Using group level factor models to resolve high dimensionality in model-based sampling

Stevenson, N.; Innes, R.J.; Gronau, Q.F.; Miletic, S.; Heathcote, A.; Forstmann, B. U.; Brown, S.D.

Psychological Methods

Using Group Level Factor Models to Resolve High Dimensionality in Model-Based Sampling

Niek Stevenson, Reilly J. Innes, Quentin F. Gronau, Steven Miletić, Andrew Heathcote, Birte U. Forstmann, and Scott D. Brown

Online First Publication, June 24, 2024. https://dx.doi.org/10.1037/met0000618

AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Methods

ISSN: 1082-989X

https://doi.org/10.1037/met0000618

Using Group Level Factor Models to Resolve High Dimensionality in Model-Based Sampling

Niek Stevenson[1], Reilly J. Innes[1], Quentin F. Gronau[1, 2], Steven Miletić[1], Andrew Heathcote[1, 2], Birte U. Forstmann[1], and Scott D. Brown[2]

[1] Department of Psychology, University of Amsterdam

[2] School of Psychology, University of Newcastle

Abstract

Joint modeling of decisions and neural activation poses the potential to provide significant advances in linking brain and behavior. However, methods of joint modeling have been limited by difficulties in estimation, often due to high dimensionality and simultaneous estimation challenges. In the current article, we propose a method of model estimation that draws on state-of-the-art Bayesian hierarchical modeling techniques and uses factor analysis as a means of dimensionality reduction and inference at the group level. This hierarchical factor approach can adopt any model for the individual and distill the relationships of its parameters across individuals through a factor structure. We demonstrate the significant dimensionality reduction gained by factor analysis and good parameter recovery, and illustrate a variety of factor loading constraints that can be used for different purposes and research questions, as well as three applications of the method to previously analyzed data. We conclude that this method provides a flexible and usable approach with interpretable outcomes that are primarily data-driven, in contrast to the largely hypothesis-driven methods often used in joint modeling. Although we focus on joint modeling methods, this model-based estimation approach could be used for any high dimensional modeling problem. We provide open-source code and accompanying tutorial documentation to make the method accessible to any researchers.

Translational Abstract

In this article, we present a method of model estimation—which is useful for estimating complicated models—and more importantly, allows estimation of relationships between factors within the model. Here, we focus on joint modeling applications, where researchers have used a variety of methods (generally correlational approaches) to show links between components—such as the correlations between brain activity and observed behavior. In estimating a factor structure within our model, we can use the factor loading matrix to assess the extent to which latent behavioral factors (such as speed of processing or caution in decision making) relate to observed neural activity. This method is underpinned by a sophisticated model estimation technique and allows the user to specify any subject-level model (where the likelihood function is available; such as decision-making models, signal detection theory models, reinforcement learning models, models of blood-oxygenation level dependent responses, etc.). We provide freely available complementary code and analysis scripts for simulation and recovery (which can be used for sample size planning). Furthermore, we highlight alternate use-cases for factor loading parameterization, so that the user is well informed on which method is best suited for their application. Finally, we provide accessible code for model comparison methods, which allows the user to compare subject-level models (such as different parameterizations representing different hypotheses) and group level models (such as the ideal number of factors needed). We hope that researchers find this method accessible and informative for a variety of model-based research questions.

Keywords: joint modeling, factor analysis, dimensionality reduction, Hierarchical estimation

Niek Stevenson https://orcid.org/0000-0003-3206-7544

Reilly J. Innes https://orcid.org/0000-0003-0382-0154

Reilly J. Innes and Niek Stevenson share joint first authorship. This research was supported by European Research Council (ERC) Consolidator (to Birte U Forstmann), NWO Vici Grant (to Birte U Forstmann), and Australian Research Council Discovery Project (to Scott D Brown; ARC DP210103873). Andrew Heathcote was supported by an Advanced ERC grant (Wagenmakers 743086 UNIFY) and a Vidi grant (VI.Vidi.191.091) from the Dutch Research Council (NWO).

Code for the simulations and applications presented in the current article can be found at https://osf.io/pn3ca/. Code to apply the methods to future research can be found at https://github.com/niekstevenson/hierarchical-factor-analysis.

Correspondence concerning this article should be addressed to Niek Stevenson, Department of Psychology, University of Amsterdam, Amsterdam, Netherlands. Email: n.r.stevenson@uva.nl

Theory

Understanding the links between the brain and behavior is key to the field of model-based cognitive neuroscience. It uses computational cognitive models to understand the relationship between latent cognitive processes (represented by model parameters) and neural activity, to provide a "bridge locus" (Teller, 1984). For example, Forstmann et al. (2008) and van Maanen et al. (2011) showed correlations between activity in the striatum and caution (i.e., requiring more evidence before making a decision) in a choice response time task. There has been increasing interest in relating models of different tasks (Guan et al., 2020; Steyvers & Benjamin, 2019) or different modalities (Forstmann et al., 2008) to one another using correlational approaches. Most such studies correlated the parameters of the models to each other in a second, independent, step of analysis. However, these post hoc correlational approaches are susceptible to attenuation of the true correlation (Matzke et al., 2017; Rouder et al., 2023). Recent work has highlighted that in order to best account for measurement error and uncertainty about relations between different tasks or modalities the relationships should be estimated simultaneously with the models (Turner, Forstmann, et al., 2013; Wall et al., 2021). This approach is commonly referred to as "joint modeling" (Turner et al., 2019).

Joint modeling is a focus of model-based cognitive neuroscience, where links are drawn between neuroimaging data (from methods such as electroencephalography and functional magnetic resonance imaging [fMRI]) and nonlinear cognitive models of behavior (Forstmann & Wagenmakers, 2015; Palestro et al., 2018; Turner et al., 2019). Estimation methods have improved over time, with recent advances targeting efficient Bayesian hierarchical joint models (Turner, Forstmann, et al., 2017; Turner, Wang, et al., 2017). The strength of this approach is that the parameters of cognitive models are constrained by patterns in neural data, and simultaneously models of neural data are constrained by behavior. This reciprocal relationship provides a testable link between brain and behavior (Turner et al., 2019). Despite the notable advantages of this approach, there are as yet only a few applications likely due to its interdisciplinary nature, requiring an understanding of behavioral and neural models, complicated or inaccessible model estimation methods, and difficulty in estimation with lower numbers of subjects and trials due to the nonlinearity in cognitive models and the high dimensionality of the joint model.

In this article, we propose a method of estimating hierarchical Bayesian models that is broadly applicable and enables a substantial dimensionality reduction compared to current state-of-the-art joint-modeling methods through a factor analysis at the group level. It assumes no specific subject-level model, allows for factor analysis on the model parameters, estimates efficiently and without bias for both group and individual level effects even with nonlinear cognitive models, and provides deeper psychological interpretability. Furthermore, we provide easily implementable methods for model comparison and show various ways to specify constraints to more accurately represent the researcher's hypothesis. We apply the method to previously analyzed data to demonstrate its benefits and estimation properties. We provide freely available accompanying code and user guides.[1]

Approaches to Joint Modeling

Throughout this article, we will provide an illustrative example in which participants complete a binary-choice task with two difficulty conditions while in an magnetic resonance imaging (MRI) scanner. The behavior is described by a simple version of the most prominent model of response time in binary choice, the diffusion decision model (DDM; Ratcliff et al., 2004). The DDM assumes a single decision process that accumulates noisy evidence for and against each of two choices until the total reaches a boundary associated with one of the choices, triggering the corresponding response (see Donkin & Brown, 2018; Ratcliff et al., 2016, for introductions to the DDM and other related models, which are collectively referred to as "evidence-accumulation models (EAMs)"). The neural activity is described with a linear regression model of the observed activity of the prefrontal cortex (PFC) as measured by the fMRI signal (Poldrack, 2007). In our illustrative example, we are interested in whether the parameters of the DDM are statistically related to the activation of the PFC.

