$$

Subject parameters for Pm and Pt were drawn from Beta distributions, with parameters $A_{Pm} = 10*\text{M}(Pm)$ and $B_{Pm} = 10*(1 - M(Pm))$ for Pm, and analogously for Pt. Subject parameters for κ were drawn from a Gamma distribution with mean $= M(\kappa)$ and variance $= 25$.

Applying the model returns Markov-Chain Monte-Carlo (MCMC) chains for each parameter. We obtained chains of 10,000 samples, preceded by 4,000 adaptation steps. From these chains we computed the 95% highest density interval (HDI) using the method and R code of Kruschke (2011). The HDI gives the smallest interval of parameter values that is covered by 95% of the posterior density. They mean that, given our priors and the data, we can be 95% certain that the true parameter value lies within the HDI. Figures 2 to 5 plot

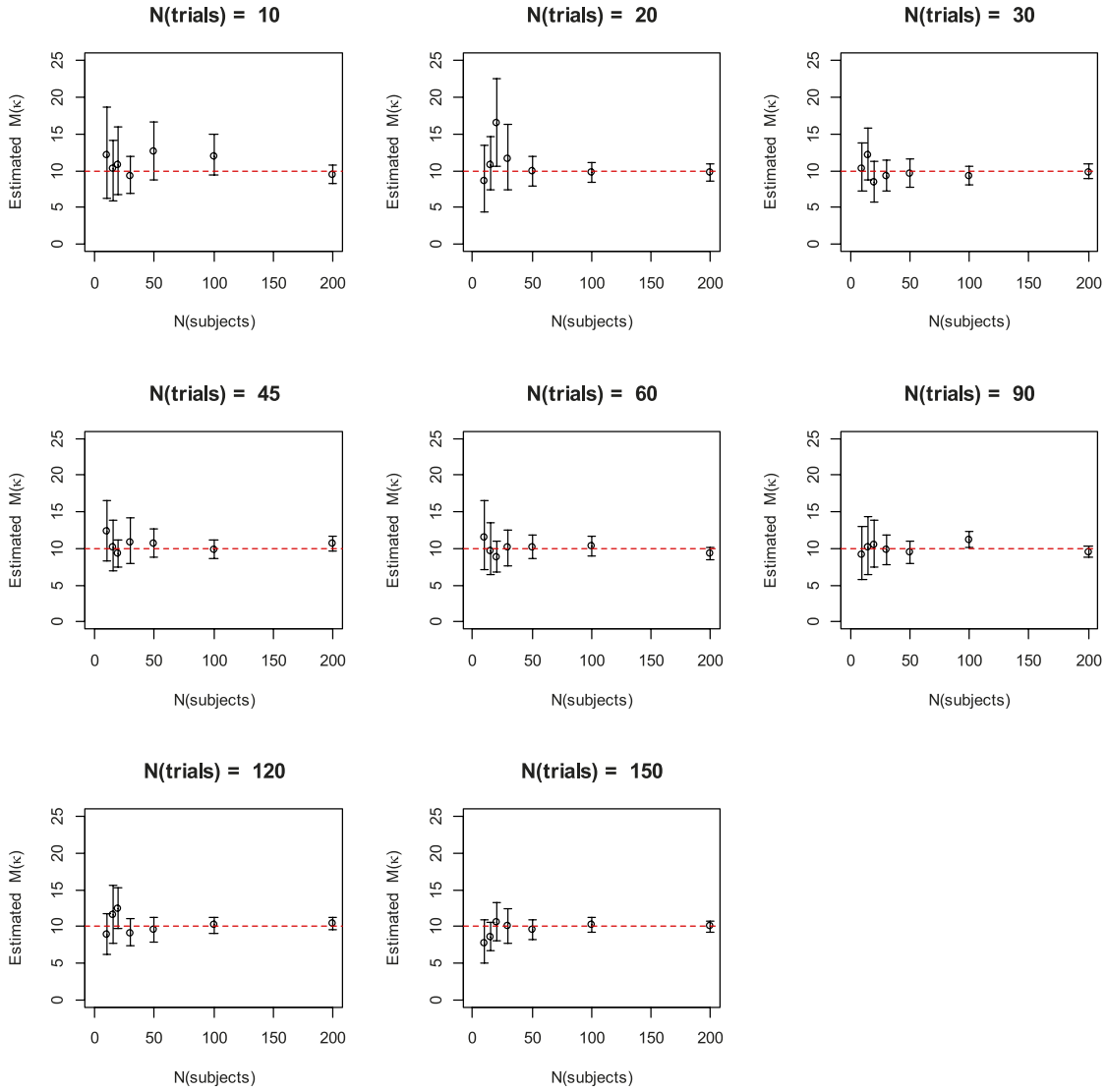

Figure 10. Recovery of the $M(\kappa)$ parameter of the three-parameter mixture model as a function of number of trials and of subjects (Simulation 3), with $M(Pm) = 0.8$, and $M(Pt) = 0.8$. The Figure shows the means and the 95% highest-density intervals of the posterior of the group-mean parameter. The broken red line represents the true value of $M(\kappa)$.

the means and the HDIs of the posteriors of the group-level parameters as a function of the corresponding parameter's true values, separately for each number of trials. The true and estimated values of $M(Pmem)$ and $M(Pnt)$ were calculated from the true and estimated values, respectively, of $M(Pm)$ and $M(Pt)$.

These results show that recovery of the probability parameters $M(Pmem)$ and $M(Pnt)$ is accurate (i.e., unbiased) and fairly precise already with 60 trials per person, and hardly improves with larger numbers of trials. Recovery of the precision parameter $M(\kappa)$, and the corresponding values of $M(SD)$, is less satisfactory at 60 trials per person, but improves with larger number of trials.

The results in Figures 2 to 5 are from simulations in which the two nonfocal parameters are set to typical values (i.e., $M(Pm) = 0.8$, $M(Pt) = 0.8$, and $M(\kappa) = 10$, corresponding to $SD = 18.6$). Changing the values of $M(Pt)$ had no noticeable impact on the recovery of $M(Pm)$, and vice versa. Higher values of κ resulted in somewhat smaller HDIs for recovery of $M(Pm)$ and $M(Pt)$. In contrast, the values of $M(Pm)$ and $M(Pt)$ had a pronounced influence on how well $M(\kappa)$, and hence, $M(SD)$ was recovered.

Figure 6, plotting the $M(\kappa)$ estimates for varying levels of $M(Pm)$, shows that the precision of estimating $M(\kappa)$ decreases noticeably as M(Pm) decreases. The reason for this is that with smaller $M(Pm)$, a larger proportion of trials are drawn from the uniform component, and these trials are uninformative for estimating the precision of the von-Mises component. Figure 6 also shows that, as long as $M(Pm)$ is not too small, reasonably precise estimates of M(κ) can already be obtained with as little as 60 trials per subject,

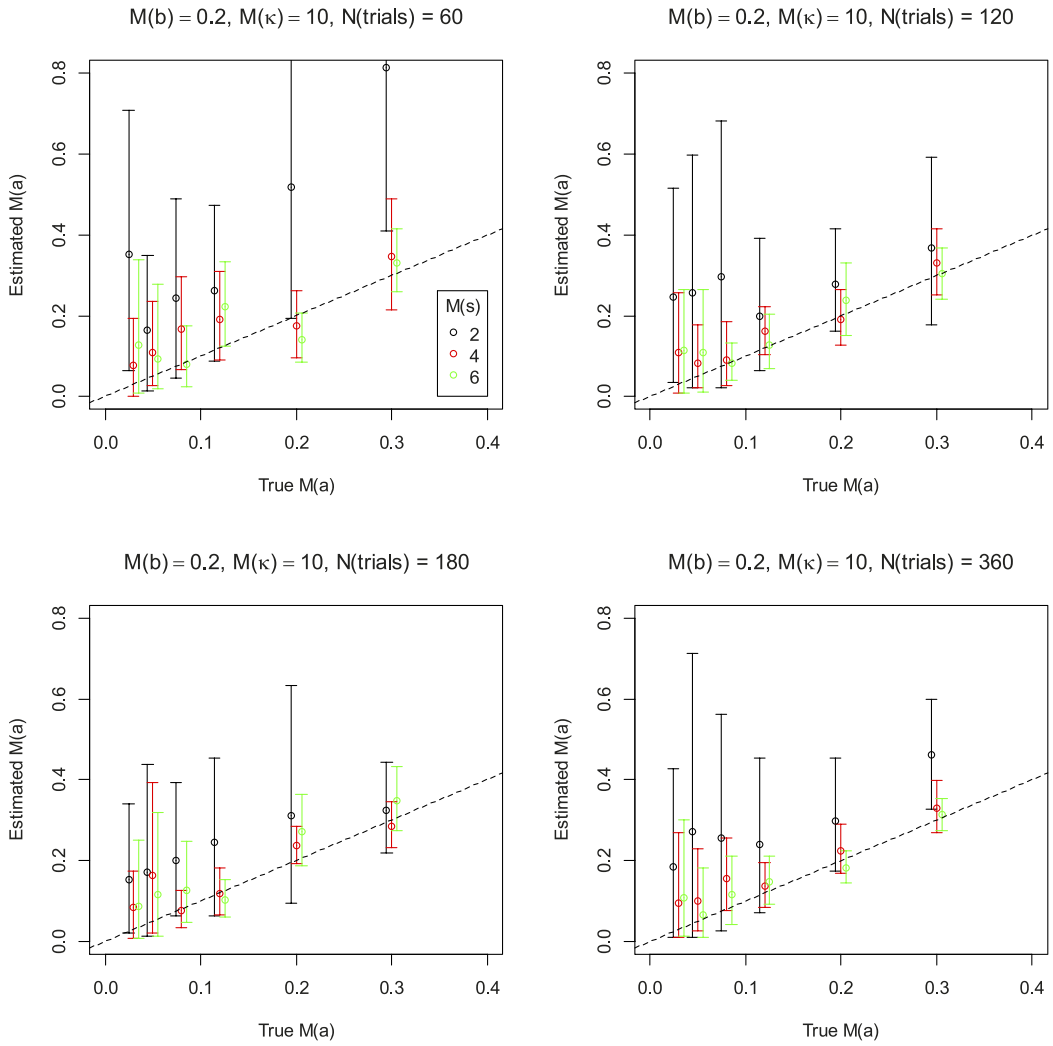

Figure 11. Recovery of the *M*(*a*) parameter of the IMM as a function of *M*(*s*), with other parameters set to typical values. The Figure shows the means and the 95% highest-density intervals of the posterior of the group-mean parameter. The broken diagonal line represents perfect recovery. Data points of different colors are offset to improve readability.

assuming a subject sample size as is typical for experiments in cognitive psychology.

