HHS Public Access

Author manuscript

A Bayesian Nonlinear Mixed-Effects Location Scale Model for Learning

Donald R. Williams,
University of California, Davis

Daniel R. Zimprich,
Ulm University

Philippe Rast
University of California, Davis

Abstract

We present a Bayesian nonlinear mixed-effects location scale model (NL-MELSM). The NL-MELSM allows for fitting nonlinear functions to the location, or individual means, and the scale, or within-person variance. Specifically, in the context of learning, this model allows the within-person variance to follow a nonlinear trajectory, where it can be determined whether variability reduces during learning. It incorporates a sub-model that can predict nonlinear parameters for both, the location and scale. This specification estimates random effects for all nonlinear location and scale parameters that are drawn from a common multivariate distribution. This allows estimation of covariances among the random effects, *within* and *across* the location and the scale. These covariances offer new insights into the interplay between individual mean structures and intra-individual variability in nonlinear parameters. We take a *fully* Bayesian approach, not only for ease of estimation, but also for inference because it provides the necessary and consistent information for use in psychological applications, such as model selection and hypothesis testing. To illustrate the model, we use data from 333 individuals, consisting of three age groups, who participated in five learning trials that assessed verbal memory. In an exploratory context we demonstrate that fitting a nonlinear function to the within-person variance, and allowing for individual variation therein, improves predictive accuracy compared to customary modeling techniques (e.g., assuming constant variance). We conclude by discussing the usefulness, limitations, and future directions of the NL-MELSM.

Keywords

Nonlinear Mixed-effects Location Scale Model; Verbal Learning; Intra-Individual Variability; Bayesian inference

The ability to learn is of central importance throughout the lifespan and it serves as a main ingredient for long-term development of individuals, as well as their adaptation to changing environmental demands (Freund, 2008; Rast, 2011). While learning abilities are most evident in young age, they remain of importance into adulthood where they, however, are subject to decline with advancing age (Kausler, 1994). That is, most people will experience losses in their learning abilities, sometimes beginning in the mid-20s (Salthouse, 2009). At the same time, a defining feature of cognitive decline is potentially large and reliable individual differences therein. For example, some individuals will decrease dramatically in their ability to store and recall new information, but others will maintain remarkable levels of cognitive functioning (Harada, Love, & Triebel, 2013). Accordingly, describing and explaining these individual differences has generated a large volume of scientific interest over the last decades, because these differences might shed light on the mechanisms that allow for maintaining the ability to learn across the lifespan (Bush & Mosteller, 1955; Craik & Lockhart, 1972; Estes, 1950; Kausler, 1994; Rast, 2011; Rast & Zimprich, 2009; Z. Zhang, Davis, Salthouse, & Tucker-Drob, 2007; Zimprich, Rast, & Martin, 2008)

Typically, performance on a learning task improves with repetition. However, with every repetition the amount of performance improvement decreases. As a consequence, performance improvements, or learning of a task, may be described as a process that benefits from further practice, but with diminishing returns. If performance is diagrammed as a function of the number of practice repetitions, the so-called learning curve emerges that follows a gradually increasing, albeit negatively accelerated trajectory (Ritter & Schooler, 2001). The relation between performance improvement and repeated practice, as described in the learning curve, is so ubiquitous that it applies to a broad variety of performance increments in human behavior, including the acquisition of new skills (Ackerman, 1988), gaining knowledge of statistics (e.g., Smith, 1998) and, of course, verbal learning (Tulving & Madigan, 1970; Tulving & Pearlstone, 1966).

Generally, different forms of learning will lend themselves to different functions that reflect theoretical expectations, such as diminishing returns with repeated practice, resulting in an approximation of an idealized asymptote of maximal learning. As such, a number of nonlinear models have been proposed that capture these distinct features. For example, Heathcote, Brown, and Mewhort (2000) discussed an exponential curve for skill acquisition while Mazur and Hastie (1978) put forward a hyperbolic function for free recall, perceptual, and motor learning. Yet another approach, the power curve, was discussed by Logan (1988) and in its more geneearl form by Newell and Rosenbloom (1981). More recent work has focused on refining some of these classic learning functions. For example, N. J. Evans, Brown, Mewhort, and Heathcote (2018) added an additional parameter to the power and the exponential function to address phases of little change early in the learning process.

The different nonlinear functions mentioned above do not only affect the curvature of learning trajectories, but they also have important theoretical implications as data-generative models. For example, as detailed by Mazur and Hastie (1978), the exponential curve may be interpreted as being based on a "replacement model" of learning. It suggests that learning is a process through which incorrect response tendencies are replaced by more and more correct response tendencies. The exponential model implies a constant learning rate relative

to the amount left to be learned and the replacement process is assumed to occur at a constant rate. By contrast, the hyperbolic curve is based on an "accumulation model" of learning. According to the accumulation model, learning is a process by which correct response tendencies increase steadily with practice and compete with incorrect response tendencies, which remain constant across trials. Unlike the exponential model, the amount of accumulation per trial is considered a constant proportion of the amount or duration of the study. Finally, the power curve is based on the assumption that "…some mechanism is slowing down the rate of learning" (Newell & Rosenbloom, 1981, p. 18). Thus, if learning follows a power law, learning slows down across trials.

The common element among these models is that they contain parameters for lower and upper boundaries and a parameter that relates the speed of learning accumulation, or learning rate, that connects the lower to the upper boundary (Anzanello & Fogliatto, 2011). Fast learners will typically have steep initial increases in performance, but a slower accumulation once maximal performance has been attained, resulting in a markedly curved learning trajectory (Rast, 2011). Slow learners, on the other hand, tend to produce more shallow trajectories that may appear linear. As a consequence, they will need more trials to master the same material, compared to fast learners, which results in learning curves that appear considerably less "bent" (Feldman, Cao, Andalib, Fraser, & Fried, 2009; Ram, Stollery, Rabbitt, & Nesselroade, 2005).

Nonlinear models are not the only approach for modeling learning data. In fact, polynomial models are a common choice due to their flexibility and ease of implementation (Blachstein & Vakil, 2016; Jones et al., 2005). However, these models are typically limited to retrodiction: that is, they do not allow for extrapolation beyond the observed data range (Bonate, 2009). Any extrapolation beyond the observed range typically results in useless predictions as they will be dominated by the leading term of the polynomial equation. More importantly for applied researchers, higher order polynomial models result in parameters that are difficult to interpret and that are highly contingent on data handling features such as centering choices (Dalal & Zickar, 2012). Nonlinear models, in turn, often result in more interpretable parameters and, if the functions represent a theoretical model, they allow for extrapolations that result in more meaningful predictions compared to polynomial models (Bonate, 2009; Grimm, Ram, & Hamagami, 2011)

While most work revolved around individual differences, in recent years the focus has widened to include differences within individuals (typically over time or across similar situations) as a source of information on the stability of a persons' behavioral system (Brose, Voelkle, Lövdén, Lindenberger, & Schmiedek, 2015; Rast, Hofer, & Sparks, 2012; Sliwinski, Hoffman, & Hofer, 2010; Voelkle, Brose, Schmiedek, & Lindenberger, 2014). The central idea is that within-person variability should not be regarded as reflecting mere measurement error but that it conveys systematic information that is not contained in individual means. These ideas are not new to psychology as they have been discussed for almost a century now (Cattell, Cattell, & Rhymer, 1947; Fiske & Rice, 1955; Horn, 1972; Woodrow, 1932). However, only over the past 15 to 20 years has interest in intra-individual variability (IIV) increased (Eizenman, Nesselroade, Featherman, & Rowe, 1997; Hultsch, Hertzog, Small, McDonald-Miszczak, & Dixon, 1992; Nesselroade & Salthouse, 2004),

probably also due to the availability of abundant within-person-level data, and it is now commonly used to describe the number of reversible, short-term behavioral fluctuations that are observed over time (Ram & Gerstorf, 2009). These fluctuations can also occur in different situations and are interpreted to carry information about short-term adaptive processes, regulative mechanisms, and potential vulnerability to the system (Nesselroade, 1991; Röcke & Brose, 2013).

An important assumption about IIV is that it reflects other aspects of behavioral outcomes compared to individual levels or rates of change, such as individual means (Fagot et al., 2018, *iM*). Also, from a theoretical perspective, IIV is given the potential to be both the explanans and explanandum. For example, two individuals may perform, on average, equally well on a task but their performance pattern on single items within the task might be quite different (Garrett, MacDonald, & Craik, 2012). Such differences in the distribution of scores about the average performance are considered to reflect a meaningful aspect of individual performance and behavior, for example, in the prediction of cognitive decline (Lövdén, Li, Shing, & Lindenberger, 2007).

In terms of statistical modeling, mixed-effects models have been a frequent choice as they partition between- and within-person variability and thus allow for clustering effects, common in repeated measurements, such as learning tasks (Chu et al., 2007). Recently, mixed-effects models have been expanded to mixed-effects location scale models (MELSM; Hedeker, Mermelstein, & Demirtas, 2008; Watts, Walters, Hoffman, & Templin, 2016) that allow the inclusion of submodels for both the between- and within-person variance – or the location and the scale of the generating data process. With these models it is possible to model and estimate between and within person differences simultaneously. So far, the MELSM has been, for the most part, defined and used for intrinsically linear applications (Hedeker, Mermelstein, & Demirtas, 2012; Leckie, French, Charlton, & Browne, 2014; Rast & Ferrer, 2018). While Hedeker, Demirtas, and Mermelstein (2009) and Walters (2015) discussed a MELMS for a logit response, we are currently only aware of one inherently non-linear application based on the SITAR age-shifted model for adolescent growth (Cole, Donaldson, & Ben-Shlomo, 2010) described by Goldstein, Leckie, Charlton, Tilling, and Browne (2018). In that context the authors describe a MELSM that includes the non-linear term in the location part of the model. In this article, we present an approach to address and explain between- and within-person differences in verbal learning simultaneously using nonlinear models for *both* the location and the scale. Hence, the aim of this paper is to define nonlinear mixed effect models in the context of MELSM to accommodate nonlinear within-person variance changes, conditional on explanatory variables. By doing so, we extend the MELSM to inherently nonlinear applications, both in the location and scale.

From a substantive perspective, we build on previous work that investigated individual differences in learning but we add the possibility to address and model within-person variability with a mixed-effects (non)-linear sub-model resulting in the same benefits in terms of interpretability and theoretical consistence as described earlier. Hence, the approach presented here not only allows for describing within-person variance changes with nonlinear forms, but it also adds the potential to assess covariances among nonlinear location and scale parameters, thereby providing novel information for inference in psychological applications.

The remainder of the manuscript is organized as follows. First, we formally describe a general nonlinear MELSM (NL-MELSM). Second, we provide an example involving empirical data where three age groups participated in five learning trials assessing verbal memory. In our experience, classical methods (e.g., expectation-maximization) cannot readily estimate NL-MELSM's yet due to the nonlinear error variance structure which violates the assumption of normally distributed residuals thus making convergence very unlikely. Thus, hypothesis testing and model comparison with *p*-values are not currently feasible. As such, we take a *fully* Bayesian approach that does not only make estimation possible, but also provides the necessary tools for use in substantive applications. For example, we use leave-one-out cross validation LOOCV for model comparison and Bayes factors for hypothesis testing at the level of the individual parameters. Third, we discuss the findings in the context of the learning literature. Fourth, while introducing Bayesian inference is beyond the present scope, we provide several references to complement this work. We end by listing shortcomings as well as possible extensions of the NL-MELSM model.