In a Bayesian context, these behavioral and neural models are often estimated in a hierarchical framework, in which individual parameter estimates (random effects) are related through an overarching group-level distribution of each parameter. Besides having desirable effects on the estimation of the parameters of the model (Rouder & Lu, 2005; Scheibehenne & Pachur, 2015), the group level also facilitates inference at the level which is usually the target for analysis in psychological science—the population.

As the majority of model-based estimation procedures tend to assume a group-level distribution only consisting of parameter mean and variance estimates (Turner, Sederberg, et al., 2013), parameters are correlated post hoc, which causes aforementioned attenuation (Matzke et al., 2017). More recent analyses estimate the full group covariance (and by extension correlation) matrix assuming a multivariate normal distribution, alleviating attenuation issues (Turner, Forstmann, et al., 2013; Wall et al., 2021).

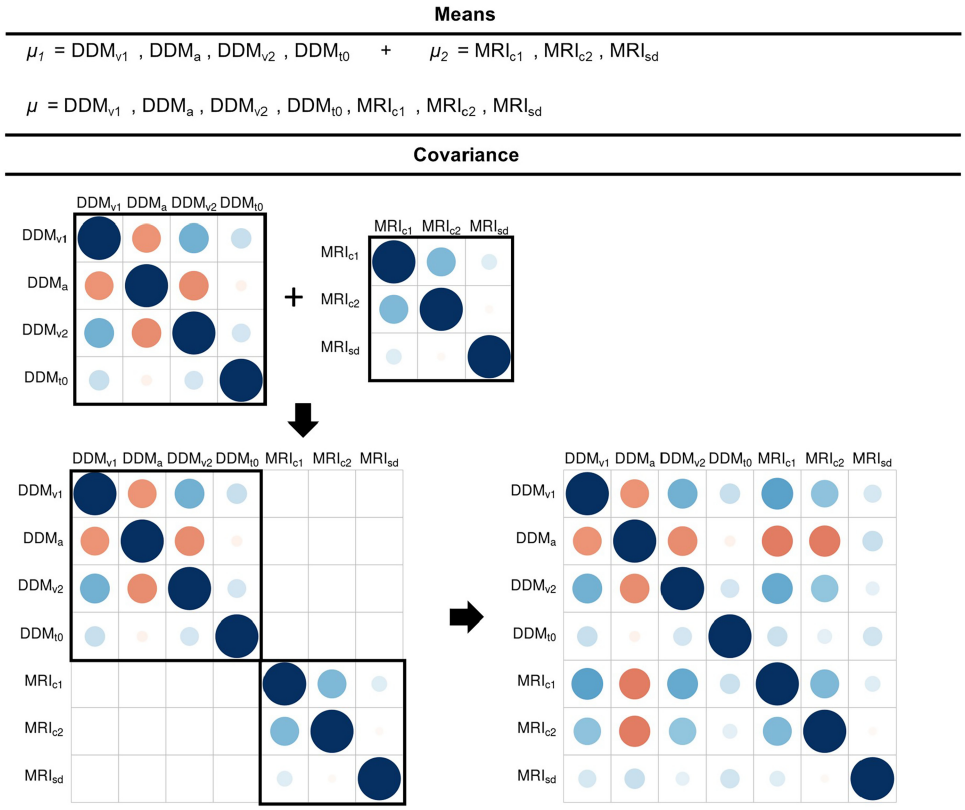
The covariance matrix provides a flexible method of linking between as well as within the components of a joint model as Figure 1 illustrates with our running example. Here, we simply extend the vector of group-level means to estimate, where the vector now includes the group-level means of both the neural and the decision-making model. Note, that also the vector of estimated random effects now encompasses parameters of both model components. Similarly, we not only estimate the covariances between parameters of the same model, but also between the parameters of the different models. This allows us to test, for example, for both difficulty conditions the correlation between neural activity in the PFC and the parameter determining the rate at which evidence accumulates (the "drift rate"), which is assumed to be larger for easier decisions and for participants with greater ability.

Wall et al. (2021) report a joint model that uses a larger covariance matrix to link three behavioral tasks. They employed a relatively new method of parameter estimation (Gunawan et al., 2017, 2020)—Particle Metropolis within Gibbs (PMwG) Markov Chain Monte Carlo sampling that we make use of here with some extensions. The PMwG method uses particle metropolis to efficiently estimate the subject-level random effects, and a Gibbs step for estimating group-level parameters. The particle sampling method simultaneously evaluates many parameter proposals on each iteration in a way that increases sampling efficiency by minimizing autocorrelation. Computational efficiency is obtained by evaluating each random effect in parallel and

[1] https://github.com/niekstevenson/hierarchical-factor-analysis. This includes several improvements to the pmwg R package that was the basis of our initial development.

Figure 1

A Graphical Example of a Joint Model Using the Covariance Approach



Note. We estimate a vector of group-level means that comprises the parameters from both the neural and behavioral model. Similarly, the covariance matrix of the joint model estimates the relationships within and between the components. Here MRI refers to the magnetic resonance imaging model parameters and DDM to the diffusion decision model parameters. See the online article for the color version of this figure.

potentially running several independent chains for the entire model in parallel, where the chains can be compared to assess convergence.

However, in the covariance approach, the number of group-level parameters grows proportionally to the square of the number of random effect parameters. This increase in dimensionality leads to lower precision in group-level estimation unless the number of participants is increased to compensate (Haines et al., 2020; Rouder et al., 2023). This problem afflicts complex behavioral models, joint models of many tasks, and especially neural models, which often examine a large number of regions of interest (ROIs), while only informed by a small number of participants (Ratcliff & Childers, 2015; Wiecki et al., 2013). To address these problems we employ a commonly used statistical method of dimensionality reduction, factor analysis. Factor analysis fits particularly well with joint neural models, where mapping behavioral processes to ROIs may not necessarily be one-to-one, but could be one-to-many (i.e., one process to many regions), many-to-one (i.e., many processes to one region) or many-to-many (Keuken et al., 2021).

Factor Analysis in Joint Modeling

Recently, Turner, Wang, et al. (2017) proposed a linking method that estimates a factor structure on the covariance matrix of a joint neural model. This method provides factors on which inference can be made and further provides a method of examining large numbers of ROIs by means of a dimensionality reduction. In particular factor loadings provide clearer inference compared to the covariance matrix, for questions such as "are regions *x*, *y* and *z* underpinned by the same latent factor." Turner, Wang, et al. (2017) show how activity from different regions loads onto factors related to speed or accuracy from behavior in a decision-making task. Importantly, the dimensionality problem is significantly reduced, with fewer parameters to estimate (as is explained later in "Dimensionality Reduction").

The factor loading matrix can be used to obtain more interpretable group-level posterior estimates. For example, it can be used to examine specific hypotheses where parameters load onto shared underlying latent factors which may underpin similar cognitive-behavioral mechanisms. To take our illustrative example, we do not formulate a hypothesis about separate covariances between DDM drift rates and neural activity for the different difficulty conditions, rather we are more interested in a common factor linking drift rates and neural activity.