Simulation 2

Using the traditional maximum-likelihood based method of fitting the mixture model to individual subject's data, Sutterer and Awh (2016) observed that for *Pmem* values of 0.4 or smaller, estimates of κ are biased downward, and hence, estimates of *SD* are biased upward. This bias limits the usability of the mixture model in conditions of poor memory, such as VWM tests with large set sizes, or tests of visual long-term memory. Simulation 2 served to analyze whether the hierarchical Bayesian framework protects against such bias. We generated data from the three-parameter mixture model with $M(SD) = 20$, $M(Pt) = 0.8$ or 1, and varying $M(Pmem)$ in small steps between 0.05 and 0.5. We simulated data for 100 trials from each of 20 subjects. Figure 7 (left panels) shows the parameter recovery for M(*SD*). An upward bias in $M(SD)$ is noticeable, but only for $M(Pmem) < 0.1$. With $M(Pmem) = 0.2$ or higher, the recovery of $M(SD)$ was unbiased and had a reasonably narrow HDI.

The right panels of Figure 7 show the results of fitting the mixture model to individual subject's data (generated by an equivalent simulation). This simulation replicates the upward bias in estimates of SD observed by Sutterer and Awh (2016).

Simulation 3

The primary advantage of hierarchical modeling is that it obtains group-level parameter estimates while allowing for potentially important individual differences. This enables researchers to estimate parameters even when each subject contributes few data points. Simulation 3 explores the trade-off between number of

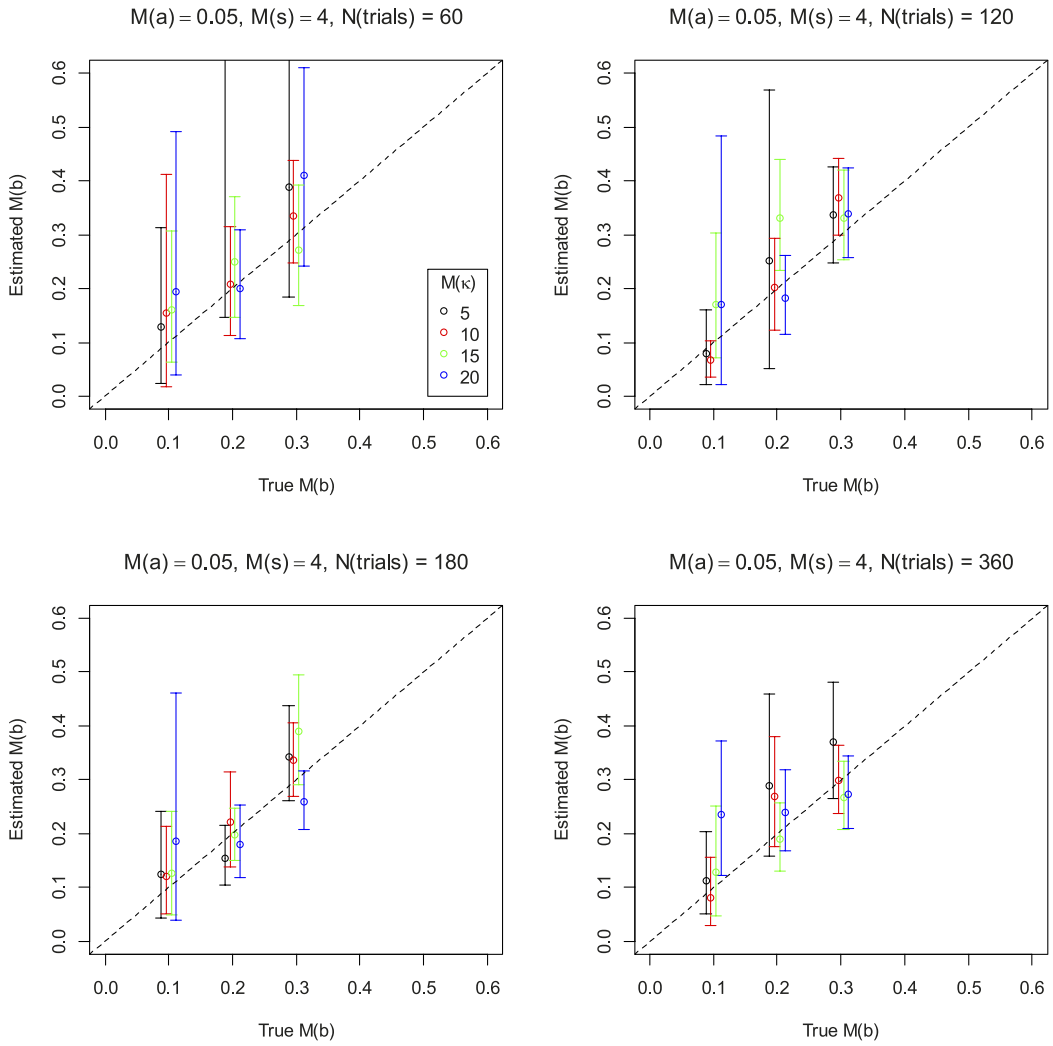

Figure 12. Recovery of the *M*(*b*) parameter of the IMM as a function of *M*(κ), with other parameters set to typical values. The Figure shows the means and the 95% highest-density intervals of the posterior of the group-mean parameter. The broken diagonal line represents perfect recovery. Data points of different colors are offset to improve readability.

subjects and number of trials per subject. We fixed the true parameters to typical values, $M(Pm) = 0.8$, $M(Pt) = 0.8$, $M(\kappa) = 10$, and varied the number of subjects (from 10 to 200) and the number of trials per subject (from 10 to 150). Figures 8 to 10 show the parameter recovery for $M(Pmem)$, $M(Pnt)$, and $M(\kappa)$, respectively. The perhaps surprising finding is that with a moderately large number of subjects ($N = 50$ or more), accurate and reasonably precise parameter estimates can be obtained even with very small numbers of trials per subject. For instance, HDIs of parameter estimates for $N = 50$ and 20 trials per subject are about as narrow as those for $N = 20$ and 90 trials per subject. This means that researchers can use the hierarchical mixture model for experiments with many conditions, for which each subject can contribute only a small number of trials (Bae, Olkkonen, Allred, & Flombaum, 2015), or for data from online experiments, in which many subjects participate for only a short time.

The hierarchical interference measurement model (HIMM)

The IMM models the probability of choosing each response $\hat{x}$ as the normalized distribution of activation over the response scale. This distribution is difficult to handle because it changes with the constellation of features in each memory array, and is often multimodal. Fortunately, we can break the activation distribution down into a weighted mixture of von-Mises distributions:

$$A(x|L_\theta) = \sum_{i=1}^{n} w_i VM(x; x_i, \kappa) + w_{n+1} VM(x; 0, 0)$$

The weights w_i for i between 1 and set size n are given by the combination of the weights of cue-based retrieval (governed by parameters $c = 1$, and s), and of cue-independent memory (governed by parameter a):