Nonlinear Mixed-Effects Location Scale Model

In this section, we describe a generic nonlinear mixed-effects model, in addition to the customary MELSM that generally assumes linearity. We then describe how the models can be combined, resulting in the proposed NL-MELSM. Along with the specification, we also discuss practical challenges (i.e., convergence issues) and advantages of this model. For example, while describing the covariance structure among location and scale random effects, we relate this to psychological applications. In the following section, using this general notation, we fit a NL-MELSM with a specific nonlinear function.

Nonlinear Mixed-Effects Model

A generic nonlinear mixed-effects (NLME) model can be defined as

$$y \sim \mathrm{N}\left(\mu_i, \sigma^2\right),$$

$$(1)$$

$$\mu_i = f\left(x_{it}, \psi_i\right),$$

where the location μ is assumed to follow some nonlinear function *f*. Here x_{it} is a matrix of individual-specific values. For example, in the context of learning or longitudinal models, x_{it} may denote the elapsed time, scaled as learning trials or recall time points $t \in (1, \dots T_i)$, respectively, for the *i*th individual. Of course, depending on the research question, x_{it} could be any variable that is expected to describe the outcome *y* with some function *f*. The nonlinear parameters are then

$$\psi_{ik} = \mathbf{X}\phi_k + \mathbf{Z}_i b_i, \ k \in \left(1, \ldots, K_\mu\right),$$

(2)

which follows the customary notation for linear mixed-effects models. Here $\mathbf{X}$ and $\mathbf{Z}$ are the fixed and random effect design matrices, with K denoting the number of nonlinear parameters to be estimated. For K, we have used the subscript μ, which denotes that these nonlinear parameters are specifically for the location, or mean structure. The random effects b_i are then assumed to follow

$$b_i \sim \mathrm{N}(0, \boldsymbol{\Theta}),$$

(3)

where b_i denotes the individual deviations from the population average. This distribution has mean zero and and a $K_\mu \times K_\mu$ dimensional covariance matrix $\boldsymbol{\Theta}$. The covariances $\theta_{ij} \in \boldsymbol{\Theta}$ can be restricted to be independent of one another, or a maximal model can be fitted in which all of the covariances between random effects are estimated (Barr, Levy, Scheepers, & Tily, 2013). This is often determined by the complexity of the model, since convergence issues can arise when fitting fully parametrized NLME models (Comets, Lavenu, & Lavielle, 2017; Hall & Bailey, 2001).

Mixed-Effects Location Scale Model.—The traditional NLME model, described in Equation (2), is known for its flexibility and conservative use of parameters compared to nth degree polynomial linear models. However, to our knowledge, an approach to explicitly model the within-person variance $e_{it} \sim N(0, \sigma^2)$ with its own functional form, in addition to allowing for individual variation therein, has not been described. In the context of linearity, this is known as mixed-effects location scale model (MELSM; Hedeker et al., 2008; Leckie et al., 2014; Rast et al., 2012), which is defined as

$$y_i \sim \mathrm{N}\left(\mu_i, \sigma_i^2\right),$$

(4)

$$\mu_i = \mathbf{X}_i \beta + \mathbf{Z}_i b_i,$$

$$\sigma_i^2 = \exp\left(\mathbf{W}_i \eta + \mathbf{V}_i u_i\right).$$

Here, rather than assuming that the residuals have a constant variance σ^2, they are described by their own mixed-effects model. The design matrices, $\mathbf{W}$ and $\mathbf{V}$, are thought to predict σ_i^2 with the fixed effects η and random effects u_i. For example, when predicting change over time, η can be the fixed effect of the time metric, which is then allowed to vary among individual (i.e., random intercepts and/or slopes). In other words, this model essentially treats the residual variance as the response variable. With computational advancements, this model has become increasingly popular in both applied (Watts et al., 2016) and methodological literatures (Lin, Mermelstein, & Hedeker, 2018; Rast & Ferrer, 2018; Walters, Hoffman, & Templin, 2018). However, to our knowledge the MELSM has not been extended to allow for fitting nonlinear functions to the location μ and the scale σ_i^2. The contribution of the current paper is to show the generality of the MELSM by bringing together both nonlinear and location scale mixed- effects models into one coherent framework.

Within-Person Variance.—We extend the traditional nonlinear mixed-effects model, described in Equation 2, by allowing σ_i^2 to have its own functional form

$$\sigma_i^2 = \exp\left[f\left(z_{it}, \zeta_i\right)\right],$$

(5)

where ζ_i denotes the nonlinear parameters for the ith individual. The entire function f is exponentiated to ensure that the variance is restricted to positive values $\left(0 < \sigma_i^2\right)$. This function can either be the same as that fitted to the location or a different functional form can be assumed. For example, it is possible to describe μ_i with an exponential function, whereas σ_i^2 can follow a Weibull growth curve. The fixed and random effects for predicting σ_i^2 are then defined as

$$\zeta_{ik} = \mathbf{W}\gamma_k + \mathbf{V}_i u_i, \quad k \in \left(1, ..., K_{\sigma^2}\right),$$

(6)

where K is the number of nonlinear parameters to be estimated. Note that in this model σ_i^2 is a response variable that is simultaneously being estimated with the location μ_i. This is a particularly important feature because such within-person variance can be examined in

relation to variables external to the system (Rast & Ferrer, 2018). For example, a researcher interested in learning may want to use variables to predict the (in-)consistency in learning. In addition, there might be variables external to the modeled system (e.g., age, learning strategy, etc.) that could potentially explain part of the variance in learning performance that are not accounted for by the model fitted to the location (Equation 2).

Random Location and Scale Effects.—With the introduction of a mixed-effects model for the within-person variance, we also allow for random effects to be estimated simultaneously for the location and the scale part. By estimating these two components within the same NLME model, we are able to account for possible covariances that arise among iMs and iSDs, which ensures that we can make valid inferences about our parameter estimates (Verbeke & Davidian, 2009). From a substantive perspective, in the context of learning, these models allow for determining the relationship among individual location and scale parameters. For example, one might investigate whether individuals with faster learning rates tend to show smaller within-person variance in their asymptote. Of course, this requires allowing for covariances between the random location b_i and scale effects u_i, respectively. Let us assume nonlinear functions for the location (μ_i) and scale $\left(\sigma_i^2\right)$, which result in a $\left(K_\mu + K_{\sigma^2}\right) \times \left(K_{\sigma^2} + K_\mu\right)$ covariance matrix Σ. The random effects for both the location and the scale come from the same multivariate distribution

$$\begin{bmatrix} b_i \\ u_i \end{bmatrix} \sim N(0, \Sigma)$$

with mean zero. Following common practice (Barnard, McCulloch, & Meng, 2000), the covariance matrix can then be re-expressed as

$$\Sigma = \tau \Omega \tau',$$

(7)

where Ω is a $\left(K_\mu + K_{\sigma^2}\right) \times \left(K_{\sigma^2} + K_\mu\right)$ correlation matrix and τ is the diagonal of a matrix with the same dimension containing the random effect SDs. This parameterization offers several advantages compared to hierarchical Bayesian models that depend on conjugate priors, such as, for example the inverse Wishart distribution (Barnard et al., 2000). The approach of partitioning the covariance matrix into a diagonal matrix of SDs and a correlation matrix is more efficient and numerically stable (Rast & Ferrer, 2018), in addition to more intuitive prior specification on the size of the correlations, it readily allows for Bayesian hypothesis testing via Bayes factors (Marsman & Wagenmakers, 2017).

For clarity, suppose we have two nonlinear parameters for the location ($K_\mu = 2$) and scale $\left(K_{\sigma^2} = 2\right)$. This parametrization, described in Equation (7), would result in a 4×4 correlation matrix

$$
\boldsymbol{\Omega} = \begin{bmatrix} 1 & & & \\ \phi_1, \phi_0 & 1 & & \\ \gamma_0, \phi_0 & \gamma_0, \phi_1 & 1 & \\ \gamma_1, \phi_0 & \gamma_1, \phi_1 & \gamma_1, \gamma_0 & 1 \end{bmatrix}.
$$

(8)

Here we have depicted the lower-triangle. The off-diagonal entry at $\boldsymbol{\Omega}_{2,1} = \phi_1,\ \phi_0$ denotes the correlation between the location random effects while the $\boldsymbol{\Omega}_{4,3} = \gamma_1,\ \gamma_0$ captures the correlation among the scale random effects. Notably, the four other elements $\boldsymbol{\Omega}_{3:4,1:2}$ correspond to the correlations among the location ϕ and scale γ random effects.

A Nonlinear Mixed-Effects Location Scale Model for Learning

In this section, using the generic specification above, we describe a specific NL-MELSM applied to learning data. The assumed functional form is the asymptotic parameterization that is described in Pinheiro and Bates (2006). We thus extend the asymptotic nonlinear mixed-effects model for the location to modeling the within-person variance with the same nonlinear function, in addition to allowing for individual variation in the location and scale parameters. Further, we include age groups as a predictor for the nonlinear parameters in the NL-MELSM (Section Predicting Nonlinear Parameters).

Illustrative Data

Study Participants.—Study participants were recruited in the Nuremberg and Erlangen Metropolitan regions. The total was $N = 333$, including three age groups young ($n = 111$, 56 female), mean age 21.5 (SD 1.72), middle-aged ($n = 111$, 56 female), mean age 46.9 (SD 2.62), and old ($n = 111$, 56 female), mean age 73.8 (SD 2.28).

Verbal learning.—Verbal learning (VL) was assessed by five consecutive trials of a word list recall task. The task comprised 27 meaningful, but unrelated, two- to three-syllabic German words taken from the Handbook of German Word Norms (Hager & Hasselhorn, 1994). The words appeared on a computer screen at a rate of 2 seconds each, and participants had to read them aloud. After the presentation of all 27 words, participants recalled as many words as possible in any order. This procedure was repeated five times using the same list of 27 words but with a different order of word presentation for each trial. The number of correctly recalled words was scored after each trial, which could range between 0 and 27.

Model Specification

The learning process was modeled with an asymptotic function that has three parameters to be estimated

$$f(\text{trial}, \phi) = \phi_0 + (\phi_1 - \phi_0) \times \exp\left[-\exp(\phi_2) \times \text{ trial}\right].$$

(9)

Here, trials is scaled as $0, \ldots, T-1$ which reduces Equation (9) to ϕ_1, the initial performance when trial = 0. The parameter ϕ_0 is the upper asymptote of recall performance as the number of trials $\rightarrow \infty$ 1, and ϕ_2 refers to the rate of learning which also defines the curvature in the learning trajectories. The asymptotic function described here and reparameterizations thereof have been used previously to model learning related data (e.g. Heathcote et al., 2000; Rast, 2011). If one considers nonlinear functions from a theoretical perspective, functions with asymptotic boundaries are compatible with learning processes in which limiting behavior is expected. That is, common performance in VL increases with every additional presentation of the stimuli. However, as known from "testing the limits" studies (Baltes & Kliegl, 1992), after a certain number of trials the performance remains constant, and increasing the number of trials will not improve performance. This behavior is best modeled by nonlinear functions, which incorporate an upper asymptote that captures the theoretical limit for individual learning. Note however, that these are not, by any means, the only functions to model learning data. Especially in situations where learning is delayed initially but then accelerates quickly (Rickard, 2004), standard exponential models are not well suited to capture the learning process. For example, N. J. Evans et al. (2018) formulated a delayed exponential model with an additional parameter to capture initial delay in learning. Hence, it is important to emphasize that Equation (9) can take any nonlinear form. The selection of an appropriate functional form will depend on the underlying process or may result from an exploratory comparison among different candidate functions.