Some studies have used a two-step approach to relate parameter estimates of cognitive models to one another using factor analysis (Lerche et al., 2020; Schmiedek et al., 2007; Schmitz & Wilhelm, 2016; Weigard et al., 2021). However, just as with simple correlation methods, this two-step factor analysis approach will suffer from

attenuation (Matzke et al., 2017; Rouder et al., 2023). The joint model used by Turner, Wang, et al. (2017) and Kang et al. (2021) does not suffer from such attenuation. However, the authors employ simplifications of the factors and estimated model parameters that make very strong assumptions, specifically that the parameters of the cognitive model dictate the architecture of the linking function. Hence, the cognitive model drives both the neural and behavioral data, which in turn means that this method is difficult to generalize to alternate model-based questions. Using a method where factors are estimated as part of the group-level model, as opposed to the individual-level cognitive model, enables a more flexible exploration of the pattern of factor loadings, and leads to informed group-level factor structures that are not explicitly constrained by one modality of the data.

Additionally, in Turner, Wang, et al. (2017), a method of data pooling was used, where single-trial parameter estimates are obtained, and the data were analyzed as one single subject, where each trial informs the "group" level. This method of pooling increases data, and consequently certainty, at the group level, but ignores clustering in the data due to individual differences, which can affect the precision of the estimates (Gelman, 2006; Rouder & Lu, 2005). This model is poorly informed at the "lower" level of the hierarchy, as single-trial parameters are constrained by only one observation, compared to estimating individual subject effects where each individual has many observations.

An alternative method of introducing factor analysis in joint modeling, which also addresses the post hoc factor analysis concerns raised above, is to use a hierarchical Bayesian framework. In this framework, we estimate psychological model(s) for each individual, and across individuals, these model parameters are reciprocally linked to a group level, which describes the distribution of the individual level parameter estimates. Here we propose to use a factor structure as this group-level distribution. By using a hierarchical approach, rather than pooling different subjects' averages, or relying on individual subject fits, we respect the individual differences between participants and their effect on the group-level estimates (Lee, 2011).

To construct such a hierarchical Bayesian factor model, we employ the PMwG sampling method as presented in Gunawan et al. (2020). Rather than assuming a multivariate normal distribution with a variance–covariance matrix as they assumed, the relationships between parameters are captured in a factor structure, which can be formalized as the researcher deems appropriate. We label the multivariate normal model "PMwG-MVN" and the factor model "PMwG-FA." The model assumed at the group level is then a factor decomposition of the variance–covariance matrix. For the sampling method, the only change to Gunawan et al. (2020) PMwG-MVN is an alteration to the Gibbs step. Hence, this represents a change only to the group-level model, meaning desirable properties, such as the efficient particle sampling regime, are maintained.[2]

We also provide a method that approximates integration over the prior to obtain the marginal likelihood (otherwise known as the probability of the data given the model), which allows the researcher to compare different group-level models (e.g., with different numbers of factors) using Bayes Factors, the golden standard in Bayesian model comparison (Kass & Raftery, 1995). We supply accompanying user-friendly code for constructing these hierarchical factor models using PMwG, where the individual-level model can be any model specified by the user.

Aims and Article Overview

Our aim is to provide a suite of approaches for model estimation that complement, and extend on, existing methods, and give researchers the freedom to tailor an analysis to their needs, drawing from a toolbox of easily implementable joint modeling techniques. The main body of the article highlights the hierarchical factor analysis method, which is applicable to any subject-level model. The method describes a standard factor analysis approach conducted on the group-level model using PMwG, adding to the Gibbs step to estimate a decomposition of the covariance matrix which allows for accurate estimation with less information. Furthermore, we show default settings for factor loading constraints to enable the identification of the factor loading matrix, as well as more hypothesis-driven constraints. In the appendices, we highlight several other hierarchical models for overcoming the dimensionality problem. These are less applicable to joint modeling, but may be useful in alternative applications.

The remainder of the article is structured as follows; first, we detail each of the methods. Secondly, we show simulation and recovery exercises across these methods and for a variety of questions. Finally, we show applications of these methods to previously collected data, before concluding the article.

The methods outlined in the following section can be broken into three overlapping parts; factor analysis, factor constraint, and model comparison. The primary method discussed in the article proposes the use of a factor analysis decomposition for the group level of a hierarchical model. The contribution of this method is dimensionality reduction and estimating underlying factor structures between parameters of a model. The next section provides information about the different ways factor analysis can be constrained. The main contribution here is to provide an understanding of the best use cases across different factor parameterizations. Finally, model comparison builds on marginal likelihood estimation methods by Tran et al. (2021), which enable researchers to unbiasedly estimate the marginal likelihood of models of varying complexity (i.e., two factor compared to three factor models) and with different group level models (i.e., PMwG-MVN compared to PMwG-FA). Additional estimation methods, such as single-subject estimation using methods based on PMwG, are shown in the Appendix, with these methods representing smaller contributions for more case-specific applications.

Factor Analysis Methods

Factor analysis models are commonly used tools for reducing the number of parameters estimated in a variance–covariance matrix through estimating latent "factors" that generate the observed data (Lawley & Maxwell, 1971). These methods provide an approximation for a covariance (or correlation) matrix. This means that we can approximate the matrix with many fewer parameters, as well as

[2] In the PMwG algorithm, each participant's random effect vector is estimated using particle metropolis, where on each iteration, particles are proposed from a mix of the group distribution (assumed multivariate normal with group mean and group variance) and the individual distribution (assumed multivariate normal with random effect mean and group variance weighted by a tuning parameter labeled as ϵ). This allows many possible parameter values to be compared on each iteration, with both individual and group distributions informing the subsequent particle draws.

identify latent factors. Factor analysis methods have a long history in psychological sciences, and have more recently been applied to problems in cognitive science (Lerche et al., 2020; Schmiedek et al., 2007; Schmitz & Wilhelm, 2016; Weigard et al., 2021), since they provide a representation of latent constructs that can help to better account for interrelationships between sources of data or model parameters. In estimating a specified number of underlying factors, information is provided about the relationships between the estimated parameters and the latent factors, where different factor loadings represent the strength of associations.

We propose a factor structure decomposition for the group level variance–covariance matrix assumed by Gunawan et al. (2020). The advantage of this method compared to a regular variance–covariance matrix is that we can still observe an approximation of the variance–covariance matrix (through reconstructing the elements of the factor structure), but fewer parameters are estimated. We can also test the number of underlying factors, and we can use the factor loadings to aid interpretation. We use the same underlying sampling mechanics as PMwG (Gunawan et al., 2020), with group mean and variance parameters denoted θ_μ and θ_Σ.

For the implementation of the factor analysis on the group level, we rely on the Gibbs sampler as described by Ghosh and Dunson (2009). The authors show that this parameter-expanded Gibbs sampler has better sampling properties compared to standard approaches. Similar to other studies that implemented Bayesian factor analysis we relied on Gibbs sampling rather than accept-reject sampling, because of the increased estimation performance, especially under high dimensionality (Bhattacharya & Dunson, 2011; Ghosh & Dunson, 2009; Lopes & West, 2004; Smith & Roberts, 1993; Song & Lee, 2001).