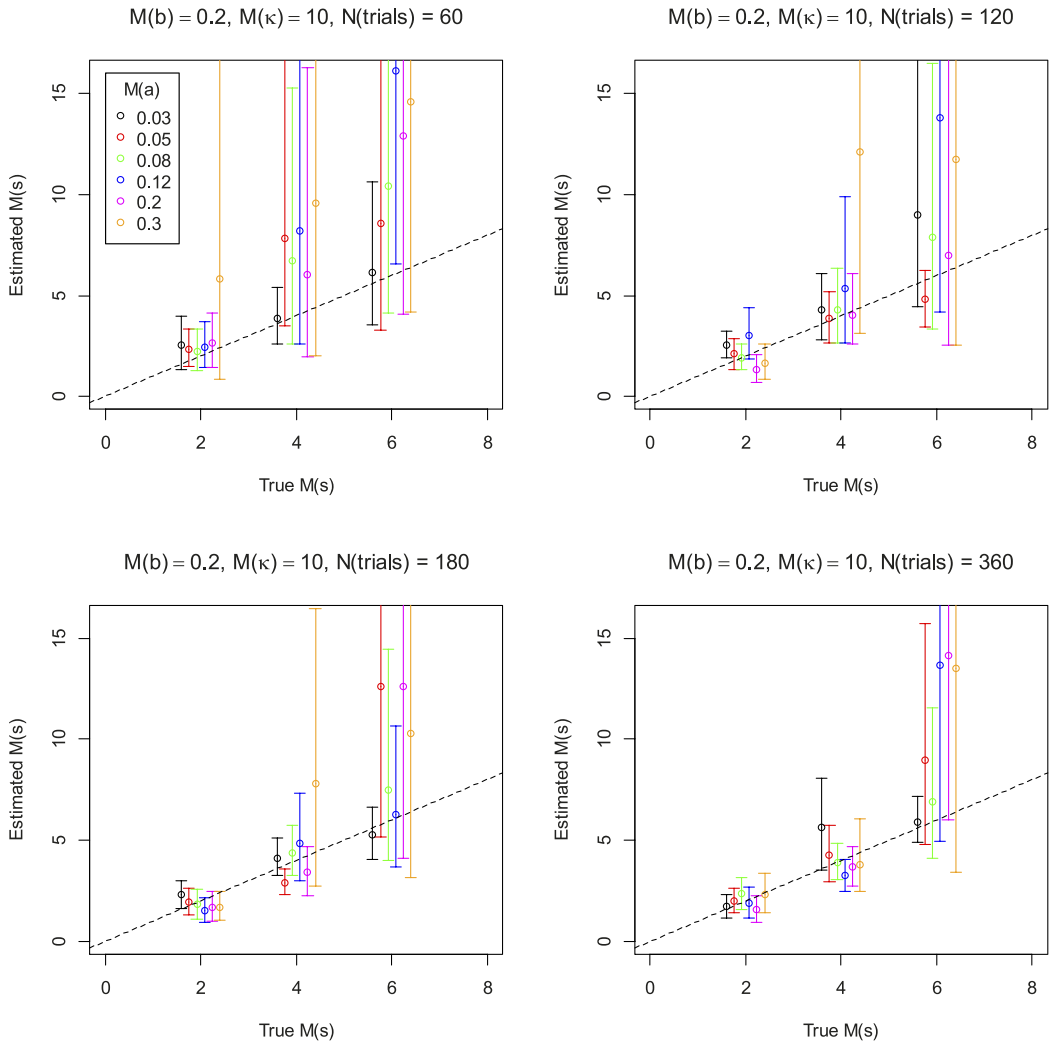

Figure 13. Recovery of the *M*(*s*) parameter of the IMM as a function of *M*(*a*), with other parameters set to typical values. The Figure shows the means and the 95% highest-density intervals of the posterior of the group-mean parameter. The broken diagonal line represents perfect recovery. Data points of different colors are offset to improve readability.

$$w_i = a + c\exp[-s \cdot D(L_i, L_\theta)]$$

The weight w_{n+1} is the contribution of uniform background noise, b. The von-Mises distribution with precision = 0 is the uniform distribution on the circle, so that we can model the entire activation distribution as a mixture of von-Mises distributions with different mean and precision parameters. The mixture weights can be translated into mixture probabilities by normalizing:

$$p_i(x|L_\theta) = \frac{w_i(x|L_\theta)}{\sum_{j=1}^{n+1} w_j(x|L_\theta)}$$

We can now express the likelihood of the IMM as a probability mixture of von-Mises distributions:

$$P(\hat{x}|L_\theta) = \sum_{i=1}^{n} p_i VM(x; x_i, \kappa) + p_{n+1} VM(x; 0, 0)$$

In the hierarchical Bayesian IMM (HIMM) we model response $\hat{x}_{i,j}$ of participant j in trial i as distributed according to a von-Mises distribution:

$$\hat{x}_{i,j} \sim VM\left(m_{i,j}; z_{i,j} \cdot \kappa_j\right)$$

The mean of the von-Mises, $m_{i,j}$, is the true feature of one array object, sampled from the array objects according to a categorical distribution with probabilities $\mathbf{P}_{i,j}$, which is the vector of mixture probabilities p_1 to p_{n+1}. The precision κ_j is multiplied by an indicator variable $z_{i,j}$ that equals 1 for values of $object_{i,j}$ between 1 and setsize n, and equals 0 in case $object_{i,j} = n{+}1$, so that precision is set to 0, and the von-Mises turns into a uniform distribution, whenever category $n{+}1$ is sampled

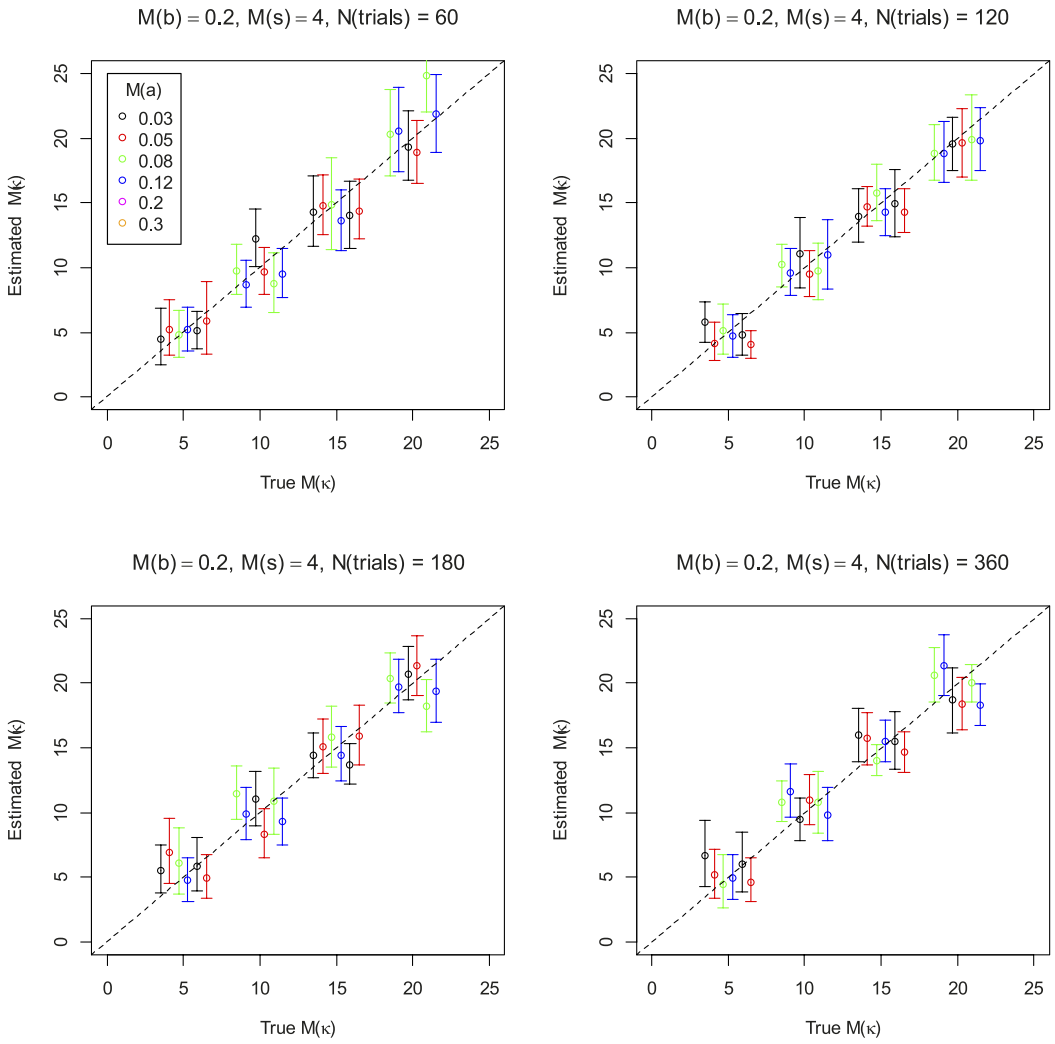

Figure 14. Recovery of the $M(\kappa)$ parameter of the IMM as a function of $M(a)$, with other parameters set to typical values. The Figure shows the means and the 95% highest-density intervals of the posterior of the group-mean parameter. The broken diagonal line represents perfect recovery. Data points of different colors are offset to improve readability.

from the categorical distribution.

$$m_{i,j} = \mathbf{M}_{i,j}(object_{i,j})$$
$$z_{i,j} = \mathbf{Z}(object_{i,j})$$
$$object_{i,j} \sim Cat(\mathbf{P}_{i,j})$$

$\mathbf{M}_{i,j}$ is the vector of memory features in array i presented to person j; $\mathbf{Z}$ is the vector of indicator variables, and $\mathbf{P}_{i,j}$ is the vector of mixture probabilities.

The person-level parameters a_j, b_j, and κ_j were drawn from Gamma distributions:

$$a_j \sim Gamma(S_a, R_a)$$
$$b_j \sim Gamma(S_b, R_b)$$
$$\kappa_j \sim Gamma(S_\kappa, R_\kappa)$$

The parameter s of the exponential generalization gradient on the cue-feature dimension was drawn for each person from a normal distribution

$$s_j \sim Normal(M(s), \sigma(s))$$

As priors for the group-level parameters we used moderately informative priors, informed by maximum-likelihood fits of the IMM to several data sets:

$$M(a) \sim Gamma(1, 2) \qquad M(b) \sim Gamma(1, 0.5)$$
$$\sigma(a) \sim Gamma(1, 2) \qquad \sigma(b) \sim Gamma(1, 0.5)$$
$$S_a = \frac{M(a)^2}{\sigma(a)^2}; R_a = \frac{M(a)}{\sigma(a)^2} \qquad S_b = \frac{M(b)^2}{\sigma(b)^2}; R_b = \frac{M(b)}{\sigma(b)^2}$$
$$M(\kappa) \sim Gamma(1, 0.1) \qquad M(s) \sim Normal(3, 100)$$
$$\sigma(\kappa) \sim Gamma(1, 0.1) \qquad \sigma(s) \sim Gamma(1, 0.1)$$
$$S_\kappa = \frac{M(\kappa)^2}{\sigma(\kappa)^2}; R_\kappa = \frac{M(\kappa)}{\sigma(\kappa)^2}$$