In a mixed model framework, assuming that the nonlinear parameters vary between individuals, Equation (9) is extended to

$$y_{it} = \phi_{0i} + (\phi_{1i} - \phi_{0i}) \times \exp\left(-\exp(\phi_{2i}) \times \text{ trial}_{it}\right) + \epsilon_{it}.$$

(10)

Rather than assume the customary $\epsilon_i \sim N(0, \sigma^2)$ for the within-person variance σ^2, we can instead assume that it follows a nonlinear function as well and that it takes different values across measurement occasions t and indvidual i. In other words, we treat σ_{it}^2 as a person and time-varying response variable to be predicted by a set of nonlinear parameters. While it is

possible to allow the nonlinear functions (e.g., exponential vs. weibull) to differ between the location and the scale, we assume

$$\sigma_{it}^2 = \exp\left[\gamma_{0i} + (\gamma_{1i} - \gamma_{0i}) \times \exp(-\exp(\gamma_{2i}) \times \text{trial}_{it})\right],$$

(11)

where σ_{it}^2 contains all error variances for individuals i and for each each trial t. Furthermore, the nonlinear parameters $\gamma \in (1, \ldots, 3)$ have the same interpretation as in equation (9) that describes the location. Of course, an interesting extension is to then predict ϕ and γ with explanatory variables. That is, one can assess hypothesized group differences in the nonlinear parameters. Without loss of generality, this results in the following specification

$$\phi_{ki} = \mathbf{X}_i\beta + \mathbf{Z}_ib_i, \quad k \in (1, \ldots, 3),$$

(12)

$$\gamma_{ki} = \mathbf{W}_i\eta + \mathbf{V}_iu_i, \quad k \in (1, \ldots, 3),$$

where $\mathbf{X}$ and $\mathbf{W}$ are the fixed effect design matrices that are thought to influence the nonlinear parameters (ϕ and γ), whereas $\mathbf{Z}$ and $\mathbf{V}$ are the random effect components. In the present example, for demonstrative purposes, we first assume the design matrices contain a column of 1's ($n \times 1$), thereby denoting the population averages for each nonlinear parameter. Importantly, this general NL-MELSM, herein referred to as NL-MELSM 1, can easily be extended to include explanatory variables thought to predict the nonlinear parameters (Section Predicting Nonlinear Parameters).

Although a common approach for building mixed effect models is to use some variety of stepwise selection (Zuur et al., 2007), we take the approach described in Barr et al. (2013). Here we begin by fitting the maximal covariance structure, in which all random effects are allowed to correlate with one another. This allows for not only estimating the correlations between random location and scale effects, but also correlations between the two, in addition to limiting data dependent decisions (Gelman & Loken, 2014). For the asymptotic function, fit to both the location and scale, there are 6 nonlinear parameters in total which results in

$$\begin{bmatrix} b_i \\ u_i \end{bmatrix} \sim N(0, \boldsymbol{\Sigma}),$$

$$\Sigma = \tau \Omega \tau'$$

with the means as zero and the covariance defined as Σ. We re-expressed Σ as $\tau \Omega \tau'$, where Ω is a $(q + p) \times (q + p)$ correlation matrix and τ are the SDs. The estimation of the full covariance structure can be challenging, even for *state-of-the-art* Bayesian software, such as Stan. As such, to ensure good quality of the parameter estimates, convergence of the MCMC algorithm needs to be monitored closely.

Prior Specification.—In order to keep the parameter estimates within a reasonable range, we specified priors that were informed by previous work, by prior predictive distributions and by the data. Since the nonlinear parameters have clear meanings, for example ϕ_1 as learning at trial one, we therefore decided upon the priors by visualizing the data and referring to past research (Rast, 2011). This was accomplished by not only looking at the observed data, but also by examining the prior predictive distribution. For the latter, without conditioning on the data, hypothetical datasets were replicated from the assumed prior distributions to ensure adequacy for the goal at hand–convergence with minimal informativeness. While beyond the scope of this work, we refer to the following references for visualizing Bayesian models (including prior predictive checks; Gabry, Simpson, Vehtari, Betancourt, & Gelman, 2019), checking prior-data conflict (M. Evans & Moshonov, 2006), and Gelman, Simpson, and Betancourt (2017) that provides several ways to think about prior specification.

We first describe the assumed priors for the fixed effects, where

$$\phi_0 \sim N(25, 2),$$

(13)

$$\phi_1 \sim N(8, 2),$$

$$\phi_2 \sim N(-1, 2),$$

denote the location parameters. Specifically, ϕ_0 denotes the upper asymptote of 25, with a scale of 2, which was chosen with respect to the maximum score (i.e., 27) possible on the recall task. On the other hand, the prior on ϕ_1 reflects recall at trial 1, whereas the scale encoded a 95% prior probability in the range of $4 - 12$. This choice was informed based on past research, for example Rast (2011), where the reported recall at trial 1 was 6.27, so we did not want to be overly restrictive. The learning rate ϕ_2 is not straightforward to infer from visualizations. We thus centered the prior distribution based on Rast (2011), but the wide scale ensured the prior was minimally informative. We viewed this as a reasonable choice,

since the the rate of learning is of primary interest. The priors for the scale parameters followed

$$\gamma_0 \sim N(0, 1),$$

$$(14)$$

$$\gamma_1 \sim N(1, 1),$$

$$\gamma_2 \sim N(0, 1),$$

which have the same interpretation as for the location parameters. We justify these priors, *a priori*, as not expecting substantial variability in the nonlinear parameters fitted to the within-person variance. However, to ensure there is some benefit from fitting a nonlinear function to the scale, we provide a model comparison to a linear function (Section: Model Selection).

The random effect SDs also require a prior distribution, which were similarly chosen to to be minimally informative. We denote the random location effect SDs $\tau_{\phi_1, \ldots, \phi_3}$, whereas the random scale effects SDs are denoted $\tau_{\gamma_1, \ldots, \gamma_3}$. The assumed priors are then

$$\tau_{\phi_1, \ldots, \phi_3} \sim \text{Student-t}(3, 0, 5),$$

$$(15)$$

$$\tau_{\gamma_1, \ldots, \gamma_3} \sim \text{Student-t}(3, 0, 1),$$

$$\Omega \sim \text{LKjcorr}(\eta = 1),$$

where LJKcorr is the Lewandowski, Kurowicka, and Joe prior on the correlation matrix (Lewandowski, Kurowicka, & Joe, 2009). This distribution for LKJcorr is governed by a single parameter η, where a value of one places a uniform prior over all correlation matrices, assuming a 2×2 matrix. In larger dimensions, even with $\eta = 1$, there is shrinkage towards an identity matrix, due to the positive definiteness restriction. This model serves as the foundation for the remainder of this work (NL-MELSM 1), and all proceeding models

(Section: Model Selection) are either more restricted special cases (thus all the priors remain the same), or we implicitly assume *so-called* non-informative prior distributions. The latter choice was explicitly made for simplicity, such that the primary focus remains on introducing the NL-MELSM.

Estimation and Software.—The fitted models included four chains of 5,000 iterations, excluding a warm-up period of the same size. This number of iterations provided a good quality of the parameter estimates in which the models converged with potential scale reduction factors $\hat{R}$ smaller than 1.1 (Gelman, 2006). As measures of relative model fit, we report the deviance and the Pareto smoothed importance sampling-Leave-one-out cross-validation (LOO) with the corresponding standard errors (Vehtari, Gelman, & Gabry, 2017). The fitted models were evaluated based on the their respective differences in the values of LOO and standard errors. Additionally, we also computed pseudo-BMA+ weights, describe in Yao, Vehtari, Simpson, and Gelman (2017), which provide a similar interpretation as Akaike weights (Wagenmakers & Farrell, 2004). That is, the probability that a given model will provide the most accurate out-of-sample predictions. Alternative Bayesian approaches for variable selection are described elsewhere (O'Hara & Sillanpää, 2009). Further, it is customary to evaluate nonlinear effects with significance tests which are not possible for the present model. For demonstrative purposes we used Bayes factors (BF_{10}), with the Savage-Dickey ratio, to evaluate evidence for not only non-zero effects but also the null hypothesis. We follow Lee and Wagenmakers (2013), where $1 \quad BF_{10} \quad 3$ is anecdotal, $3 \quad BF_{10} \quad 10$ is moderate, and $10 \quad BF_{10}$ is considered strong evidence in favor of the alternative hypothesis (i.e., the prior distribution). The reciprocal values, in turn, provide evidence in favor of the null hypothesis. Further, we summarize the posterior distributions with equal tailed 90% credible intervals (CrI), but note this is mostly an arbitrary choice that can be justified in two ways. The tail regions make up 10% of the posterior samples which leads to greater stability. Second, when an interval excludes zero, there is a directional posterior probability of at least 95%. When making posterior based inference, we also emphasize density regions of interest.

All computations were done in R version 3.4.2 (R Core Team, 2017) The models were fitted with the the package **brms** (Bürkner, 2017b), which serves as a front-end to the probabilistic programming language Stan (Stan Development Team, 2016). There are several advantages of the package **brms**. The model specification follows **lme4** (Bates, Mächler, Bolker, & Walker, 2015), although **brms** allows for fitting a much wider range of models. Additionally, there are several post-processing features for model checking (Appendix C1) and Bayesian hypothesis testing. We refer interested readers to Stegmann, Jacobucci, Harring, and Grimm (2018), where **brms** was compared to alternative R packages for fitting nonlinear mixed-effects models. There are also several tutorials using **brms**, including ordinal models (Bürkner & Vuorre, 2017), distributional regression (Estimating Distributional Models with brms), and supporting code for an introductory Bayesian textbook (Statistical Rethinking with brms). Model comparisons were made with the package **loo** (Vehtari, Gabry, Yao, & Gelman, 2018). For this work, we have provided annotated R code for fitting NL-MELSM 1 (Appendix R-code). The final model converged after 3.7 h on a Windows 10 operated system with an IntelCore i7 processor at 3.6 GHz, four cores (8 threads) and 16 GiB RAM.

Model Selection

As noted before, we take the maximal perspective of Barr et al. (2013), where we are primarily interested in fitting a fully parameterized model. This is in contrast to more common stepwise model building approaches, but offers several advantages. First, stepwise strategies can introduce bias in that the stochastic nature of model selection is typically ignored altogether (Harrell, 2001; Kabaila & Leeb, 2006; Leeb, Pötscher, & Ewald, 2015; Raftery, Madigan, & Hoeting, 1997; Tibshirani, Taylor, Lockhart, & Tibshirani, 2016). Second, assuming the model can be estimated, we do not think restricting parameters to zero is a reasonable approach, because estimating the posterior distributions provides useful information (Figure 1). That is, leaving all variables in the model, especially in a Bayesian context, ensures all uncertainties are averaged across when obtaining the marginal posterior distributions. To demonstrate the utility of the proposed NL-MELSM for learning, however, we do make model comparisons with LOO. Of note, the goal of cross-validation is not necessarily to identity the *true* model, which assumes the $\mathscr{M}$-closed setting (Bernardo & Smith, 2001; Lee & Vanpaemel, 2017; Piironen & Vehtari, 2017). That is, it is assumed that the *true* model *is* in the candidate set. On the other hand, in the $\mathscr{M}$-open setting, where it is not assumed that the *true* model necessarily is in the candidate set, the objective is to minimize prediction error. This means that LOO does not inherently favor parsimonious models, but can nonetheless be used to inform model building, revision, and ultimately selection. Additionally, for each of the candidate models, we computed psuedo-BMA (Bayesian model averaging) weights with the approach described in Yao et al. (2017). This provides an additional quantity for model comparison, and in the event no model is preferable, would allow for averaging the posterior and predictive distributions (Section Substantive Applications).