The group level factor model for $N_{1,\ldots,n}$ subjects, $P_{1,\ldots,p}$ parameters, and $K_{1,\ldots,k}$ factors is:

$$Y = H\Lambda^{\mathrm{T}} + \epsilon, \tag{1}$$

Y is a matrix of the random effects (i.e., the subject level) with N rows and P columns. Λ is the factor loading matrix of dimensions $P \times K$. H is a matrix of $N \times K$ where the rows are the factor scores η_n for each individual. The factor scores are distributed as $\eta_N \sim \mathcal{N}_K(0, \Psi)$, where Ψ is a diagonal matrix with K elements, $\psi_{1..K}$. Lastly, $\epsilon \sim \mathcal{N}_p(0, \Sigma)$ describes the residual errors, with diagonal elements $\Sigma = \sigma^2_{1,\ldots,P}$. We further estimate μ, a vector with length P, which contains the mean difference from 0 for each parameter, since factor analysis assumes zero-centered input. Importantly, μ describes the group level mean θ_μ for our hierarchical factor model. The covariance matrix of Y, Ω, is then described by:

$$\Omega = \Lambda\Psi\Lambda^{\mathrm{T}} + \Sigma, \tag{2}$$

Λ represents the factor loading matrix of size $P \times K$. Ψ represents the diagonal variance–covariance matrix of the latent factors of size $K \times K$, which is the parameter expansion proposed by Ghosh and Dunson (2009) to improve estimation properties. Λ^{T} is the transpose of Λ. Lastly, Σ represents the residuals matrix of size $P \times P$, where generally only the diagonal elements are estimated. More details on factor analysis estimation are shown below. Due to the parameter expansion from Ghosh and Dunson (2009), for interpretation, the sampled Λ is generally converted back to its nonparameter expanded equivalent Λ_{true}:

$$\Lambda_{\text{true}} = \Lambda\Psi^{1/2}. \tag{3}$$

The Gibbs steps are described in Table 1. We label this method "PMwG-FA," however, note that the random effects sampling method is unchanged (i.e., general PMwG as described in Gunawan et al. (2020) and Gunawan et al. (2017). Here, "FA" merely represents the factor analysis assumed for the group-level model.

To implement the factor structure in PMwG-FA, the only changes made are at the group level Gibbs step, where, rather than three steps (i.e., estimate mean, variance, and weights), there are five steps (i.e., estimate means, factor loadings, factor variance, factor scores, and residual variance), as shown in Algorithm 1. Before the subsequent iteration (at the end of the Gibbs step), we reconstruct θ_Σ (using Equation 2) so that it can be used as in PMwG-MVN. The group level means θ_μ, together with the group-level variance–covariance matrix θ_Σ, which is reconstructed from the factor structure, are used in the hierarchical model as the prior for the random effects α (Lee, 2011). We showcase this decomposition of the group level variance–covariance matrix for our earlier introduced illustrative example in Figure 2, although given the small parameter space, the gain in reduced parameters is small in PMwG-FA compared to PMwG-MVN.

By specifying an adequate number of factors given the full number of parameters, the reconstructed covariance matrix provides a good approximation to the full covariance matrix (Lopes, 2014). As shown in the simulation and parameter recovery section and online appendix the associated parameters are well estimated. Indeed, the recovered covariance matrix is often more accurate in PMwG-FA than PMwG-MVN, due to fewer parameters being estimated. Consequently, the algorithm enables simultaneous estimation of group-level mean parameters θ_μ and random effects α, as well as the underlying factor structure Λ and an approximation of the group-level covariance matrix θ_Σ.

All of these components are useful for model-based inference for psychological questions, however, the factor structure is especially useful in joint-modeling approaches. This is because the user can specify and test hypotheses about expected factor loadings or the number of potential latent factors. Alternatively, the user could analyze the full covariance matrix, which may be more precisely estimated in some applications (compared to using PMwG-MVN) when there is less group-level information available. In the present version of PMwG-FA, the number of factors must be specified, and models with different numbers of factors can be evaluated through model comparison methods (which are detailed below). Future work will address automatic methods of choosing a number of factors.

Table 1
Sampling Steps

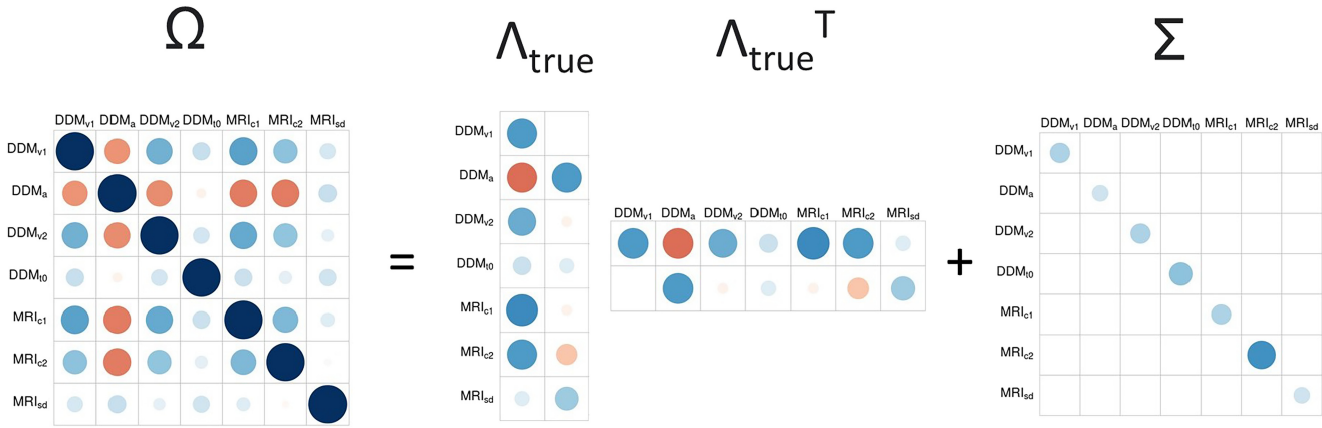
Algorithm 1 Simplified sampling procedure

1. Select initial values for $\alpha_{1:\,S}$
2. Sample $\mu \mid \alpha_{1:\,S}, H, \Lambda, \Sigma^{-1}$ from $\mathcal{N}(\mu_\mu, \sigma_\mu)$
3. Sample rows of $H \mid \alpha_{1:\,S}, \Psi^{-1}, \Lambda, \Sigma^{-1}$ from $\mathcal{N}(\mu_H, \sigma_H)$
4. Sample diagonal elements of $\Sigma^{-1} \mid \alpha_{1:\,S}, H, \Lambda$ from $Gamma(a_{\Sigma^{-1}}, b_{\Sigma^{-1}})$
5. Sample rows of $\Lambda \mid \alpha_{1:\,S}, H, \Sigma^{-1}$ from $\mathcal{N}(\mu_\Lambda, \sigma_\Lambda)$
6. Sample diagonal elements of $\Psi^{-1} \mid H$ from $Gamma(a_{\Psi^{-1}}, b_{\Psi^{-1}})$
7. Construct θ_Σ using $\Omega = \Lambda\Psi\Lambda^{\mathrm{T}} + \Sigma$
8. Sample $\alpha_{1:\,S}$
9. The sampled $\alpha_{1:\,S}$ now serve as new input for steps 2:6

Note. For the full sampling distributions at each step, see Appendix B. Steps 1, 8, and 9 are as described in Gunawan et al. (2020).

Figure 2

A Graphical Example of How We Can Reconstruct the Covariance Matrix of Our Illustrative Example Using the Factor Components



Note. Note that superscript T denotes the transpose. Here MRI refers to the magnetic resonance imaging model parameters and DDM to the diffusion decision model parameters. See the online article for the color version of this figure.