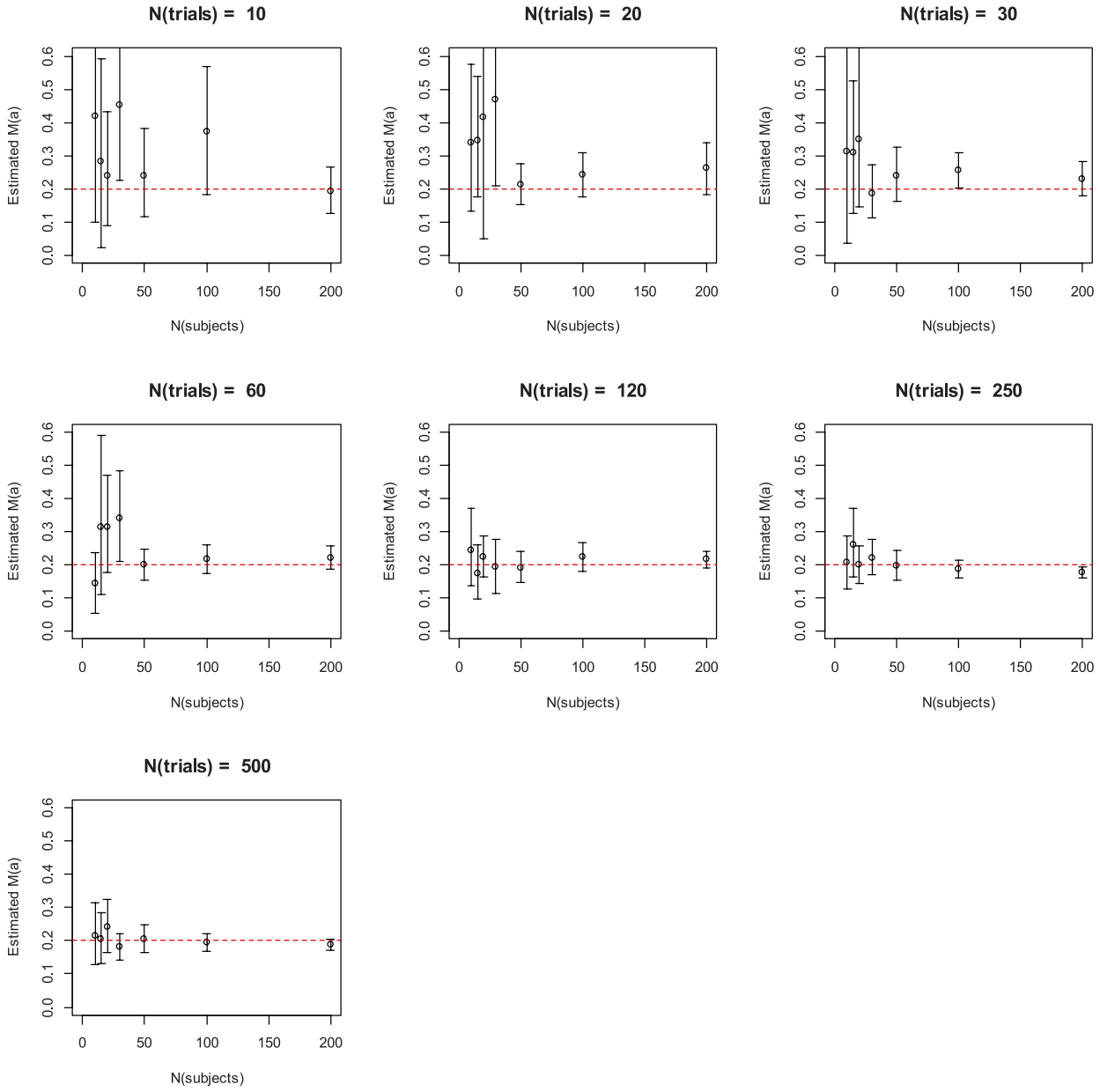

Figure 15. Recovery of the *M*(*a*) parameter of the IMM as a function of number of trials and number of subjects. The Figure shows the means and the 95% highest-density intervals of the posterior of the group-mean parameter. The broken red line is the true parameter value.

Parameter recovery simulations with the HIMM

We ran two parameter recovery simulations with the HIMM. The first simulation served to test the model's ability to recover parameters from data of a typical VWM experiment. The second simulation explored the parameter recovery for small numbers of trials per subject.

Simulation 1

We generated data from all combinations of the following values for the four free group-level parameters of the HIMM:

$$M(a) = [0.03,\ 0.05,\ 0.08,\ 0.12,\ 0.2,\ 0.3]$$

$$M(b) = [0.1,\ 0.2,\ 0.3]$$

$$M(s) = [2,\ 4,\ 6]$$

$$M(\kappa) = [5,\ 10,\ 15,\ 20].$$

The between-subject variances were set to $\mathrm{Var}(a) = 0.01$, $\mathrm{Var}(b) = 0.015$, $\mathrm{Var}(s) = 3$, and $\mathrm{Var}(\kappa) = 15$. These parameter means and variances were chosen

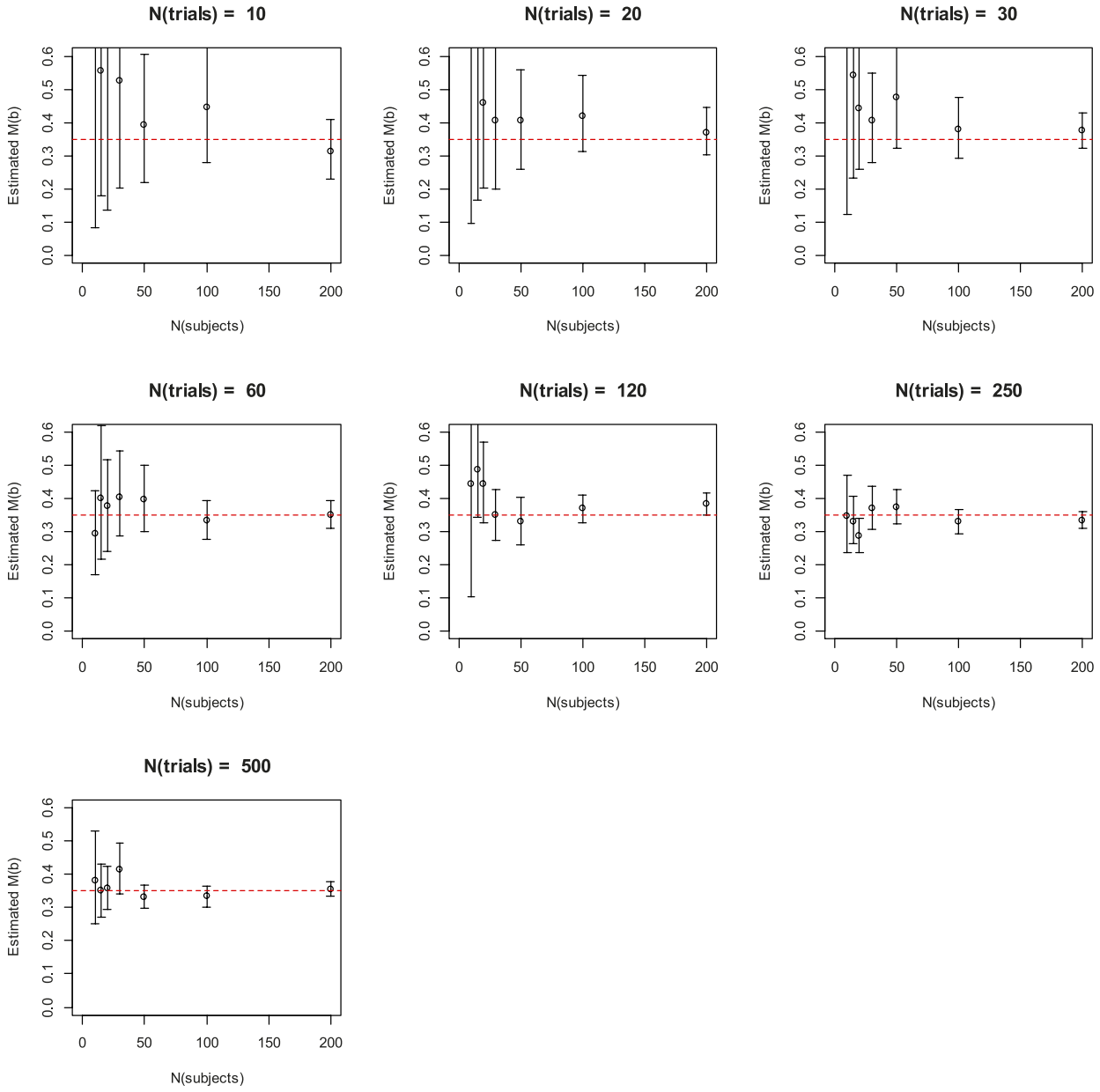

Figure 16. Recovery of the *M*(*b*) parameter of the IMM as a function of number of trials and number of subjects. The Figure shows the means and the 95% highest-density intervals of the posterior of the group-mean parameter. The broken red line is the true parameter value.

for representing the distributions of typical values that were estimated when we fitted the IMM separately to data from individual participants in a VWM experiment (set size = 4), using a maximum-likelihood algorithm. Each simulation generated data for 20 subjects for a variable number of trials, N(trials) = 60, 120, 180, or 360. Parameter values of individual subjects were drawn from Gamma distributions with means set to the values of M(a), M(b), M(s), and M(κ) chosen for the current simulation, and the corresponding between-subject variances.