The LOO results are reported in Table 1. While there are no set rules for interpreting LOO, the differences can be interpreted relative to the uncertainty, expressed in its standard errors. We first compared NL-MELSM 1 to a standard linear mixed-effects location scale model (MELSM). This provided evidence in favor of the proposed nonlinear model, thereby suggesting verbal learning is better predicted by an asymptotic function. We then compared predictive accuracy to a standard nonlinear mixed-effects model (NLME), where the fully parameterized model was favored, thus demonstrating the assumption of constant within-person variance is too restrictive. We then expanded NLME, by fitting an asymptotic function to the scale, but not allowing for individual variation therein (NL-MELSM 2). The fully parametrized model minimized the errors, which indicates that allowing for individual variability in the variance leads to improved prediction accuracy. The next comparison was to a LME model fitted to the within-person variance (NL-MELSM 3), which indicated that the nonlinear function provided more accurate predictions. The final comparison was to a model in which correlations between the random location and scale effects were restricted to zero (NL-MELSM 4), where there was model selection uncertainty, as indicated by the standard error. This highlights an advantage of fully Bayesian model comparison, in that there is an automatic measure of uncertainty that guards against overconfident inferences. It would be further possible to fit models to check whether particular correlations would provide a better fit. On the other hand, due to providing more information for inference we

summarize the fully parameterized model, although in practice model averaging could be used to account for uncertainty (Section: Substantive Applications).

Model Interpretation

In this section, we describe the results for NL-MELSM 1. Since a defining feature of the proposed model is quantifying individual variability, we have also plotted the results in Figure 1. This shows the individual effects, summarized with 90% CrIs, where blue bars denote estimates that excluded the population average. Thus, this plot visualizes whether (and how many) individuals differed from the population average. Additionally, Figure 2 shows the predicted values, both for the location of verbal memory and the within-person variability, or scale. In practice, where nonlinear parameters do not always have intuitive interpretations (specially for the scale parameters), these visualizations add clarity to the fitted model.

Location Parameters.—The location parameters of the NL-MELSM capture the average means (fixed effects) and individual departures (random effects) and typically correspond to the same parameters obtained from the standard NLME model. On average, the asymptote (ϕ_0) was 25.95 (90% CrI = [24.97, 26.84]), with anecdotal evidence for the absence of individual variability ($BF_{10} = 0.48$, $\tau_{\phi_0} = 0.93$, 90% CrI = [0.20, 1.79]). Indeed, from the top left panel in Figure 1, it is clear that most individuals had a similar asymptote. The initial recall performance after the first trial is captured by the ϕ_1 parameter ($\phi_1 = 8.58$, 90% CrI = 8.28, 8.90). As suggested by Figure 1, top middle panel, there was strong evidence for individual variability ($BF_{10} = 9.5 \times 10^{25}$, $\tau_{\phi_1} = 2.94$, 90% CrI = [2.64, 3.24]). The learning rate (ϕ_3) that links the initial performance to the asymptote and defines the curvature was, on average, −1.12 (90% CrI = [−1.23, −1.04]), with strong evidence for individual variability ($BF_{10} = 4.3 \times 10^8$, $\tau_{\phi_1} = 0.23$, 90% CrI = [0.10, 0.39]). This can be seen in Figure 1, top left panel, where it is clear that many individuals differed from the population average. That is, some individuals had very "bent" learning trajectories while others showed minimal curvature (Figure 2). These differences are also visible in the predicted trajectories in the upper panel of Figure 2.

Scale Parameters.—The scale parameters capture the within-person variance and just as the location parameters they provide fixed and random effects. Given the model definition, we report all scale parameters on the logarithmic scale (e.g., exp(0.70) = 2.01). The asymptote was, on average, 0.70 (90% crI = [0.25, 0.53]), which provides the theoretically upper bound of within-person variance when the number of trials approach infinity. Interestingly, there was strong evidence for individual variation ($BF_{10} = 2.1 \times 10^3$, $\tau_{\gamma_0} = 0.63$, 90% CrI = [0.23, 1.42]) indicating that some individuals became more variable as trials progressed, while others became less so (see lower panel of Figure 2). The average variability at trial one was 0.43 (90% CrI = [0.25, 0.56]), with anecdotal evidence for individual variability ($BF_{10} = 1.65$, $\tau_{\gamma_1} = 0.23$, 90% CrI = [0.10, 0.39]). However, as shown in the lower panels in Figure 1, it should be noted that several individuals differed from the

average value. The degree of curvature is again defined by the change rate (γ_3) which was, on average, -1.31 (90% CrI = $[-2.94, 0.50]$), with moderate evidence that individuals varied ($BF_{10} = 3.90$, 90% CrI = $[0.25, 2.36]$). As is evident from Figure 2, changes in the within-person SD ranged from nonlinear to practically linear. However, the asymptotic function provided better fit according to LOO (NL-MELSM 1 vs. NL-MELSM 3). Together, by extending the standard NLME model, we demonstrated that within-person SD is best described by a nonlinear trajectory, in addition to showing individual variation therein.

Correlations.—An additional advantage of the proposed model is that correlations between location and scale random effects can be estimated. For example, it is possible to assess whether the asymptote for within-person variance $\left(\tau_{\gamma_0}\right)$ is related to verbal memory at trial 1 $\left(\tau_{\phi_1}\right)$. The full correlation matrix is provided in Table 2 and visually separated in quadrants that indicate the source of correlations. That is, the top left quadrant reports correlations among the location parameters, the lower right quadrant describes correlations among the scale parameters, and the lower left quadrant reports all nine correlations among the location and scale random effects. For the location random effects, there was strong evidence for a positive correlation between the initial performance and learning rate ($r = 0.71$, $BF_{10} = 10 \times 10^7$). In other words, those who began the study with greater verbal ability apparently memorized words at a faster rate and attained their asymptote more quickly compared to those who started with lower recall performance. This is visible in the fan spread of the individual trajectories in Figure 2. At the beginning of the study, there was also strong evidence for a positive correlation between verbal memory and within-person SD ($r = 0.63$, $BF_{10} = 16.16$). Interestingly, there was suggestive evidence that the location and scale rate parameters were positively correlated ($r = 0.46$, $BF_{10} = 3.14$), which suggests that individuals with a "bent" trajectory for verbal memory also showed increased curvature with respect to within-person variability. We note that there was anecdotal evidence for a negative relationship between variability at trial one and the asymptote for within-person SD ($r = -0.37$, $BF_{10} = 1.91$). While the effect was not large, this correlation denotes that more variable individuals, at the beginning of the study, also had a smaller estimate for the within-person SD asymptote.

Predicting Nonlinear Parameters

In this section, we extend NL-MELSM 1 by predicting the nonlinear parameters. In practical applications, this allows for answering specific questions relating to the nonlinear function. For these illustrative data, we predicted each nonlinear parameter with a fixed effect that included three age groups (Section Illustrative Data). We assumed a dummy coded matrix, where the youngest group was the reference category (location: β_0 and scale: η_0). With this parametrization, we estimated differences for the middle (β_1 and η_1) and the oldest age group (β_2 and η_2). In other words, in reference to Equation 12, the design matrices for the location $\mathbf{X}$ and scale $\mathbf{W}$ included age group as the predictor variable. For simplicity, we assumed *so-called* non-informative priors, and summarized the posterior distributions with directional probabilities. For example, $p(\theta > 0 | y, M)$ denotes the posterior probability of a positive effect.

Model Interpretation

To understand the fitted model, we again rely on visualization, where the predicted values for each group are plotted in Figure 3. Additionally, the fixed effects are reported in Table A1. Before describing these results it should be noted that LOO did not favor this model compared to NL-MELSM 1 ($LOO_{diff} = 0.3$, SE = 8.3), and according to the pseudo-BMA+ weights, each model received about equal support.

Location Parameters.—As evident from Figure 3, the age groups differed at trial 1, with the youngest group showing the highest verbal memory. Indeed, in comparison to the youngest age group, the posterior probability that both age groups had lower scores was 99.9% (Table A1). A similar pattern was observed for the learning rate, where it appears that the youngest group showed the most curvature in their learning trajectory. Indeed, compared to the youngest study participants, there was a 99.9% posterior probability that the older age groups learned words at a slower rate.

For within-person variability, the posterior probability that the middle age group had a higher asymptote than the youngest age group was 97.5%, thereby suggesting learning variability may be related to age. Although the 90% CrI included zero, it should be noted there was a 91.3% posterior probability that the oldest age group also had a larger asymptote. Further, as illustrated in the lower panels of Figure 3, it is clear that only the youngest group had a negative within-person SD trajectory, in that variance reduced while in the process of learning. On the other hand, for the older groups, the learning curves were slightly concave indicating that the asymptote was higher than the variability at the beginning of the learning trials. In comparison to the youngest group, the middle age group apparently had less within-person SD at trial 1, with a posterior probability of 98.8%. We emphasize that the findings are far from confirmatory, but nonetheless demonstrate the utility of the proposed model. That is, by modeling within-person SD we were able to suggest a relationship between learning and variability, in addition to group differences therein.

Discussion

In this paper we expanded upon the standard nonlinear mixed-effects model, by fitting a nonlinear function to the within-person variance, and allowing for individual variation therein. This model was specifically built to address the goal of identifying and accounting for IIV in a nonlinear mixed effects model, with a specific application for learning. Nonlinear modeling and explaining learning at both the mean and variance level requires repeated trials and flexible methods that are able to capture changes within and differences between individuals, of which the NL-MELSM is one such model. This approach can simultaneously model both the mean nonlinear structure, in the location component, and also the variance nonlinear structure, in the scale part. Moreover, variables can be included to predict either (or both) the location and scale of nonlinear parameters. The model also results in the estimation of random effects for both the location and the scale components, thus allowing for characterizing individual effects. This does not only control for the

influence of mean and variance dependency, but also allows for obtaining correlations between the nonlinear location and scale random effects.

The information provided by the correlations across the nonlinear location and scale effects is a unique feature of the NL-MELSM. In our illustrative example, at trial 1, we found that individuals with higher verbal memory also had the most variability in verbal memory. Further, while the evidence was not strong, there was an interesting negative relationship between variability at trial 1 and the asymptote for within-person SD. Participants with larger initial variability tended to decrease over time while participants with smaller initial variability tend to be associated to increasing variability over time. While this could be an actual feature of how people learn, it might also be due to a ceiling effect in the observed data. Participants who started higher initially, reached the full list length faster and with it, had less room to be variable. Based on visual inspection, extending the model to predict the nonlinear parameters suggested that only the youngest age group showed a declining curve, in that variability reduced when learning (Figure 3) – which could again be due to the generally higher performance and upper limit of 27 words. Of note, these types of inferences, with a fine-grained level of detail about IIV, are unique to the NL-MELSM.

Substantive Applications

In addition to learning, there are many psychological applications where the NL-MELSM may be fruitfully applied. For example, there is now a large body of research using growth curve models in the structural equation framework for estimation (Grimm et al., 2011). Paradoxically, as the name explicitly assumes a curved trajectory, extending beyond a linear function has proven non-trivial, often requiring approximations such as the Taylor series expansion (Hall & Bailey, 2001; Rast, 2011). And we are not aware of any examples that have fit a sub-model to the within-person variability. On the other hand, growth curve models can be estimated in a mixed-effects framework (Curran, 2003; Curran, Obeidat, & Losardo, 2010). Thus, the presented model provides a flexible alternative that also allows for asking novel research questions. As such, the NL-MELSM can be applied in the broader context of developmental research, to elucidate both inter and intra-individual variability in curved trajectories.