Factor Analysis Constraint and Sign-Switching Problem

Constraint is an important consideration when estimating the factor decomposition of the variance–covariance matrix. The factor loadings matrix Λ is often constrained in some way to ensure it is not invariant under rotations (Ghosh & Dunson, 2009). This occurs because one can obtain an identical Ω by multiplying Λ by an orthonormal matrix (Seber, 2009). To ensure an identifiable Λ, common approaches assume that Λ is estimated as a lower triangular matrix, with positive diagonal elements (Aguilar & West, 2000; Geweke & Zhou, 1996).

Two approaches exist to enforce strictly positive diagonal elements. First, one could constrain the diagonal elements of Λ to be fixed to a constant (often 1), which ensures that all other elements of that loading column are relative to the first entry. We will refer to this approach as a diagonal constraint. Second, Ghosh and Dunson (2009) describe a column-wise post hoc solution to enforce positive diagonal elements, which is to switch the sign of all elements in a column of Λ if the diagonal entry of that column is negative. Consequently, all elements of the loadings column are only relative in sign to the first element of that column. We will refer to this approach as applying minimal constraint.

Using minimal constraint allows more estimation flexibility than constraining the diagonal elements to 1. However, this breaks down when the first element of the column corresponding to that factor is close to 0, since its entries will switch signs between iterations, where it is negative compared to positive, leading to a bimodal distribution of estimated factor loadings. To ensure that there is no "sign-switching problem" (i.e., bimodality in the posterior samples when using minimal constraint), it is often useful to make a density plot of the posterior samples of the diagonal entries of the loadings matrix. If these densities contain 0, the post hoc solution to flip all nondiagonal loadings relative in sign to the diagonal loadings, could lead to bimodal loading distributions. In these cases, it is often safer to employ diagonal constraints.

Both forms of constraints often provide useful interpretation of the factors, since the factors can be described relative to the fixed variables (Lopes & West, 2004). Even further constraints can be imposed for more hypothesis-driven research questions. For example, if the researcher has specific hypotheses that two elements are not part of the same latent factor, specific loadings can be fixed to 0. We will refer to this approach as applying extra constraints. We allow the user to set their own constraint depending on whether they are using it for data-driven or hypothesis-driven approaches.

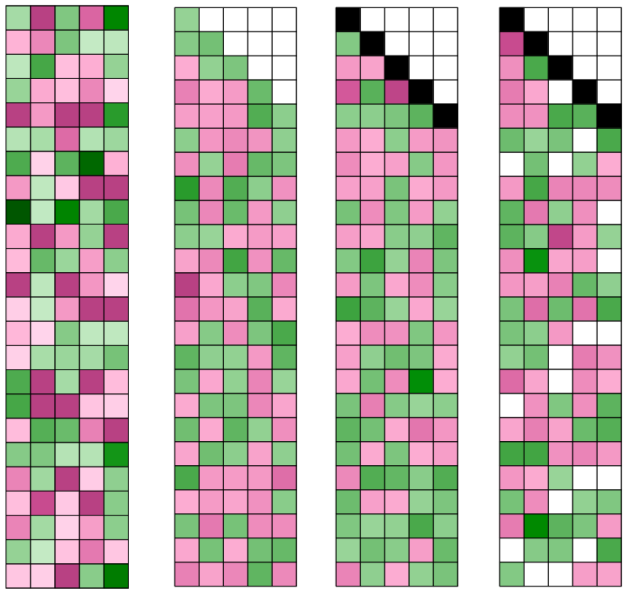
Lastly, if dimensionality, and not the underlying latent factor structure, is of importance—and no hypotheses regarding variables to fix exist—Λ can be left unconstrained. Loadings cannot be interpreted since the posterior of Λ will be subject to aforementioned rotations, however, θ_Σ is unaffected. This unconstrained method uses factor analysis purely as a tool for dimensionality reduction, and since rotations in Λ have no effect on the covariance matrix (because of the outer product with its transpose), θ_Σ can be efficiently approximated (Bhattacharya & Dunson, 2011) We represent the different methods of constraint for Λ in Figure 3.

In summary, different factor constraints can lead to different outcomes, particularly in regards to the identifiability of Λ. Which is appropriate is dependent on the desired analysis outcome or specific hypotheses. For instance, if the identifiability of Λ (i.e., the factor loadings) is important, the diagonal constraint or minimal constraint should be imposed, although for the latter this can be compromised by bimodality in the diagonal loadings. Extra constraints should be imposed if there are parameters that have no reason to load on certain factors. In the case of joint models, diagonal or minimal constraint is likely preferred as analyzing factor loadings could be key to analysis. However, when including higher dimensional data (such as neural measures), extra constraints may be imposed to handle nuisance parameters (such as drift in fMRI analyses). For safety against the sign-switching problem, and to ensure good recovery properties, we set the default constraint to the diagonal constraint case.[3] For our illustrative example we showcase how a user can specify the different forms of constraint, and check for bimodality with minimal constraint, in the file toy-example.R at https://github.com/niekstevenson/hierarchical-factor-analysis.

[3] We also save both the original Λ and transformed Λ_{true} estimates.

Figure 3

Examples of Different Kinds of Constraint, Increasing in Constraint From Right to Left



Note. White cells indicate cells fixed to 0, black cells indicate cells fixed to 1 and colored cells represent cells that are estimated. The upper triangle must always be estimated as zero if Λ is to be identified, as the first K elements are independent factors. (a) No constraint, where all elements are freely estimated. This can lead to identifiability and sign-switching issues. (b) Minimal constraint where the diagonal is freely estimated. This can lead to the sign-switching problem, however, θ_Σ shows slightly improved recovery. (c) Diagonal constraint where the diagonal is fixed at 1. This avoids the sign-switching problem. (d) Diagonal and extra constraint, where the diagonal is set to 1 and specific elements in the lower rectangle are fixed at 0. Elements fixed at 0 make specific assumptions that certain parameters do not load on factors. See the online article for the color version of this figure.

PMwG-FA Priors

In PMwG-MVN, the prior for the mean of the group level is assumed to follow a multivariate normal distribution. We use the prior on the covariance matrix specified by Huang and Wand (2013), using a scale mixture representation of the inverse-Wishart distribution with mixture weights given by an inverse-gamma distribution:

$$\begin{aligned}
&\Sigma|a_1, \ldots, a_p \sim \text{Inverse} - \text{Wishart}\ (v + p - 1,\\
&\quad 2v\text{diag}(1/a_1, \ldots, 1/a_p),\\
&a_p \sim \text{Inverse} - \text{Gamma}(1/2, 1/A_p^2),\\
&\mu_p \sim \mathcal{N}(\mu_{0\mu,p}, \Sigma_{0\mu,p})
\end{aligned} \tag{4}$$

(with $P_{1,\ldots,p}$ parameters). Using this prior structure allows us to separate the implied prior on the correlations from the prior on the (co)-variances (Huang & Wand, 2013). Here we follow Gunawan et al. (2020) and set $A = 1$ and $v = 2$, which leads to broad priors on the (co)-variances and uniform priors on the correlations. The larger A the weaker the prior on the (co)-variances. Increasing v leads to more zero-centered priors on the correlations. Decreasing v leads to more prior support for correlations away from 0 in both directions.