Figures 11 to 14 show the means and 95% HDIs of the posterior parameter estimates for the four group-level parameters, respectively. In each panel, the two not-varied parameters were set to an intermediate value. The simulation shows good recovery for parameters κ and *b*, but difficulties with recovering *a* (in particular at low values of *s*) and *s* (in particular at high values of *a*). This difficulty arises from the fact that both *a* and *s* account for nontarget intrusions, one (*s*) in a distance-dependent, the other (*a*) in a distance-independent way. The data from an experiment of typical size provide only very sparse information to

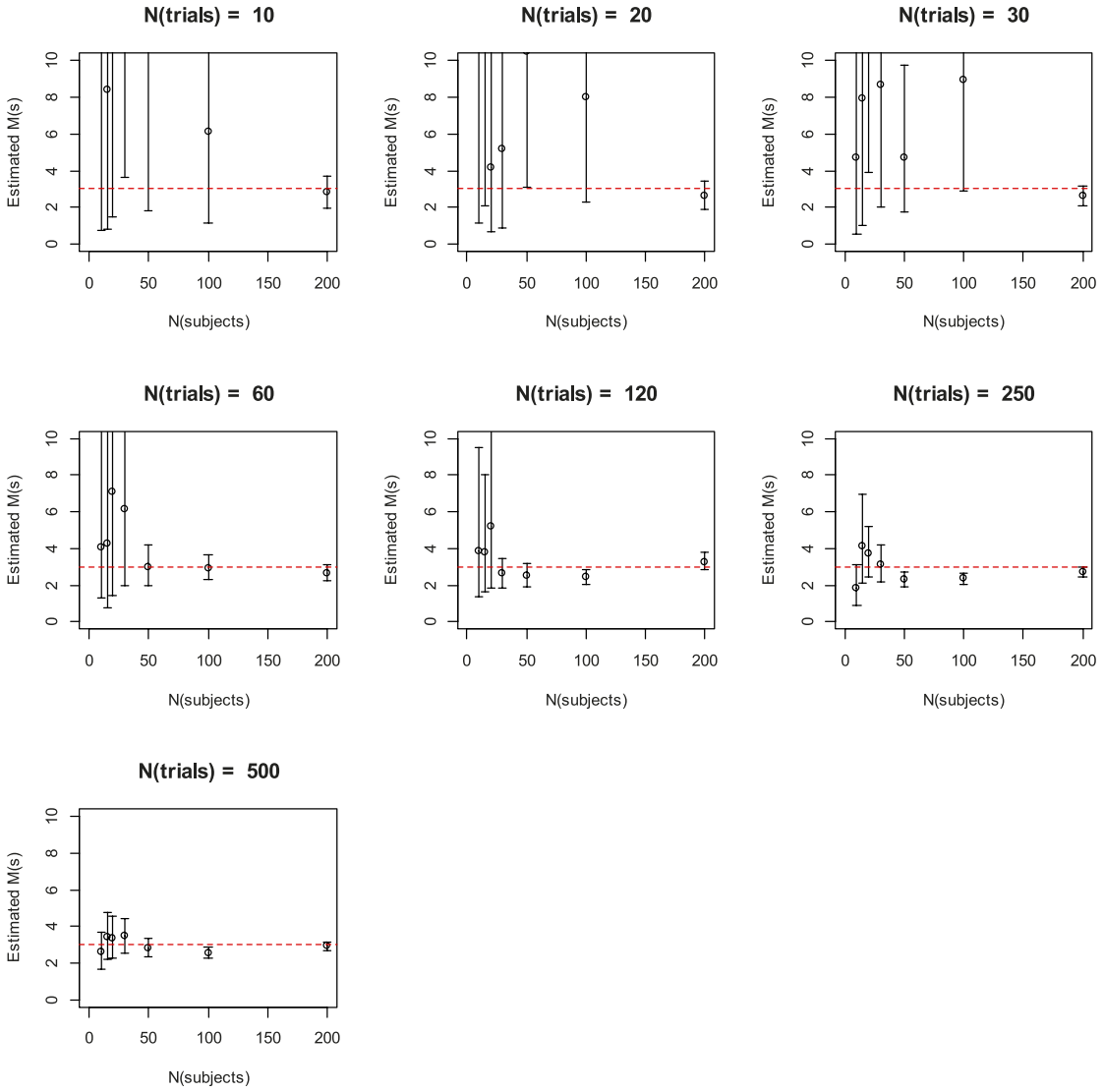

Figure 17. Recovery of the *M*(*s*) parameter of the IMM as a function of number of trials and number of subjects. The Figure shows the means and the 95% highest-density intervals of the posterior of the group-mean parameter. The broken red line is the true parameter value.

distinguish between distance-dependent and distance-independent intrusions.

For a given dataset, a simpler model may be more appropriate than the full IMM. IMM version *abc* (*s* constrained to 20) does not include distance-dependent intrusions of nontargets. Conversely, IMM version *bsc* (*a* constrained to 0) does not include distance-independent intrusions of nontargets. The most appropriate model out of these three can be determined via model comparison using a criterion such as the Deviance Information Criterion (DIC) (Spiegelhalter, Best, Carlin, & van der Linde, 2002) or the Widely Applicable Information Criterion (WAIC; Watanabe, 2010), which measure model fit for hierarchical models, taking model complexity into account. Following the advice of Gelman, Hwang, and Vehtari (2014), we recommend the WAIC because it is the only information criterion that is fully Bayesian. In addition, a simulation study comparing DIC and WAIC for the IMM found that WAIC is more stable by two orders of magnitude.[2] We also found that model selection using DIC recovered the true model (i.e., the full version of HIMM) in only half the simulation runs of Simulation 1, whereas model selection using WAIC recovered the true model in almost all simulation runs.

Simulation 2

The second simulation investigated how parameter recovery changed as a function of number of subjects and of number of trials per subject. We set the model parameters to typical values ($a = 0.2$, $b = 0.35$, $s = 3$, $\kappa = 10$) and varied the number of subjects (from 10 to 200) and the number of trials per subject (from 10 to 500).

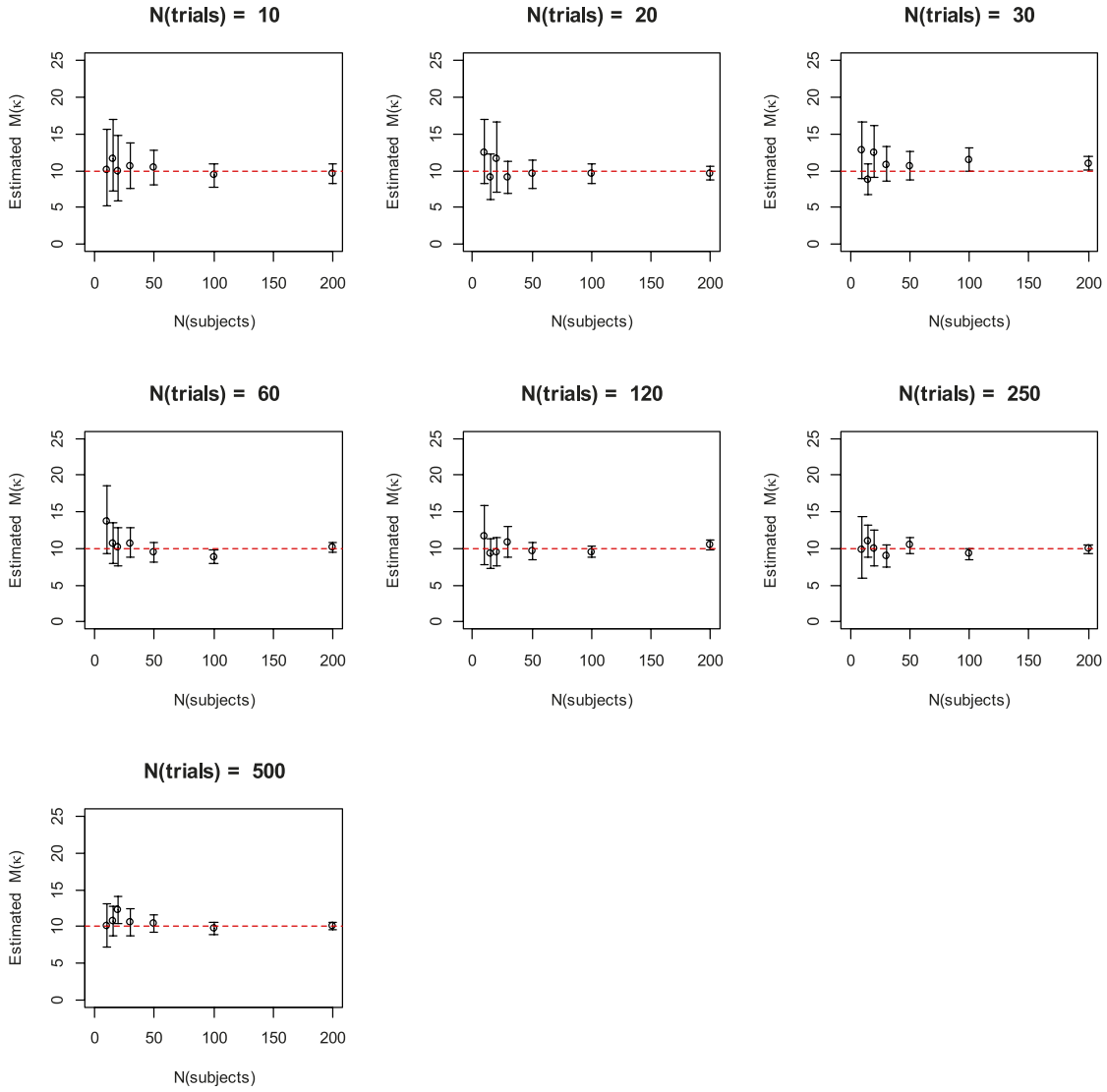

Figure 18. Recovery of the $M(\kappa)$ parameter of the IMM as a function of number of trials and number of subjects. The Figure shows the means and the 95% highest-density intervals of the posterior of the group-mean parameter. The broken red line is the true parameter value.

Figures 15 to 18 show the results. They confirm the difficulty of obtaining precise estimates of parameters a, b, and s with 20 subjects and less than 100 trials. The situation becomes much better, however, with 30 or more subjects and more than 100 trials. Moreover, the hierarchical modeling framework enables a trade-off between number of trials and number of subjects: Decent estimates can even be obtained with 10 trials when the sample size is increased to 200 subjects.