Bayesian estimation has become more accessible and popular over the last years as it entered mainstream software packages. In fact, the models described here were fit with the R package **brms** which uses similar syntax as **lme4** (Bürkner, 2017a). In our experience, **brms** is sufficiently flexible to fit most models in psychology, but Stan can be used directly if needed. For example, it is possible for another sub-model predicting the between-person variance (Rast & Ferrer, 2018). While Bayesian estimation techniques have become more widespread, the same can not be said about Bayesian inference. While a thorough discussion on that topic (e.g., on hypothesis testing) was beyond the scope of this work we illustrated some of the possibilities that Bayesian inference is able to offer. There are now several introductions for Bayesian inference specifically for psychological applications (Quintana & Williams, 2018). These are typically geared towards simpler models (e.g., *t*-test; Rouder, Speckman, Sun, Morey, & Iverson, 2009), but the techniques can be used with the NL-MELSM. We refer to Wagenmakers, Marsman, et al. (2018) and Wagenmakers, Love, et al.

(2018), in addition to Wagenmakers, Lodewyckx, Kuriyal, and Grasman (2010) which is specifically about the Savage-Dickey ratio. Importantly, Bayes factors depend critically upon the prior distribution, which should ideally be informed by relevant theory. In the absence of precise theoretical predictions, it is common place to assume defaults (Rouder & Morey, 2012). In practice, sensitivity analyses should be performed, although they were excluded from this work for simplicity. We refer to Carlsson, Schimmack, Williams, and Bürkner (2017), where the prior distributions were varied to investigate whether the effect was robust.

For NLME models, it is customary to compare the fit of several functions (Rast, 2011). As such, we included some information about model selection with LOO (Vehtari et al., 2017). We reemphasize that LOO can guide model evaluation, and model selection (if desired), but it is important to appreciate potential limitations (Section Model Selection; Gronau & Wagenmakers, 2018). For example, while LOO is a fully Bayesian approach to assess predictive accuracy, it is asymptotically equivalent to the widely applicable information criterion (Watanabe, 2010) which is, in turn, asymptotically equivalent to the Akaike information criterion (Akaike, 1974). This means that LOO is not consistent for the purpose of model selection (Y. Zhang & Yang, 2015), having a similar justification as AIC (Burnham & Anderson, 2004), but with respect to the posterior distribution (Vehtari et al., 2017). Further, while beyond the scope of this work, uncertainty in model selection should be considered with averaging the posterior or predictive distributions (Yao et al., 2017, e.g., with pseudo-BMA+ weights). We refer to Hoeting, Madigan, Raftery, and Volinsky (1999) for an introduction, and two examples of Bayesian model averaging in the psychological literature (Gronau et al., 2017; Scheibehenne, Gronau, Jamil, & Wagenmakers, 2017).

Limitations

There are several notable limitations of this work in particular, as well as the NL-MELSM in general. For the latter, when the variances are of interest, it should be noted that their magnitude is also defined by the location of the average response. In other words, with bounded variables that are common in psychology, the variance will be a function of the person's mean (Baird, Le, & Lucas, 2006; Eid & Diener, 1999; Kalmijn & Veenhoven, 2005; Rouder, Tuerlinckx, Speckman, Lu, & Gomez, 2008). This problem is known in MELSM applications (Rast & Ferrer, 2018), but also applies to the NL-MELSM. In the present work, this correlation was large ($r = 0.63$ among τ_{ϕ_1} and τ_{γ_1}) which indicates that

individuals further away from the lower boundary also had more variability at trial 1. This could be an effect of substantive interest, or dictated by aspects of the study design. Importantly, this is not necessarily a problem for the NL-MELSM, but it should be considered when making inference from the random effect correlations. On the other hand, it is often argued that variability predicts important life outcomes (even death) while ignoring the location altogether (MacDonald, Hultsch, & Dixon, 2008), whereas the NL-MELSM accounts for both sources of uncertainty

Moreover, while we provided a *fully* Bayesian approach (i.e., posterior uncertainty was accounted for in model selection), we did not make use of the marginal likelihood and posterior model probabilities. For these models, we were not able to successfully specify what could reasonably be considered well-defined hypotheses of the data generating process.

We thus used pseudo-BMA+ weights (Yao et al., 2017), in addition to the Savage-Dickey ratio which is specifically for nested hypotheses that we found less demanding to specify (and justify). Second, because our goal was to introduce a Bayesian NL-MELSM, many choices were made for simplicity. For example, we did not provide an in-depth example of model checking, but note this is important in practical applications. Another potential limitation, considering we also offered some guidance for applied researchers, is the advocacy for fitting fully-parametrized models. This does stand in contrast to using "significance" to include variables in a model, but removing based on "non-significance" assumes evidence for the null hypothesis. We reiterate our preference in accounting for all sources of uncertainty rather than removing them from the model altogether.

Regarding the underlying data, we would like to point out that five repeated measurements is certainly at the lower end of data requirements, both in terms of accurately modeling the mean and/or the variance structure. While this current work served as an illustration on the use of a nonlinear MELSM in psychological data, for the purpose of an actual substantive application we would recommend at least 20 repeated measures (Rast & Zimprich, 2010). In our experience, recovering the fixed effects parameters for the location and scale is not typically an issue. However, fewer data increases uncertainty in the estimation of the elements in the covariance matrix. For example, Rast and Ferrer (2018) suggest that, in a linear MELSM, large correlations of approximately .40 require at least 75 participants with 75 repeated measurements.

Future Directions and Conclusion

The introduced model can be extended in several ways. We only fitted one type of nonlinear function, but of course, the model readily allows for the implementation of other functions, both for the location and the scale. While there is a substantial body of literature on nonlinear functions for the mean structure (e.g. N. J. Evans et al., 2018; Heathcote et al., 2000; Mazur & Hastie, 1978; Newell & Rosenbloom, 1981; Restle & Greeno, 1970; Ritter & Schooler, 2001; Zimprich et al., 2008), research in functions that best describe within-person variance in the context of learning are not available yet. The NL-MELSM might prove fruitful in expanding the current body of literature on modeling learning and other nonlinear change processes.

The purpose of this paper was to present the NL-MELSM as a flexible model for psychological applications. Our proposed model is suited for learning, as presented in this work, but can also be used more generally in gerontological and/or developmental areas. By focusing on the within-person variance, this approach opened up possibilities for modeling a component that is often disregarded as unexplained residuals (i.e., "noise"). The provided example illustrated such possibilities and highlighted that the residual variance may show systematic, nonlinear patterns, that are important for understanding psychological processes.

Acknowledgments

Appendix A: Predicting Nonlinear Parameters

Table A1

Parameter estimates.

	Location				**Scale**		
Parameter	**M**	**SD**	**90% CrI**	*Parameter*	**M**	**SD**	**90% CrI**
$\beta_0\phi_0$	26.15	0.46	[25.40, 26.93]	$\eta_0\gamma_0$	0.34	0.28	[−0.16, 0.74]
$\beta_1\phi_0$	−0.38	0.74	[−1.56, 0.90]	$\eta_1\gamma_0$	**0.75**	**0.53**	**[0.05, 1.73]**
$\beta_2\phi_0$	−0.73	0.92	[−2.29, 0.71]	$\eta_2\gamma_0$	0.51	0.48	[−0.13, 1.34]
$\beta_0\phi_1$	10.37	0.30	[9.89, 10.87]	$\eta_0\gamma_1$	0.61	0.11	[0.43, 0.77]
$\beta_1\phi_1$	**−2.49**	**0.40**	**[−3.16, −1.84]**	$\eta_1\gamma_1$	**−0.31**	**0.16**	**[−0.59, −0.07]**
$\beta_2\phi_1$	**−2.91**	**0.41**	**[−3.58, −2.24]**	$\eta_2\gamma_1$	−0.15	0.15	[−0.41, 0.06]
$\beta_0\phi_2$	−0.76	0.07	[−0.88, −0.64]	$\eta_0\gamma_2$	−1.56	0.98	[−2.79, 0.49]
$\beta_1\phi_2$	**−0.42**	**0.10**	**[−0.58, −0.25]**	$\eta_1\gamma_2$	−0.48	0.66	[−1.51, 0.64]
$\beta_2\phi_2$	**−0.64**	**0.11**	**[−0.82, −0.46]**	$\eta_2\gamma_2$	−0.77	0.74	[−1.87, 0.57]

Bold indicates 90% intervals that excluded zero. β_0 and η_0: intercept (youngest age group). β_1 and η_1: middle age group β_2 and η_2: oldest age group. ϕ_0 and γ_0: asymptote. ϕ_0 and γ_0: "intercept" (estimate at trial 1). ϕ_2 and γ_2: log rate of change.

Appendix B: R-code

Example code for the final model. Data and code can be obtained from the OSF repository at https://osf.io/k3rsz/

```
library(brms)

dat ← read.csv("dat.csv")

###############################

######## define the model ########

###############################

b_mod1 ← bf(recall ~ betaMu + (alphaMu - betaMu) * exp(-exp(gammaMu) *
trial),

        betaMu ~ 1 + (1|c|subject),

        alphaMu ~ 1 + (1|c|subject),

        gammaMu ~ 1 + (1|c|subject),

        nl = TRUE)
```

```
        nlf(sigma ~ betaSc + (alphaSc - betaSc) * exp(-exp(gammaSc) * trial),

            alphaSc ~ 1 + (1|c|subject),

            betaSc ~ 1 + (1|c|subject),

            gammaSc ~ 1 + (1|c|subject))

###############################

########## define priors ##########

###############################

prior1 = c(set_prior("normal(25, 2)", nlpar = "betaMu"),

            set_prior("normal(8, 2)", nlpar = "alphaMu"),

            set_prior("normal(-1, 2)", nlpar = "gammaMu"),

            set_prior("normal(0, 1)", nlpar = "betaSc"),

            set_prior("normal(1, 1)", nlpar = "alphaSc"),

            set_prior("normal(0, 1)", nlpar = "gammaSc"),

            set_prior("student_t(3, 0, 5)" , class = "sd", nlpar = "alphaMu"),

            set_prior("student_t(3, 0, 5)" , class = "sd", nlpar = "betaMu"),

            set_prior("student_t(3, 0, 5)" , class = "sd", nlpar = "gammaMu"),

            set_prior("student_t(3, 0, 1)" , class = "sd", nlpar = "alphaSc"),

            set_prior("student_t(3, 0, 1)" , class = "sd", nlpar = "betaSc"),

            set_prior("student_t(3, 0, 1)" , class = "sd", nlpar = "gammaSc"))

## fit the model

b_fit1 = brm(b_mod1, prior = prior1, data = dat,

            control = list(adapt_delta = .9999, max_treedepth = 12),

            chains = 4, iter = 10000,

            inits = 0, cores = 4)
```

```
summary(b_fit1)

# we encourge the exploration of a variety of non-linear functions.
```

Appendix C: Posterior Predictive Check
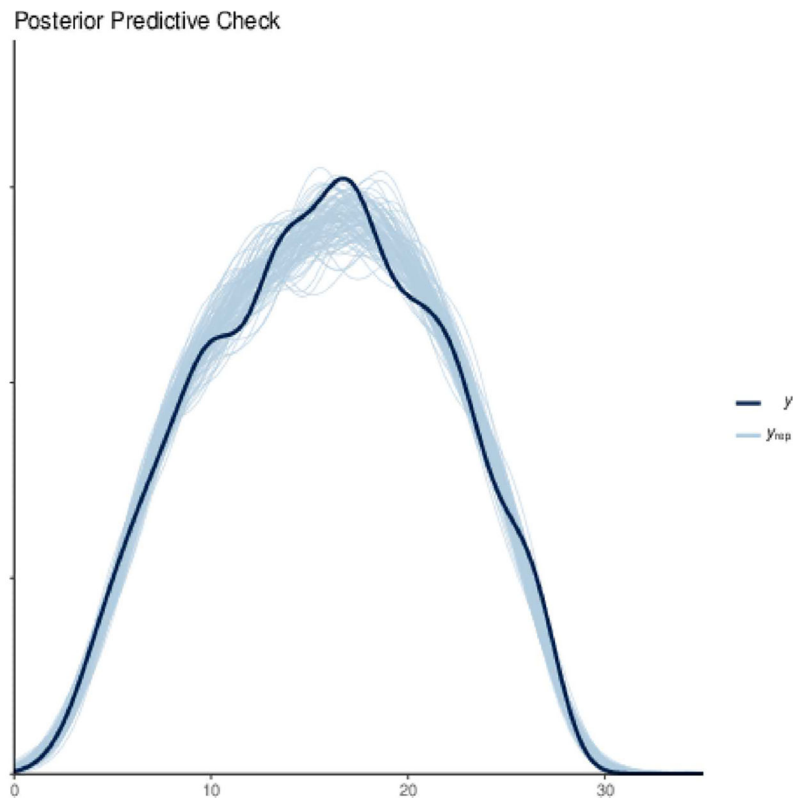
Posterior Predictive Check

Figure C1 .
Posterior predictive check for NL-MELSM 1.