In PMwG-FA a different prior needs to be specified for the factor structure, which is used instead of the variance–covariance matrix. For this prior we follow Ghosh and Dunson (2009), using truncated normal diagonal elements for Λ (only if a minimal constraint is specified), normal prior distributions for the lower rectangular elements of Λ and inverse-Gamma priors for Σ and Ψ:

The conjugate priors as described in Ghosh and Dunson (2009) are

$$\begin{aligned}
&\lambda_p \sim \mathcal{N}(\mu_{0\lambda,p}\Sigma_{0\lambda,p}),\\
&\sigma_p \sim \text{Inverse} - \text{Gamma}(a_p, b_p),\\
&\psi_k \sim \text{Inverse} - \text{Gamma}(c_k, d_k),\\
&\mu_p \sim \mathcal{N}(\mu_{0\mu,p}, \Sigma_{0\mu,p})
\end{aligned} \tag{5}$$

(with $P_{1,\ldots,p}$ parameters and $K_{1,\ldots,k}$ factors). Thus, the prior on the individual factor loadings (λ) is a normal distribution, where a zero-mean represents no prior belief in the directionality of the relationship. The prior uncertainty on the relationships is governed by the variance of the prior distribution on λ. Furthermore, we use the inverse-gamma distributions as priors for the individual between factor variances (ψ) and residual variances (σ). The first parameter of the inverse-gamma distribution is the shape, which determines how peaked the prior belief is. The second parameter is the scale (also commonly referred to as the rate, which is the inverse of the scale), which determines the spread of the prior belief.

Similar to PMwG-MVN we can partially separate the prior belief on the covariances and the correlations that can be reconstructed using the factor loadings (λ) and residual variances (σ) as in Equation 2. The prior on the implied covariances is jointly driven by the prior on the factor loadings and the prior on the residual variances. However, a higher prior on the residual variances leads to an implied prior closer to 0 on the correlations. Furthermore, the transformation to λ_{true} as described in Equation 3 results in half-t distributed priors for the diagonal elements of λ_{true} and a prior t distribution for the off-diagonal elements.

The Gibbs steps that can be derived from these priors, which are used in the sampling of the group-level of PMwG-FA, are described in Appendix B.

Lastly, although any type of psychological model can be used to describe the data of each individual, both PMwG-MVN and PMwG-FA do assume that the random effects as a group are described by a (reconstructed) multivariate normal distribution, which is almost exclusively the assumption in the linear mixed modelling field (Jiang & Nguyen, 2007), which in principle the proposed methods are extensions to.

Dimensionality Reduction

As mentioned above, factor analysis is useful in cases where the number of covariances to be estimated is large and data is limited, such that the data are unable to appropriately inform covariance estimates. To be specific, when using a multivariate normal on the group level one would estimate P parameters for the mean, P parameters for the variances, and $(P \times P - P)/2$ for the covariances. Using factor analysis with K factors this is reduced to P parameters for the mean and $P(K + 1) - K(K - 1)/2$ for the covariance matrix, minus the number of additionally constrained cells (as outlined above).

This also means that given no additional constraints, the number of factors K should be chosen such that the number of free parameters (using factor analysis) does not exceed the number of free parameters in the multivariate normal analysis (as indicated in Figure 4), as this would not decrease dimensionality. Thus $P(P+1)/2 - (P(K+1) - K(K-1)/2) \geq 0$. Figure 4 shows how dimensionality increases linearly for PMwG-FA and quadratically for PMwG-MVN given the number of subject-level model parameters.

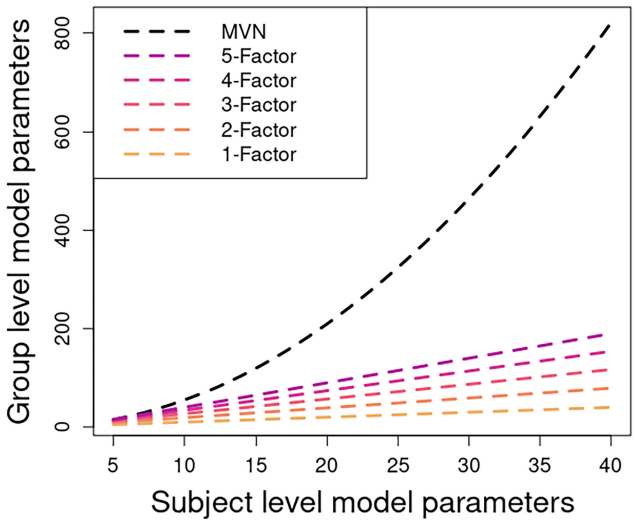
We recommend a maximum of five factors for most typical psychological modeling applications of PMwG-FA. Generally, between one and three factors is adequate. Although these guidelines are somewhat arbitrary, in most cases it is unlikely that enough latent structure exists in the data to inform six or more factors.

Model Comparison

While factor decomposition is useful for dimensionality reduction purposes, researchers face the problem of identifying the optimal number of group-level factors. Typically, when testing various hierarchical models, it is good practice to use model comparison methods (Heathcote et al., 2015; Lee et al., 2019). The challenge for model selection is assessing a model's descriptive adequacy, predictive performance, and parsimony (Gronau et al., 2020; Myung & Pitt, 1997). Given that the main comparison of interest using the PMwG-FA approach is at the group level, the model comparison method should account for both the complexity of the model, the descriptive adequacy of the model and the number of factors (which relates to parsimony and complexity of the model). Model comparison methods such as Akaike Information Criterion (Akaike, 1973), Deviance Information Criterion (Spiegelhalter et al., 2002), and Widely Applicable Information Criterion (Vehtari et al., 2017) do not account for the added complexity in the model prior of hierarchical models (Gronau et al., 2020; Tran et al., 2021).

Figure 4

Increase in the Number of Group Level Parameters to Be Estimated Relative to the Number of Subject-Level Parameters Across Different Group Level Models



Note. MVN = multivariate normal; FA = factor analysis (MVN Shown in Black, FA Shown in Colors). See the online article for the color version of this figure.

The marginal likelihood provides a solution to these issues that can be seen as a gold standard for model comparison in quantitative psychology (Kass & Raftery, 1995; Tran et al., 2021). In special cases, the marginal likelihood can be solved analytically, but generally, the marginal likelihood needs to be estimated, which can be computationally expensive. Recent advances such as bridge-sampling (Gronau et al., 2017) and importance sampling squared (IS^2; Tran et al., 2021) have enabled robust and quick methods for estimating the marginal likelihood and integrating over the prior, such that group level distributions are accounted for.

Following Tran et al. (2021), we propose IS^2 as a means of estimating the marginal likelihoods to use in comparing PMwG-FA models. The marginal likelihood is the average likelihood of the model given the data, where the average is taken over the prior distribution of model parameters. As the marginal likelihood estimate integrates across the prior space of the model, marginal likelihood estimates inherently account for both model flexibility and model complexity. To elaborate, increasing the number of estimated parameters increases the prior space, which is only worthwhile if the additionally introduced parameters increase the likelihood enough to counteract taking the average over a larger prior space. Estimating the marginal likelihood has attractive properties for the current application, in that the complexity in the factor structure is accounted for and Bayes factors can be easily calculated to quantify model comparison.