Application of the measurement models to experimental data

Here we apply the HMM and the HIMM to data from two experiments, first reported in Oberauer and Lin (2017). Experiment 1 tested 20 young adults on continuous reproduction of colors, varying set size from 1 to 8; each person completed 100 trials per set size. The colors of an array were displayed in a random subset of 13 equidistant locations on a virtual circle around the screen center. Arrays were presented for 100 ms, followed by a 1 s retention interval, after which a randomly selected object was tested by highlighting its location. Participants reproduced the target color on a color wheel with 360 colors (see Figure 19). We applied the three-parameter HMM and the four-parameter HIMM to the data of each set size separately. Figure 20 shows the resulting parameter estimates for the HMM, and Figure 21 those for the HIMM. Comparisons of model fits assessed by WAIC (Table 1) showed that the HMM won the competitions for most set sizes. Remarkably, the HMM had a systematically smaller

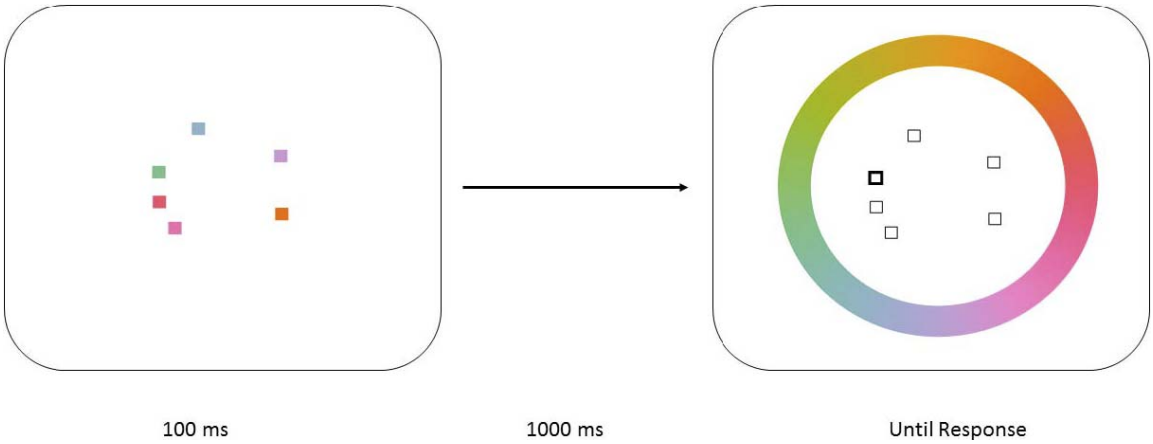

Figure 19. Example trial of Experiment 1 with set size 6.

WAIC than the constrained *abc* version of the HIMM, which is mathematically equivalent to the HMM (see Appendix 1). Apparently, the differences in how the models are parameterized affect their information criteria, as measured by WAIC. The WAIC is calculated from the likelihood of the data under the model, integrating over all possible parameter values weighted by their posterior probability. As such, it depends on the posterior distribution of parameter values, which in turn is influenced by their prior distributions. Because the two equivalent models are parameterized differently, we could not set equivalent

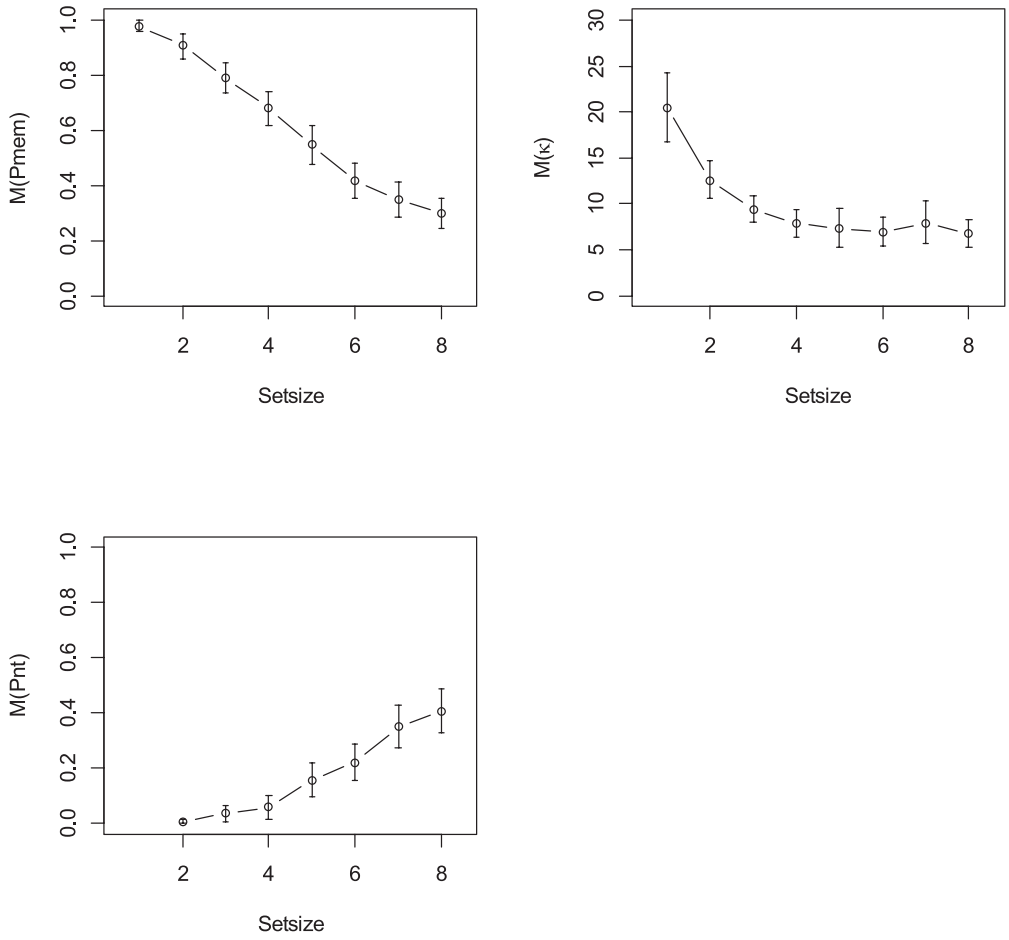

Figure 20. Estimated parameters (means and 95% HDIs of posteriors) of the hierarchical mixture model as a function of set size (Experiment 1).

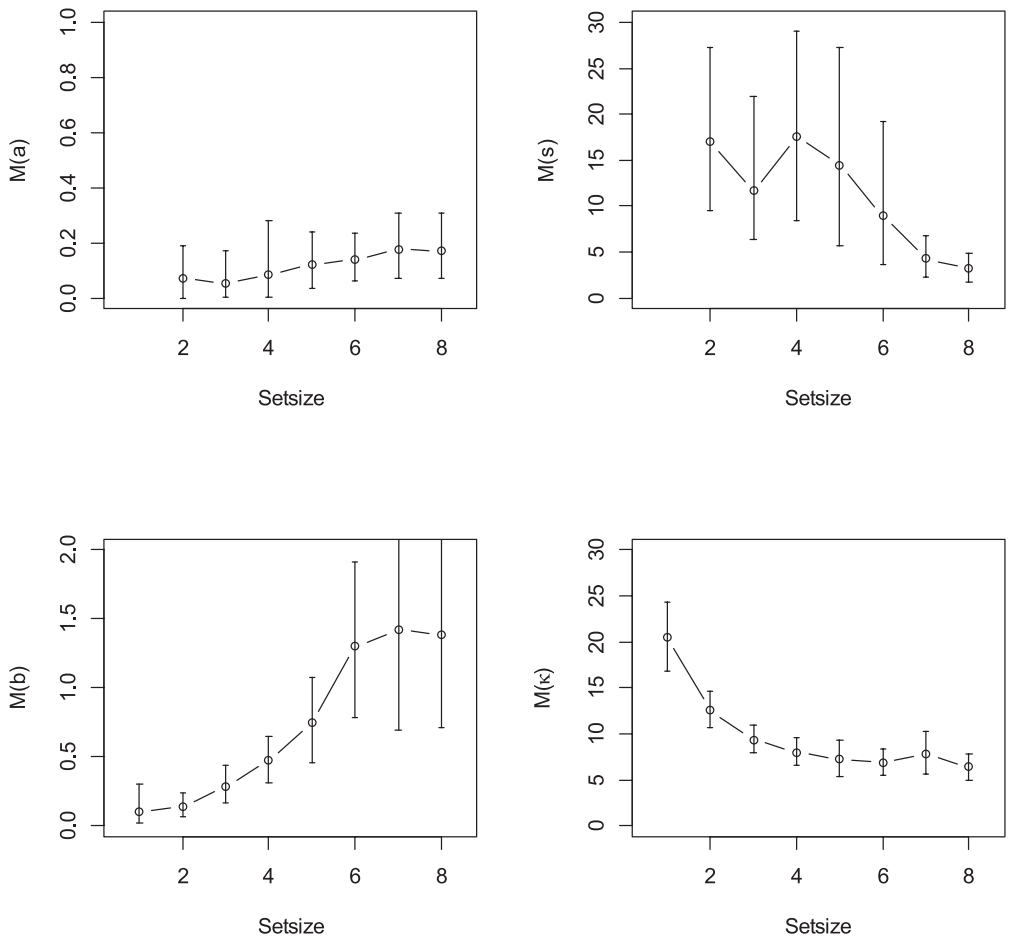

Figure 21. Estimated parameters (means and 95% HDIs of posteriors) of the hierarchical IMM as a function of set size (Experiment 1).

priors on their parameters, and as a consequence, comparing their WAIC values could be biased in favor of one model. Researchers interested in whether their data provide evidence for distance-dependent nontarget intrusions (as incorporated in the s parameter of the IMM) or evidence for distance-independent nontarget intrusions (as incorporated in the *a* parameter) should not compare the fit of the IMM to that of the mixture model but rather compare different versions of the IMM to each other.

Experiment 2 was designed to test the prediction of the IMM that nontarget intrusions become more prevalent the closer they are to the target in cue-feature space (i.e., the dimension that is used to identify the target). Twenty-one young adults worked on continuous reproduction of orientation at set size 6. Arrays consisted of six discs, each with a rectangular gap pointing in one of 360 angular directions, displayed in a random subset of 13 equidistant locations around the screen center, and rendered in six different colors randomly drawn from eight colors that were equidistant on a color circle. Arrays were presented for 1 s, followed by a 1 s retention interval. A randomly chosen object was probed for reproduction of orientation. In one condition the target was probed by its location, and in another condition it was probed by its color (see Figure 22). We fit the full HIMM and the constrained model version HIMM-*abc* (which eliminates the distance dependence of nontarget intrusions and is equivalent to the three-parameter mixture model) to each probing condition separately. For the full HIMM, we determined the distance between nontarget and target locations, $D(L_i, L_\theta)$, on the relevant cue-

Set Size	HMM	HIMM-full	HIMM-abc	HIMM-bsc
1	118	119	119	119
2	1537	1549	1555	1560
3	2762	2765	2789	2770
4	3710	3722	3853	3731
5	4426	4425	4584	4432
6	4788	4809	5077	4821
7	4636	4640	4809	4651
8	4729	4714	4940	4737

Table 1. WAIC values of the mixture model and the interference measurement model for Experiment 1. *Note*: Smaller WAIC reflects better fit.