References

Ackerman PL (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. Journal of experimental psychology: General, 117 (3), 288.

Akaike H (1974, 12). A New Look at the Statistical Model Identification. IEEE Transactions on Automatic Control, 19 (6), 716–723. doi: 10.1109/TAC.1974.1100705

Anzanello MJ, & Fogliatto FS (2011). Learning curve models and applications: Literature review and research directions. International Journal of Industrial Ergonomics, 41 (5), 573–583. Retrieved from 10.1016/j. ergon.2011.05.001 doi: 10.1016/j.ergon.2011.05.001

Baird BM, Le K, & Lucas RE (2006). On the nature of intraindividual personality variability: Reliability, validity, and associations with well-being. Journal of personality and social psychology, 90 (3), 512. [PubMed: 16594835]

Baltes PB, & Kliegl R (1992). Further testing of limits of cognitive plasticity: Negative age differences in a mnemonic skill are robust. Developmental Psychology, 28 (1), 121.

Barnard J, McCulloch R, & Meng X-L (2000). Modelling Covariance Matrices in Terms of StandardDeviations and Correlations With Applications to Shrinkage. Statistica Sinica, 10 (4), 1281–1311. doi: 10.2307/24306780

Barr DJ, Levy R, Scheepers C, & Tily HJ (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language, 68 (3), 255–278. Retrieved from http://www.sciencedirect.com/science/article/pii/S0749596X12001180 doi: 10.1016/j.jml. 2012.11.001

Bates D, Mächler M, Bolker B, & Walker S (2015). Fitting Linear Mixed-Effects Models Using {lme4}. Journal of Statistical Software, 67 (1), 1–48. doi: 10.18637/jss.v067.i01

Bernardo JM, & Smith AFM (2001). Bayesian theory. IOP Publishing.

Blachstein H, & Vakil E (2016). Verbal learning across the lifespan: an analysis of the components of the learning curve. Aging, Neuropsychology, and Cognition, 23 (2), 133–153.

Bonate PL (2009). Pharmacokinetic and Pharmacodynamic Modeling and Simulation. doi: 10.1007/978-1-4419-9485-1

Brose A, Voelkle MC, Lövdén M, Lindenberger U, & Schmiedek F (2015). Differences in the between-person and within-person structures of affect are a matter of degree. European Journal of Personality, 29 (1), 55–71.

Bürkner P-C (2017a). Advanced Bayesian Multilevel Modeling with the R Package brms. , 1–18. Retrieved from http://arxiv.org/abs/1705.11123

Bürkner P-C (2017b). brms : An R Package for Bayesian Multilevel Models Using Stan. Journal of Statistical Software, 80 (1). Retrieved from http://www.jstatsoft.org/v80/i01/ doi: 10.18637/jss.v080.i01

Bürkner P-C, & Vuorre M (2017). Ordinal Models in Psychology : A Tutorial. , 1–17.

Burnham KP, & Anderson DR (2004). Multimodel inference: Understanding AIC and BIC in model selection. Sociological Methods and Research, 33 (2), 261–304. doi:10.1177/0049124104268644

Bush RR, & Mosteller F (1955). Stochastic models for learning. New York: Wiley.

Carlsson R, Schimmack U, Williams DR, & Bürkner PC (2017). Bayes Factors From Pooled Data Are No Substitute for Bayesian Meta-Analysis: Commentary on Scheibehenne, Jamil, and Wagenmakers (2016) (Vol. 28) (No. 11). doi:10.1177/0956797616684682

Cattell RB, Cattell AKS, & Rhymer RM (1947). P-technique demonstrated in determining psychophysiological source traits in a normal individual. Psychometrika, 12 (4), 267–288. [PubMed: 18921433]

Chu B-C, Millis S, Arango-Lasprilla JC, Hanks R, Novack T, & Hart T (2007). Measuring recovery in new learning and memory following traumatic brain injury: a mixed-effects modeling approach. Journal of clinical and experimental neuropsychology, 29 (6), 617–625. [PubMed: 17691034]

Cole TJ, Donaldson MDC, & Ben-Shlomo Y (2010). SITAR—a useful instrument for growth curve analysis. International Journal of Epidemiology, 39 (6), 1558–1566. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20647267http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2992626https://academic.oup.com/ije/article-lookup/doi/10.1093/ije/dyq115. doi: 10.1093/ije/dyq115 [PubMed: 20647267]

Comets E, Lavenu A, & Lavielle M (2017). Parameter Estimation in Nonlinear Mixed Effect Models Using saemix, an R Implementation of the SAEM Algorithm. Journal of Statistical Software, 80 (3). Retrieved from http://www.jstatsoft.org/v80/i03/. doi: 10.18637/jss.v080.i03

Craik FIM, & Lockhart RS (1972). Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior, 11 (6), 671–684. doi:10.1016/S0022-5371(72)80001-X

Curran P (2003). Have multilevel models been structural equation models all along? Multivariate Behavioral Research, 38 (4), 529–569. Retrieved from http://www.leaonline.com/doi/abs/10.1207/s15327906mbr3804_5. doi: 10.1207/s15327906mbr3804 [PubMed: 26777445]

Curran P, Obeidat K, & Losardo D (2010). Twelve frequently asked questions about growth curve modeling. Journal of Cognition and Development, 11 (2), 121–136. doi: 10.1080/15248371003699969 [PubMed: 21743795]

Dalal DK, & Zickar MJ (2012). Some common myths about centering predictor variables in moderated multiple regression and polynomial regression. Organizational Research Methods, 15 (3), 339–362.

Eid M, & Diener E (1999). Intraindividual variability in affect: Reliability, validity, and personality correlates. Journal of Personality and Social Psychology, 76 (4), 662.

Eizenman DR, Nesselroade JR, Featherman DL, & Rowe JW (1997). Intraindividual variability in perceived control in a older sample: The MacArthur successful aging studies. Psychology and aging, 12 (3), 489. [PubMed: 9308096]

Estes W (1950). Toward a statistical theory of learning. Psychological Review, 57 (2), 94–107.

Evans M, & Moshonov H (2006). Checking for prior-data conflict. Bayesian Analysis, 1 (4), 893–914. doi: 10.1214/06-BA129

Evans NJ, Brown SD, Mewhort DJK, & Heathcote A (2018). Refining the law of practice. Psychological Review, 125 (4), 592–605. Retrieved from http://doi.apa.org/getdoi.cfm?doi=10.1037/rev000010. doi: 10.1037/rev0000105 [PubMed: 29952624]

Fagot D, Mella N, Borella E, Ghisletta P, Lecerf T, & De Ribaupierre A (2018). Intra-Individual Variability from a Lifespan Perspective: A Comparison of Latency and Accuracy Measures. Journal of Intelligence, 6 (1), 16.

Feldman LS, Cao J, Andalib A, Fraser S, & Fried GM (2009). A method to characterize the learning curve for performance of a fundamental laparoscopic simulator task: Defining "learning plateau" and "learning rate". Surgery, 146 (2), 381–386. Retrieved from 10.1016/j.surg.2009.02.021 doi: 10.1016/j.surg.2009.02.021 [PubMed: 19628099]

Fiske DW, & Rice L (1955). Intra-individual response variability. Psychological Bulletin, 52 (3), 217. [PubMed: 14371891]

Freund AM (2008). Successful Aging as Management of Resources: The Role of Selection, Optimization, and Compensation. Research in Human Development, 5 (2), 94–106. doi: 10.1080/15427600802034827

Gabry J, Simpson D, Vehtari A, Betancourt M, & Gelman A (2019). Visualization in Bayesian workflow. J.R. Statist. Soc. A(1), 1–14. Retrieved from http://arxiv.org/abs/1709.01449.

Garrett DD, MacDonald SWS, & Craik FIM (2012). Intraindividual reaction time variability is malleable: feedback-and education-related reductions in variability with age. Frontiers in human neuroscience, 6, 101. [PubMed: 22557954]

Gelman A (2006). Prior distributions for variance parameters in hierarchical models. Bayesian Analysis, 1 (3), 515–534. doi: 10.1214/06-BA117A

Gelman A, & Loken E (2014). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. Psychological bulletin, 140 (5), 1272–1280. doi: 10.1037/a0037714 [PubMed: 25180805]

Gelman A, Simpson D, & Betancourt M (2017). The prior can generally only be understood in the context of the likelihood. ArXiv preprint. Retrieved from http://arxiv.org/abs/1708.07487

Goldstein H, Leckie G, Charlton C, Tilling K, & Browne WJ (2018). Multilevel growth curve models that incorporate a random coefficient model for the level 1 variance function. Statistical Methods in Medical Research, 27 (11), 3478–3491. Retrieved from http://journals.sagepub.com/doi/10.1177/0962280217706728 doi: 10.1177/0962280217706728 [PubMed: 28459180]

Grimm KJ, Ram N, & Hamagami F (2011). Nonlinear growth curves in developmental research. Child Development, 82 (5), 1357–1371. doi: 10.1111/j.1467-8624.2011.01630.x [PubMed: 21824131]

Gronau QF, Van Erp S, Heck DW, Cesario J, Jonas KJ, & Wagenmakers E-J (2017). A Bayesian model-averaged meta-analysis of the power pose effect with informed and default priors: the case of felt power. Comprehensive Results in Social Psychology, 2 (1), 123–138. Retrieved from https://www.tandfonline.com/doi/full/10.1080/23743603.2017.1326760 doi: 10.1080/23743603.2017.1326760

Gronau QF, & Wagenmakers E-J (2018). Limitations of Bayesian Leave-One-Out Cross-Validation for Model Selection. Retrieved from https://psyarxiv.com/at7cx/ doi: 10.17605/OSF.IO/AT7CX

Hager W, & Hasselhorn M (1994). Handbuch Deutschsprachiger Wort- normen. [Handbook of German word norms]. Göttingen, Germany: Hogrefe & Huber.

Hall D, & Bailey R (2001). Modeling and prediction of forest growth variables based on multilevel nonlinear mixed models. Forest Science, 47 (3), 311–321.

Harada CN, Love MCN, & Triebel K (2013). Normal Cognitive Aging. Clin Geriatr Med, 29 (4), 737–752. doi: 10.1016/j.cger.2013.07.002.Normal [PubMed: 24094294]

Harrell F (2001). Regression modeling strategies, with applications to linear models, survival analysis and logistic regression. New York: Springer.