IS^2 works by first using posterior samples from PMwG-FA to inform an importance distribution for the parameters. This importance distribution is constructed by fitting a Student t distribution to posterior Markov chain Monte Carlo samples. We then construct conditional proposal parameters—labeled "particles"—for each subject. The marginal likelihood is then estimated unbiasedly for each particle, which is combined with the importance distribution. Hence, the importance sampling procedure is in itself an important sampling procedure that can be used to estimate the likelihood. In comparison to Tran et al. (2021), who used IS^2 for comparison of models estimated using PMwG-MVN, the only changes are at the group importance level, where the prior is the same as used for estimation of PMwG-FA. By including the factor analysis priors in IS^2, we are able to estimate the marginal likelihood and use this to compare alternate models—models which include different number of factors, alternate group level models (i.e., PMwG-MVN vs PMwG-FA) or different parameterizations. Code for conducting IS^2 on PMwG-FA output is included in the online appendix.

We note that IS^2 provides an estimate of the marginal likelihood. To obtain a standard error of this estimate we use bootstrapping. One way to decrease the uncertainty in the marginal likelihood estimate is to gather more samples from IS^2 until you reach the desired confidence level. Throughout the article, we adapt the strategy where we first use default settings of IS^2 as provided in the online appendix. If we cannot confidently conclude that one model is preferred over the other, we increase the number of IS^2 samples until that threshold is reached.

Simulations and Recovery

To ensure the efficacy of the proposed methods, we conducted a variety of simulation and recovery studies. We tested the PMwG-FA method, as well as IS^2, to assess the selection of the generating model and recovery of its parameters including the factor

loading matrix and the full covariance matrix. The simulations use realistic participant numbers and trials per participant for neuroscientific applications, which are often small due to the costs and time involved in data collection. Similarly, the number of parameters was based on plausible fMRI joint-modeling studies (see applications below), where the cognitive model (5–10) parameters are simultaneously estimated with neural regressors for the blood-oxygenation level dependent (BOLD) activity corresponding to ROIs under different contrasts. In the following sections we provide representative results; the full set of results can be viewed in the online appendices at https://osf.io/pn3ca/.

Mean Recovery

For the first study, we looked at the recovery of the generating mean and correlation parameters. We simulated from a normal distribution with 10–50 mean parameters (increasing by steps of 10) varying between −5 and 5, and we constrained the variances to be equal to 1. We used 25–100 subjects (increasing by steps of 25), each with 200 trials per condition. At the group level, we used one to four factors, and these were estimated with either minimal or diagonal constraints imposed. For each data set, the group level Λ_{true} values were generated from a normal distribution centered at 0, with a standard deviation of 0.3, truncated at 0.99 and −0.99 (i.e., no loading value should be higher than 1 or lower than −1).[4] The random effects were generated for each person using $\eta\lambda^t + \epsilon$, with means between −5 and 5 in equal steps for n parameters. We chose to use a multivariate normal distribution for the random effects for computational efficiency, which is an important consideration given the size of the simulation studies.

Figure 5 provides an example of the mean recovery, where estimated values are shown on the x-axis and generating values on the y-axis. Points falling along the diagonal line indicate accurate recovery, as was the case here.

Factor Loadings Recovery

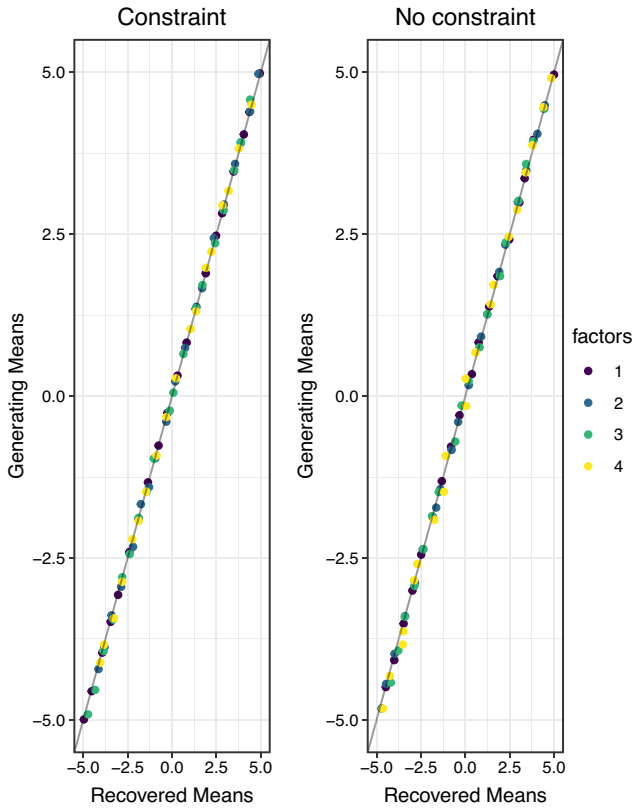
Using the same simulated data from above, we also investigated the recovered factor loadings Λ. Figure 6 provides example results. In all cases, the PMwG-FA method adequately recovered the mean factor loading, Λ_{true}, with the generating value falling within the estimated 95% credible interval for each parameter across factors (as indicated by the dashed vertical line—the mean—falling within the distributions of estimates in Figure 6).

Covariance Recovery

We also tested the recovery of the full covariance matrix using PMwG-FA. Theoretically, the choice of the minimal or diagonal constraint of the factor loadings matrix should only minimally influence the covariance recovery (Ghosh & Dunson, 2009). To check, we simulated data from an inverse-Wishart distribution for the covariance matrix. We used ranges of 10–50 parameters (incrementally increasing by 10) and 25–100 subjects (incrementally increasing by 25). We then estimated the models using PMwG-MVN, PMwG-FA, and PMwG-Diagonal (PMwG-Diag). PMwG-Diag uses PMwG sampling with independence assumed at the group level—i.e., only a diagonal estimated for the variance–covariance matrix, as shown in Appendix "PMwG-Diagonal." We compared the recovery of the PMwG-MVN to the approximated covariance from PMwG-FA given by Equation 2.

Figure 5

Parameter Recovery for Generating Parameters



Note. Generating parameters are shown on the y-axis, with corresponding estimated parameters on the x-axis. Colors of the points indicate which number of factors were used. On the left is the PMwG-FA estimated with diagonal constraint, on the right is with no diagonal constraint. This plot uses generated data from 50 subjects for 20 parameters. PMwG-FA = Particle Metropolis within Gibbs-factor model. See the online article for the color version of this figure.

We also compared these estimates to PMwG-Diag by taking the correlation between posterior parameter estimates.