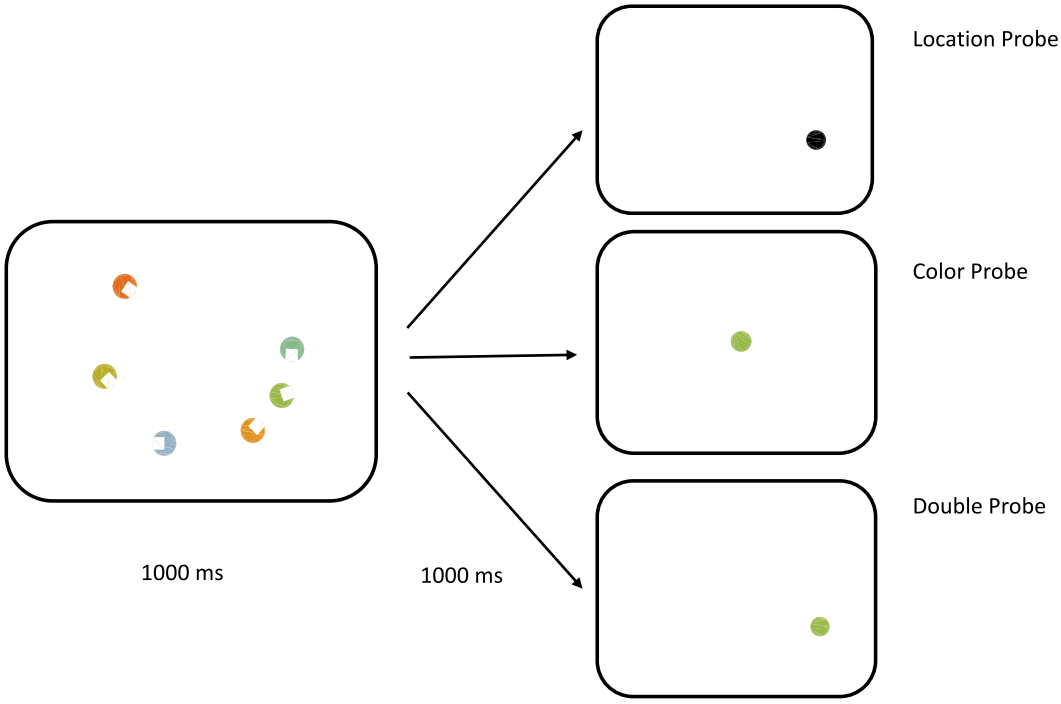

Figure 22. Example trial of Experiment 2 with three kinds of probes.

feature dimension for each condition: spatial location for the spatial-probe condition, and location in color space for the color-probe condition. Because both spatial location and colors varied on a circular scale, we measured them both in radians, so that the effects of distance on the two cue-feature dimensions can be compared with each other. For both probing conditions, the WAIC favored the full HIMM over the constrained HIMM-*abc* (color probe: WAIC for full HIMM = 5354, for HIMM-*abc* = 5398; for spatial probe: full HIMM = 6031, HIMM-*abc* = 6091). This is due to the fact that, in this data set, there is a pronounced distance gradient of nontarget intrusions: As the distance of nontargets from the target on the cue-feature dimension increased, the contribution of nontarget features to the response distribution declined. This can be seen in the histograms of responses, centered on the target and on each of the five nontargets, ordered by their distance from the target on the cue-feature dimension (Figure 23). Figure 24 shows the posteriors of the parameters of the full HIMM.

Discussion

The results presented in this article warrant methodological and substantive conclusions.

On the methodological side, we demonstrated that the popular mixture model for continuous-response data can be applied to data in a hierarchical Bayesian framework. Doing so has several advantages over the current practice of fitting data from each individual separately with maximum-likelihood techniques. One is that the hierarchical modeling framework pools information across all participants without disregarding individual differences. This enables a more robust estimate of group-level parameters. This advantage becomes most relevant when the number of trials per participant is small, or when a parameter is difficult to estimate because the data provide little information on it (e.g., the precision or *SD* parameter when *Pmem* is low). Another advantage is that Bayesian parameter estimation provides information not only about the location of a parameter (i.e., its best point estimate) but also on the credible range of values and the precision of estimation, as reflected in the highest-density intervals of the posterior distributions. More generally, the JAGS extension for the von Mises distribution enables researchers to investigate other measurement models, as well as explanatory models of continuous reproduction of features on circular dimensions (e.g., the full gamut of models compared by van den Berg et al., 2014) in a Bayesian hierarchical framework.

On the substantive side, we introduced an alternative measurement model based on assumptions that differ from those motivating the mixture model in three regards. First, whereas the mixture model builds on the assumption of discrete states in memory (i.e., an item is or is not available in memory), the IMM incorporates the idea that all items are represented in

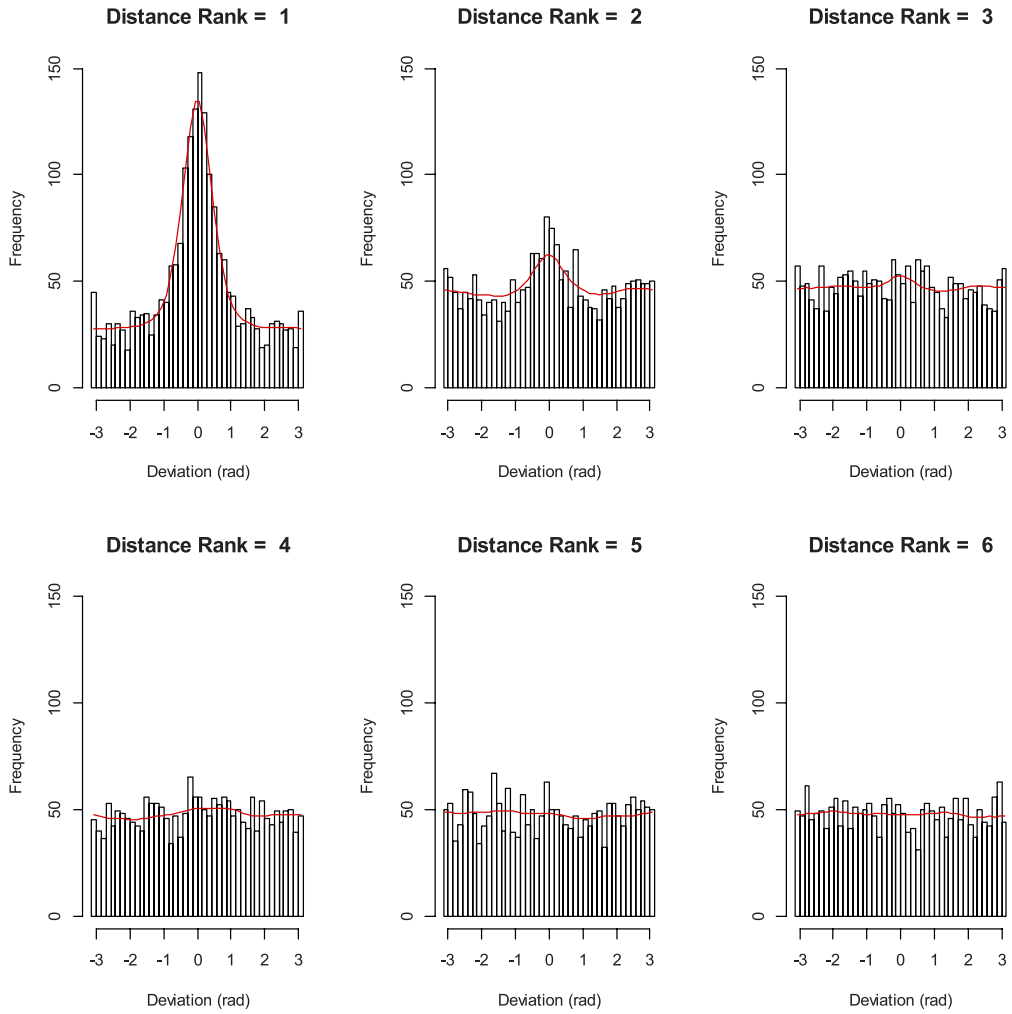

Figure 23. Histograms of responses centered around the target (top left, Distance Rank 1) and nontargets, ordered by their distance to the target on the cue-feature dimension (spatial location). Data are from the location-probe condition of Experiment 2. The red lines represent predictions of the IMM, obtained by binning samples of the posterior predictives in the same way as the data.

memory with a degree of strength varying on a continuum. Second, in contrast to the mixture model the IMM distinguishes between item information (i.e., remembering that an item has been in the current memory set) and binding information (i.e., remembering in which location an item has been presented). Third, the IMM formalizes principles of cue-based retrieval and interference, thereby providing a natural explanation for the occurrence of nontarget intrusions. Constrained versions of the IMM—versions *bac* and *bc*—*are* mathematically equivalent to the three-parameter and the two-parameter mixture model, respectively. This equivalence implies that any data fit well by the mixture model can also be fit well by a version of the IMM. Researchers can use one or the other model depending on which parameters they want to measure.