Heathcote A, Brown S, & Mewhort DJ (2000). The power law repealed: The case for an exponential law of practice. Psychonomic Bulletin and Review, 7 (2), 185–207. doi: 10.3758/BF03212979 [PubMed: 10909131]

Hedeker D, Demirtas H, & Mermelstein RJ (2009). A mixed ordinal location scale model for analysis of Ecological Momentary Assessment (EMA) data. Statistics and its interface, 2 (4), 391–401. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20357914 [PubMed: 20357914]

Hedeker D, Mermelstein RJ, & Demirtas H (2008). An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. Biometrics, 64 (2), 627–634. doi: 10.1111/j.1541-0420.2007.00924.x [PubMed: 17970819]

Hedeker D, Mermelstein RJ, & Demirtas H (2012). Modeling between-subject and within-subject variances in ecological momentary assessment data using mixed-effects location scale models. Statistics in Medicine, 31 (27), 3328–3336. doi: 10.1002/sim.5338 [PubMed: 22419604]

Hoeting JA, Madigan D, Raftery AE, & Volinsky CT (1999). Bayesian Model Averaging: A Tutorial. Statistical Science, 14 (4), 382–417. doi: 10.2307/2676803

Horn JL (1972). State, trait and change dimensions of intelligence. British Journal of Educational Psychology, 42 (2), 159–185.

Hultsch DF, Hertzog C, Small BJ, McDonald-Miszczak L, & Dixon RA (1992). Short-term longitudinal change in cognitive performance in later life. Psychology and Aging, 7 (4), 571. [PubMed: 1466826]

Jones RN, Rosenberg AL, Morris JN, Allaire JC, McCoy KJM, Marsiske M, … Malloy PF (2005). A growth curve model of learning acquisition among cognitively normal older adults. Experimental aging research, 31 (3), 291–312. [PubMed: 16036723]

Kabaila P, & Leeb H (2006). On the large-sample minimal coverage probability of confidence intervals after model selection. Journal of the American Statistical Association, 101 (474), 619–629. doi: 10.1198/016214505000001140

Kalmijn W, & Veenhoven R (2005). Measuring inequality of happiness in nations: In search for proper statistics. Journal of Happiness Studies, 6 (4), 357–396.

Kausler D (1994). Learning and memory in normal aging. San Diego: Academic Press.

Leckie G, French R, Charlton C, & Browne W (2014). Modeling heterogeneous variance–covariance components in two-level models. Journal of Educational and Behavioral Statistics, 39 (5), 307–332.

Lee MD, & Vanpaemel W (2017). Determining informative priors for cognitive models. Psychonomic Bulletin & Review, 1–14. doi: 10.3758/s13423-017-1238-3 [PubMed: 27368622]

Lee MD, & Wagenmakers E-J (2013). Bayesian cognitive modeling: A practical course. Cambridge: Cambridge University Press.

Leeb H, Pötscher BM, & Ewald K (2015). On Various Confidence Intervals Post-Model-Selection. Statistical Science, 30 (2), 216–227. doi: 10.1214/14-STS507

Lewandowski D, Kurowicka D, & Joe H (2009). Generating random correlation matrices based on vines and extended onion method. Journal of Multivariate Analysis, 100 (9), 1989–2001. Retrieved from 10.1016/j.jmva.2009.04.008 doi: 10.1016/j.jmva.2009.04.008

Lin X, Mermelstein RJ, & Hedeker D (2018). A 3-level Bayesian mixed effects location scale model with an application to ecological momentary assessment data. Statistics in medicine, 37 (13), 2108–2119. [PubMed: 29484693]

Logan GD (1988). Toward an instance theory of automatization. Psychological review, 95 (4), 492.

Lövdén M, Li S-C, Shing YL, & Lindenberger U (2007). Within-person trial-to-trial variability precedes and predicts cognitive decline in old and very old age: Longitudinal data from the Berlin Aging Study. Neuropsychologia, 45 (12), 2827–2838. [PubMed: 17575988]

MacDonald SWS, Hultsch DF, & Dixon RA (2008). Predicting impending death: inconsistency in speed is a selective and early marker. Psychology and aging, 23 (3), 595. [PubMed: 18808249]

Marsman M, & Wagenmakers EJ (2017). Bayesian benefits with JASP. European Journal of Developmental Psychology, 14 (5), 545–555. doi: 10.1080/17405629.2016.1259614

Mazur JE, & Hastie R (1978). Learning as accumulation: A reexamination of the learning curve. Psychological Bulletin, 85 (6), 1256. [PubMed: 734012]

Nesselroade JR (1991). Interindividual differences in intraindividual change.

Nesselroade JR, & Salthouse TA (2004). Methodological and theoretical implications of intraindividual variability in perceptual-motor performance. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 59 (2), P49–P55.

Newell A, & Rosenbloom PS (1981). Mechanisms of skill acquisition and the law of practice. In Anderson JR (Ed.), (pp. 1–51). Hillsdale, NJ: Erlbaum.

O'Hara RB, & Sillanpää MJ (2009). A review of bayesian variable selection methods: What, how and which. Bayesian Analysis, 4 (1), 85–118. doi: 10.1214/09-BA403

Piironen J, & Vehtari A (2017). Comparison of Bayesian predictive methods for model selection. Statistics and Computing, 27 (3), 711–735. doi: 10.1007/s11222-016-9649-y

Pinheiro J, & Bates D (2006). Mixed-Effects Models in $\$proglang{{}S{}}$ and $\$proglang{{}S-PLUS{}}$. Springer Science {&} Business Media.

Quintana DS, & Williams DR (2018). Bayesian alternatives for common null-hypothesis significance tests in psychiatry: A non-technical guide using JASP. BMC Psychiatry, 18 (1). doi: 10.1186/s12888-018-1761-4

R Core Team. (2017). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing Retrieved from https://www.r-project.org/.

Raftery AE, Madigan D, & Hoeting JA (1997). Bayesian Model Averaging for Linear Regression Models. Journal of the American Statistical Association, 92 (437), 179–191. doi: 10.1080/01621459.1997.10473615

Ram N, & Gerstorf D (2009). Time-structured and net intraindividual variability: Tools for examining the development of dynamic characteristics and processes. Psychology and aging, 24 (4), 778. [PubMed: 20025395]

Ram N, Stollery B, Rabbitt P, & Nesselroade JR (2005). Cognitive performance inconsistency: Intraindividual change and variability. Psychology and Aging, 20 (4), 623–633. doi: 10.1037/0882-7974.20.4.623 [PubMed: 16420137]

Rast P (2011). Verbal Knowledge, Working Memory, and Processing Speed as Predictors of Verbal Learning in Older Adults. Developmental Psychology, 47 (5), 1490–1498. doi: 10.1037/a0023422 [PubMed: 21574701]

Rast P, & Ferrer E (2018). A Mixed-Effects Location Scale Model for Dyadic Interactions. , 1–63. doi: 10.1080/00273171.2018.1477577

Rast P, Hofer SM, & Sparks C (2012). Modeling Individual Differences in Within-Person Variation of Negative and Positive Affect in a Mixed Effects Location Scale Model Using BUGS/JAGS. Multivariate Behavioral Research, 47 (2), 177–200. doi: 10.1080/00273171.2012.658328 [PubMed: 26734847]

Rast P, & Zimprich D (2009). Individual Differences and Reliability of Paired Associates Learning in Younger and Older Adults. Psychology and Aging, 24 (4), 1001–1006. doi: 10.1037/a0016138 [PubMed: 20025414]

Rast P, & Zimprich D (2010). Individual differences in a positional learning task across the adult lifespan. Learning and Individual Differences, 20 (1), 1–7. doi: 10.1016/j.lindif.2009.10.010

Restle F, & Greeno JG (1970). Introduction to mathematical psychology. Reading, MA: Addison-Wesley.

Rickard TC (2004). Strategy Execution in Cognitive Skill Learning: An Item-Level Test of Candidate Models. Journal of Experimental Psychology: Learning, Memory, and Cognition, 30 (1), 65–82. Retrieved from http://doi.apa.org/getdoi.cfm?doi=10. 1037/0278-7393.30.1.65 doi: 10.1037/0278-7393.30.1.65

Ritter F, & Schooler LJ (2001). The learning curve. International Encyclopedia of the Social and Behavioral Sciences, 8602–8605. doi: 10.1038/nbt0803-951

Röcke C, & Brose A (2013). Intraindividual variability and stability of affect and well-being. GeroPsych.

Rouder JN, & Morey RD (2012). Default Bayes Factors for Model Selection in Regression. Multivariate Behavioral Research, 47 (6), 877–903. doi: 10.1080/00273171.2012.734737 [PubMed: 26735007]

Rouder JN, Speckman PL, Sun D, Morey RD, & Iverson G (2009). Bayesian t tests for accepting and rejecting the null hypothesis. Psychonomic Bulletin and Review, 16 (2), 225–237. doi: 10.3758/PBR.16.2.225 [PubMed: 19293088]

Rouder JN, Tuerlinckx F, Speckman P, Lu J, & Gomez P (2008). A hierarchical approach for fitting curves to response time measurements. Psychonomic Bulletin and Review, 15 (6), 1201–1208. doi: 10.3758/PBR.15.6.1201 [PubMed: 19001591]

Salthouse TA (2009). When does age-related cognitive decline begin? Neurobiology of Aging, 30 (4), 507–514. doi: 10.1016/j.neurobiolaging.2008.09.023 [PubMed: 19231028]

Scheibehenne B, Gronau QF, Jamil T, & Wagenmakers EJ (2017). Fixed or Random? A Resolution Through Model Averaging: Reply to Carlsson, Schimmack, Williams, and Bürkner (2017). Psychological Science, 28 (11), 1698–1701. doi: 10.1177/0956797617724426 [PubMed: 28910199]

Sliwinski M, Hoffman L, & Hofer SM (2010). Evaluating convergence of within-person change and between-person age differences in age-heterogeneous longitudinal studies. Research in Human Development, 7 (1), 45–60. [PubMed: 20671986]

Smith G (1998). Learning Statistics by Doing Statistics. Journal of Statistics Education, 6 (3), 1–10. doi: 10.1080/10691898.1998.11910623

Stan Development Team. (2016). Rstan: the R interface to Stan. Retrieved from http://mc-stan.org/

Stegmann G, Jacobucci R, Harring JR, & Grimm KJ (2018). Nonlinear Mixed-Effects Modeling Programs in R. Structural Equation Modeling, 25 (1), 160–165. doi: 10.1080/10705511.2017.1396187

Tibshirani RJ, Taylor J, Lockhart R, & Tibshirani R (2016). Exact Post-Selection Inference for Sequential Regression Procedures. Journal of the American Statistical Association, 111 (514), 600–620. doi: 10.1080/01621459.2015.1108848

Tulving E, & Madigan S (1970). Memory and Verbal Learning. Annual Review of Psychology, 21 (1), 437–484. doi: 10.1146/annurev.ps.21.020170.002253

Tulving E, & Pearlstone Z (1966). Availability versus accessibility of information in memory for words. Journal of Verbal Learning and Verbal Behavior, 5 (4), 381–391.

Vehtari A, Gabry J, Yao Y, & Gelman A (2018). loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models. Retrieved from https://cran.r-project.org/package=loo.

Vehtari A, Gelman A, & Gabry J (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing, 27 (5), 1413–1432. doi: 10.1007/s11222-016-9696-4

Verbeke G, & Davidian M (2009). Joint models for longitudinal data: Introduction and overview In Fitzmaurice G, Davidian G, Verbeke G, & Molenberghs G (Eds.), Longitudinal data analysis (p. 319–326). Boca Raton: Chapman & Hall/CRC.