The IMM in its full version (and version *bsc*) makes one prediction that distinguishes it from the mixture model: Intrusions from nontargets should depend on

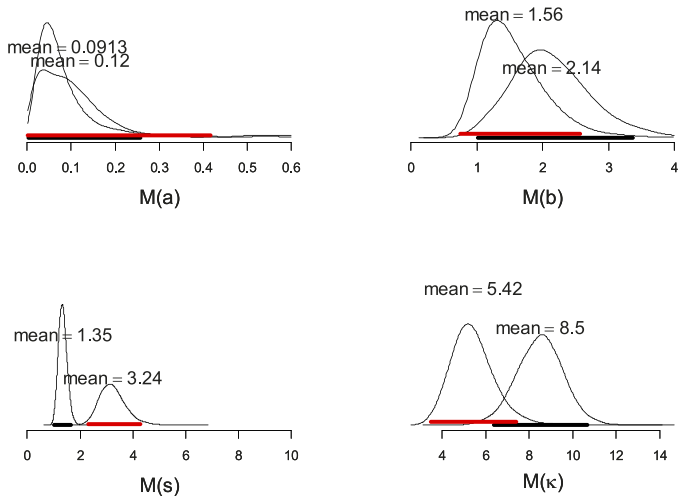

Figure 24. Posterior densities of group-level parameters of the IMM (full version) for Experiment 2. Thick horizontal bars are 95% highest-density intervals (black: color probes; red: location probes).

their proximity to the target on the cue-feature dimension. This prediction has received support in some previous experiments (Bays, 2016; Rerko, Oberauer, & Lin, 2014), and again in the experiments reanalyzed here (in particular Experiment 2). At the same time, the recovery simulations for the HIMM show that the signal of distance-dependent nontarget intrusions is likely to be weak in data from experiments with conventional numbers of participants and trials, and this was also noticeable in Experiment 1. Researchers interested in precise measurement of the parameters of the full IMM will have to invest more trials and/or participants than required for estimating the mixture-model parameters.

To conclude, we have demonstrated the feasibility of applying measurement models for continuous-response data within a hierarchical Bayesian framework. Researchers can choose between two measurement models, depending on their theoretical preferences and their goals. The mixture model can be used to obtain estimates of capacity and of precision as conceptualized in discrete-capacity theories, and to obtain an estimate of the prevalence of nontarget intrusions. The IMM is more data demanding, but in return enables a deeper analysis of the source of nontarget intrusions. In particular, the IMM can be used to measure the precision of representations not only on the target-feature dimension (parameter κ) but also on the cue-feature dimension (parameter *s*). Future research could use the IMM, for instance, to analyze data on memory for serial order of visual stimuli (Kool, Conway, & Turk-Browne, 2014). Experiments on serial-order memory for verbal stimuli have revealed a "locality constraint" on confusions between list items: The correct item in a given list position is more likely to be replaced by another item nearby in the list than by an item in a more distant position (Hurlstone, Hitch, & Baddeley, 2014). The locality constraint is an instance of distance-dependent nontarget intrusions on the dimension of list position. We could use the IMM—with list position as the cue-feature dimension—to ask whether serial recall of continuously varying visual features also shows this locality constraint. In this way, the IMM can help to bring the research traditions on working memory for visual and for verbal information closer together.

Keywords: working memory, *Bayesian hierarchical models*, *interference*

Acknowledgments

The research reported in this article was supported by a grant from the Swiss National Science Foundation (grant number 100014_126766) to the first author.

Commercial relationships: none.
Corresponding author: Klaus Oberauer.
Email: k.oberauer@psychologie.uzh.ch.
Address: University of Zurich, Department of Psychology—Cognitive Psychology, Zürich, Switzerland.

Footnotes

[1]The precision is related to the circular variance through

$$Var = 1 - \frac{I_1(\kappa)}{I_o(\kappa)}$$

[2]Both DIC and WAIC are calculated from the MCMC samples of the deviance (Gelman et al., 2014) and therefore are estimated with limited precision. We estimated the precision of DIC and WAIC by running the three versions of IMM on the same data (i.e., Experiment 1) 50 times, and calculating the *SD* of the information criteria across the 50 replications. The *SD* of DIC was 908, whereas the *SD* of WAIC was 10.

References

Bae, G. Y., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General*, *144*, 744–763, doi:10.1037/xge0000076.

Bays, P. M. (2016). Evaluating and excluding swap errors in analogue tests of working memory. *Scientific Reports, 6*, 19203, doi:10.1038/srep19203.

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10):7, 1–11, doi:10.1167/9.10.7. [PubMed] [Article]

Bays, P. M., Gorgoraptis, N., Wee, N., Marshall, L., & Husain, M. (2011). Temporal dynamics of encoding, storage, and reallocation of visual working memory. *Journal of Vision*, *11*(10):6, 1–15, doi:10.1167/11.10.6. [PubMed] [Article]

Brady, T. F., Konkle, T. F., Gill, J., Oliva, A., & Alvarez, G. A. (2013). Visual long-term memory has the same limit on fidelity as visual working memory. *Psychological Science*, *24*, 981–990, doi:10.1177/0956797612465439.

Brown, G. D. A., Neath, I., & Chater, N. (2007). A

temporal ratio model of memory. *Psychological Review*, *114*, 539–576.

Cowan, N. (2005). *Working memory capacity*. New York, NY: Psychology Press.

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin*, *53*, 134–140.

Fukuda, K., Awh, E., & Vogel, E. K. (2010). Discrete capacity limits in visual working memory. *Current Opinion in Neurobiology*, *20*, 177–182.

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, *24*, 997–1016, doi:10.1007/s11222-013-9416-2.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review*, *7*, 185–207.

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological Bulletin*, *140*, 339–373. doi:10.1037/a0034221.

Kool, W., Conway, A. R. A., & Turk-Browne, N. B. (2014). Sequential dynamics in visual short-term memory. *Attention, Perception & Psychophysics*, *76*, 1885–1901.

Kruschke, J. K. (2011). *Doing Bayesian data analysis: A tutorial with R and BUGS*. New York, NY: Academic Press.

Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology*, *55*, 1–7, doi:10.1016/j.jmp.2010.08.013.

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian modeling for cognitive science: A practical course*. Cambridge, UK: Cambridge University Press.

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience Reviews*, *17*, 347–356, doi:10.1038/nn.3655.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 103–113.

Oberauer, K., & Lin, H.-Y. (2017). An interference model of visual working memory. *Psychological Review*, *124*, 21–59.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-Effect Models in S and S-Plus*. Berlin, Germany: Springer.

Plummer, M. (2016). JAGS 4.2.0 [computer software]. Retrieved from http://mcmc-jags.sourceforge.net/

Prinzmetal, W., Amiri, H., Allen, K., & Edwards, T. (1998). Phenomenology of attention: 1. Color, location, orientation, and spatial frequency. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 261–282.

Rerko, L., Oberauer, K., & Lin, H.-Y. (2014). Spatially imprecise representations in working memory. *Quarterly Journal of Experimental Psychology*, *67*, 3–15, doi:10.1080/17470218.2013.789543.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, B*, *64*, 583–639.

Suchow, J. W., Fougnie, D., Brady, T. F., & Alvarez, G. A. (2014). Terms of the debate on the format and structure of visual memory. *Attention, Perception & Psychophysics*, *76*, 2071–2079, doi:10.3758/s13414-014-0690-7.

Surprenant, A. M., & Neath, I. (2009). *Principles of memory*. New York, NY: Taylor & Francis.

Sutterer, D. W., & Awh, E. (2016). Retrieval practice enhances the accessibility but not the quality of memory. *Psychonomic Bulletin & Review*, *23*, 831–841, doi:10.3758/s13423-015-0937-x.

van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review*, *121*, 124–149, doi:10.1037/a0035234.

Wabersich, D., & Vandekerckhove, J. (2014). Extending JAGS: A tutorial on adding custom distributions to JAGS (with a diffusion model example). *Behavioral Research Methods*, *46*, 15–28, doi:10.3758/s13428-013-0369-3.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*, 3571–3594.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12):11, 1120–1135, doi:10.1167/4.12.11. [PubMed] [Article]

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, *114*, 152–176.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*, 233–236.

Appendix 1: The equivalence between the mixture models and constrained versions of the interference measurement model

This Appendix demonstrates the equivalence between the interference measurement model (IMM) and the mixture model originally proposed as a measurement model by Zhang and Luck (2008), and subsequently expanded by Bays et al. (2009). When the spatial-gradient parameter s is fixed to a very high value, the resulting IMM version *abc* is equivalent to the three-parameter mixture model of Bays et al. (2009). In the IMM-*abc*, the distribution of activation is a sum of three components, weighted by parameters b, a, and c. In this sum, activation of the color distribution of the target comes from component weighted with c and the component weighted with a; the weight of this component is $c + a$. Activation of each of the n-1 nontargets comes only from component weighted by a, its joint weight is (n-1)a. Uniformly distributed activation for every color is added with weight b. To obtain the predicted probability distribution of responses, the activation distribution is normalized.

In the three-parameter mixture model, the probability distribution of responses is a sum of three components, weighted by parameters *Pmem* for target responses, *Pnt* for nontarget responses, and $Pg = 1 - Pmem - Pnt$ for random guessing. Thus, the only difference between the three-parameter mixture model and the IMM version *abc* is that in the former the normalization occurs before the weighted summation, whereas in the IMM it occurs afterwards. Therefore, the weights of the mixture model can be obtained from the weights of the IMM by normalization. To that end we need to divide the IMM weight parameters by their sum. For a trial with set size n, the sum of the weights equals $b + na + c$. Thus:

$$Pg = \frac{b}{b + na + c},$$
$$Pnt = \frac{(n-1)a}{b + na + c},$$
$$Pmem = \frac{c + a}{b + na + c}.$$

In both models, the precision parameter κ governs the dispersion of the activation distribution (IMM) or the response probability distribution (mixture model) over content space, which is unaffected by normalization. Therefore, the κ parameters of both models map directly onto each other.

With the additional constraint $a = 0$ we obtain the IMM version *bc*, which is equivalent to the two-parameter mixture model (Zhang & Luck, 2008). The correspondence of parameters is easily obtained by setting $a = 0$ in the equations above. Parameter *Pnt*, the probability of reporting a nontarget, drops out, resulting in the two-parameter mixture model in which $Pg = 1 - Pmem$.