Voelkle MC, Brose A, Schmiedek F, & Lindenberger U (2014). Toward a unified framework for the study of between-person and within-person structures: Building a bridge between two research paradigms. Multivariate Behavioral Research, 49 (3), 193–213. [PubMed: 26735189]

Wagenmakers EJ, & Farrell S (2004). AIC model selection using Akaike weights. Psychonomic Bulletin and Review, 11 (1), 192–196. doi: 10.3758/BF03206482 [PubMed: 15117008]

Wagenmakers EJ, Lodewyckx T, Kuriyal H, & Grasman R (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage-Dickey method. Cognitive Psychology, 60 (3), 158–189. doi: 10.1016/j.cogpsych.2009.12.001 [PubMed: 20064637]

Wagenmakers EJ, Love J, Marsman M, Jamil T, Ly A, Verhagen J, … Morey RD (2018). Bayesian inference for psychology. Part II: Example applications with JASP. Psychonomic Bulletin and Review, 25 (1), 58–76. doi: 10.3758/s13423-017-1323-7 [PubMed: 28685272]

Wagenmakers EJ, Marsman M, Jamil T, Ly A, Verhagen J, Love J, … Morey RD (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. Psychonomic Bulletin and Review, 25 (1), 35–57. doi: 10.3758/s13423-017-1343-3 [PubMed: 28779455]

Walters RW (2015). Mixed-effects location-scale models for conditionally normally distributed repeated-measures data. ProQuest Dissertations and Theses, 234.

Walters RW, Hoffman L, & Templin J (2018). The Power to Detect and Predict Individual Differences in Intra-Individual Variability Using the Mixed-Effects Location-Scale Model. Multivariate behavioral research, 53 (3), 360–374. [PubMed: 29565691]

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

Watts A, Walters RW, Hoffman L, & Templin J (2016). Intra-individual variability of physical activity in older adults with and without mild Alzheimer's disease. PLoS ONE, 11 (4). doi: 10.1371/journal.pone.0153898

Woodrow H (1932). Quotidian variability. Psychological Review, 39 (3), 245.

Yao Y, Vehtari A, Simpson D, & Gelman A (2017). Using stacking to average Bayesian predictive distributions. , 1–78. doi: 10.1214/17-BA1091

Zhang Y, & Yang Y (2015). Cross-validation for selecting a model selection procedure. Journal of Econometrics, 187 (1), 95–112. Retrieved from 10.1016/j.jeconom.2015.02.006. 1016/j.jeconom.2015.02.006 doi: 10.1016/j.jeconom.2015.02.006

Zhang Z, Davis HP, Salthouse TA, & Tucker-Drob EM (2007). Correlates of individual, and age-related, differences in short-term learning. Learning and Individual Differences, 17 (3), 231–240. doi: 10.1016/j.lindif.2007.01.004

Zimprich D, Rast P, & Martin M (2008). Individual difference in verbal learnin in old age (Hofer SM & Alwin D, Eds.). Thousand Oaks: Sage Publications.

Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM, Walker ZI, & Smith S (2007). Mixed Effects Models and Extensions in Ecology with R Extensions in Ecology with R Mixed Effects. Springer Science & Business Media. doi: 10.1111/j.1467-985X.2010.00663{_}9.x
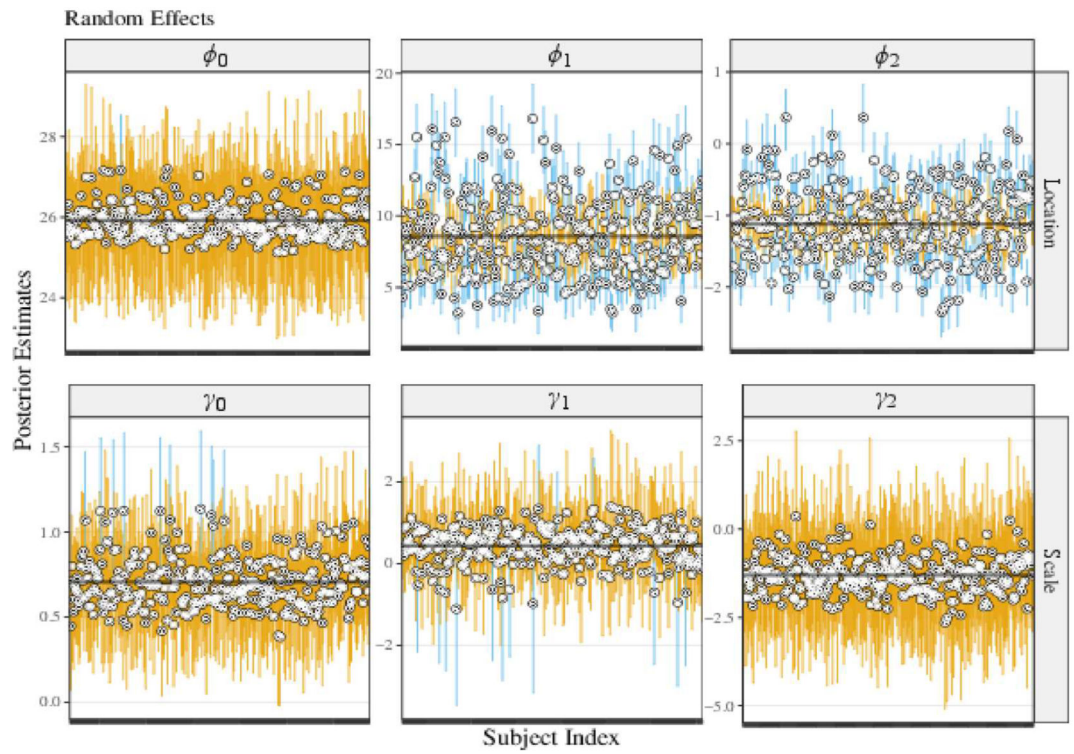
Figure 1.
Random effects centered at the population average. A blue error bar indicates that the respective 90% CrI excluded the population average. ϕ_0 and γ_0: asymptote. ϕ_1 and γ_1: "intercept" (estimate at trial 1). ϕ_2 and γ_2: rate of change.
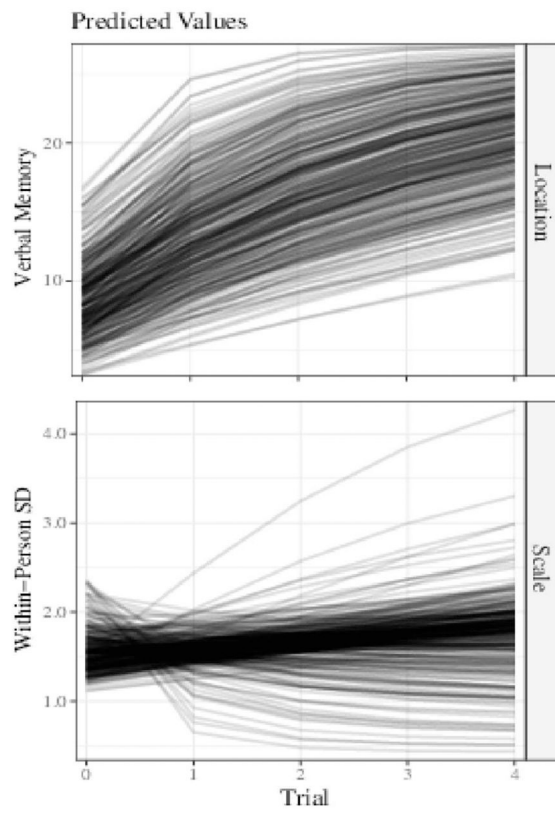
Figure 2 .
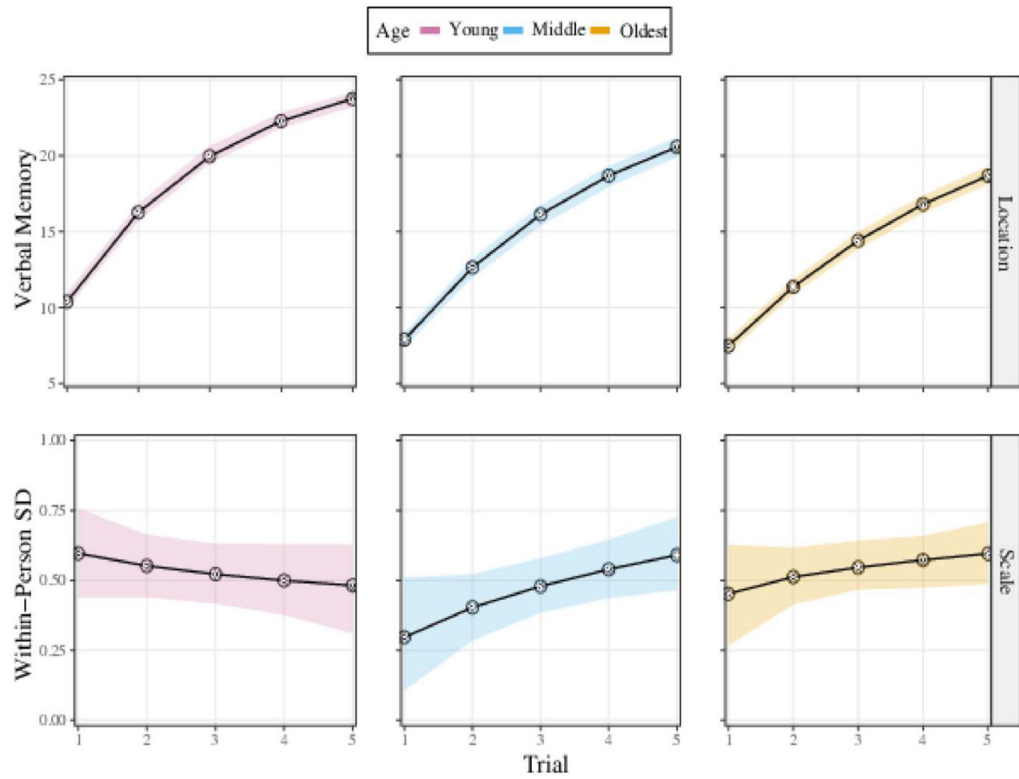Predicted values, using the mean of the posterior distribution, for the location and scale.

Figure 3 .
Predicted values, computed from the posterior distributions, for the location and scale components. The panels (and colors) correspond to the three age groups

Table 1

Model comparison results

Model	LOO$_{diff}$	SE	Pseudo-BMA+
NL-MELSM 1	-	-	0.60
MELSM	713.5	46.1	0.00
NLME	71.7	19.3	0.00
NL-MELSM 2	54.5	18.3	0.00
NL-MELSM 3	13.3	6.3	0.01
NL-MELSM 4	2.7	8.7	0.39

LOO$_{diff}$: difference in LOO. SE: standard error of the difference. Positive numbers indicate NL-MELSM 1 was more accurate. The models are identified by what differs from NL-MELSM 1. NL-MELSM 1: fully parametrized model. MELSM: asssumes linearity for the location and scale. NLME: traditional nonlinear mixed-effects model. NL-MELSM 2: individual variation not allowed for the nonlinear parameters. NL-MELSM 3: linear mixed-effects model fitted to the scale. NL-MELSM 4: covariance between location and scale random effects not estimated.

Table 2

Correlations between random location and scale effects.

	1	2	3	4	5	6
1 τ_{ϕ_0}	-		Location		Location and	
2 τ_{ϕ_1}	0.25	-				scale
3 τ_{ϕ_2}	0.26	0.71	-			
4 τ_{γ_0}	−0.52	−0.49	−0.54	-		Scale
5 τ_{γ_1}	0.22	0.63	0.48	−0.37	-	
6 τ_{γ_2}	0.30	0.28	0.46	−0.34	0.19	-

ϕ_0 and γ_0: asymptote. ϕ_1 and γ_1: "intercept" (estimate at trial 1). ϕ_2 and γ_2: log rate of change.