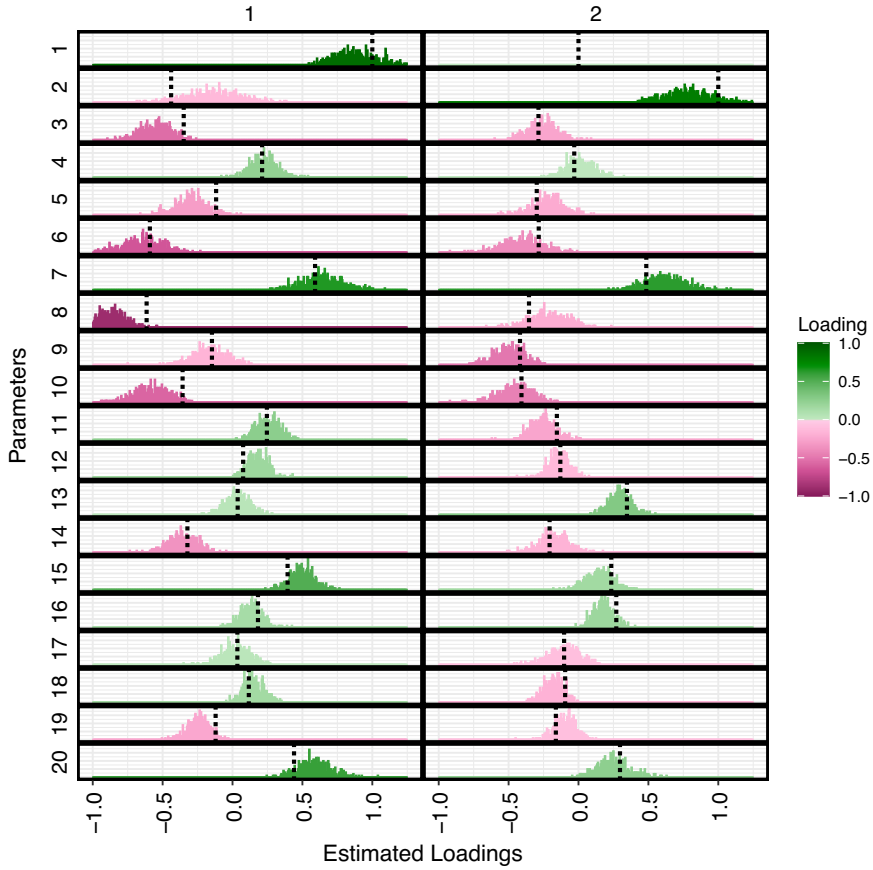
Figure 7 provides example results, where the PMwG-MVN covariance recovery is shown in the upper triangle, and PMwG-FA recovered (approximated) covariance is shown in the lower triangle, with black lines indicating the generating mean values, and red lines indicating the "recovered" PMwG-Diag covariance. To keep the figure readable we only show the first 10 parameters of the model.

As discussed by Haaf and Rouder (2019) and Matzke et al. (2017), the correlations based on the fit of the independent model are attenuated, which highlights the strength of PMwG-MVN and PMwG-FA compared to a group level model which assumes parameter independence. As expected given the generating model was a

[4] σ_{diag} was simulated from an inverse-gamma function with shape 10 and rate 2. η was generated from a multivariate normal with mean 0, standard deviation 1. ϵ was generated from a multivariate normal with mean 0 and standard deviation σ. These values were chosen such that they closely corresponded to the prior distributions. Although with our illustrative example, we show that this is not a prerequisite for recovery.

Figure 6

Posterior Samples of Λ_{true} (Factor Loadings) for a Two Factor Model (One Column for Each Factor) With 20 Parameters and 50 Subjects



Note. The dashed line shows the generating Λ_{true} values. The upper rectangle cell is at 0 due to the constraint imposed. See the online article for the color version of this figure.

multivariate normal, PMwG-MVN showed better recovery properties than PMwG-FA (i.e., it should be easier to recover the MVN if the generating values were from an MVN).

The covariance recovery study highlighted that although PMwG-MVN gave more accurate and reliable estimates, PMwG-FA adequately recovered the covariance matrix with far fewer parameters for the group-level model. Given that the generating model is MVN, the recovery of PMwG-FA is promising, especially given the significant dimensionality reduction. In psychological applications of cognitive modeling, we are unaware of the true data generating model, so given that PMwG-FA recovers the MVN generating model well, with fewer parameters, and with a latent factor structure, this represents an advance on current methods.

In the next sections, we report a model recovery study to illustrate and test the use of marginal likelihood estimators for comparing different group-level models, as well as applying our methods to actual data, where we are unaware of the true group structure.

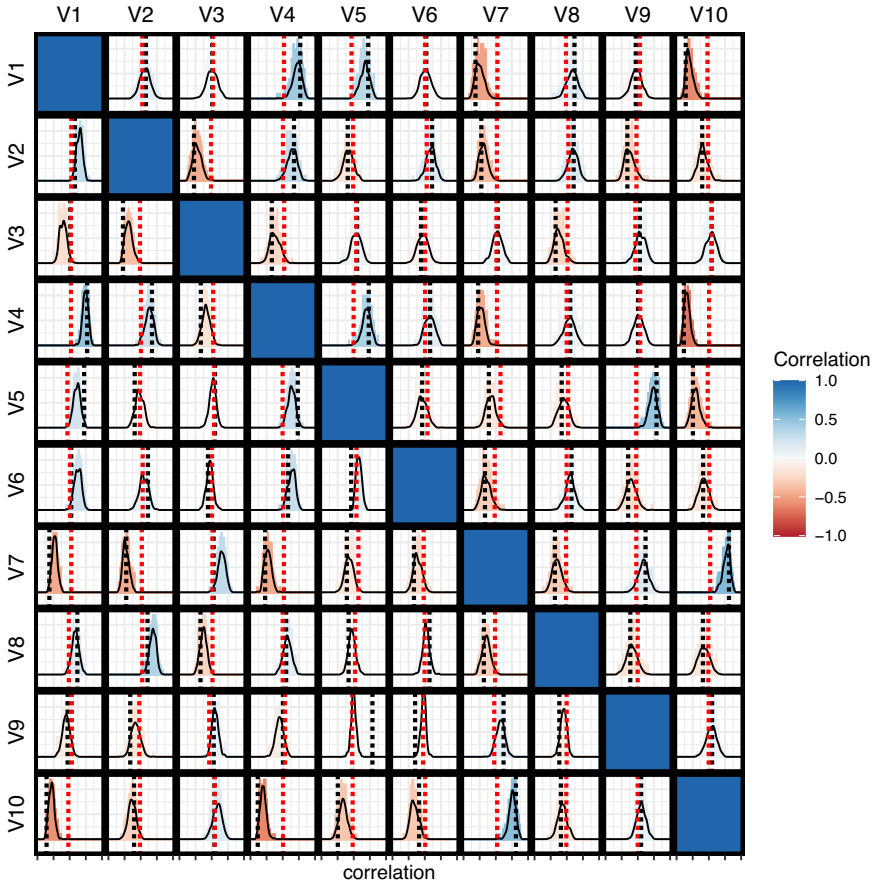
Model Recovery

This recovery exercise focused on differences in the group-level model, either MVN or FA with two factors, and also varied the number of parameters (20 or 40). The individual level model was again a normal distribution with variances constrained to 1. The group-level means varied between −5 and 5. The group-level FA and MVN models were generated with the same approach as described in Sections "Covariance Recovery" and "Factor Loadings Recovery," respectively. For each subject, 200 trials were estimated. Table 2 provides marginal likelihoods estimated by IS^2 along with bootstrapped standard errors. For each of the four simulated data sets we estimated four models; PMwG-MVN, and PMwG-FA with one, two, or three factors. Since the simulated data with a factor structure have an implied covariance matrix, PMwG-MVN should be able to recover that implied covariance matrix. However, since PMwG-MVN uses more parameters to account for the covariance matrix, the IS^2 estimate should prefer the factor models.

We found that when the data were generated from a two-factor model the two-factor model had the highest marginal likelihood estimate, indicating the good recovery properties of PMwG-FA and the precision of IS^2. Furthermore, when the data were generated from an MVN with 20 parameters, the 20-parameter MVN also had the highest marginal likelihood estimate. Lastly, when the data were simulated from an MVN with 40 parameters, PMwG-FA with three factors had the highest marginal likelihood estimate. This can be explained

Figure 7

Posterior Samples of Σ (Covariance Matrix) for a Multivariate Normal Model (Upper Triangle) Compared to the Implied Covariance for a Factor Model With Two Factors (Lower Triangle) With 20 Parameters and 50 Subjects



Note. The black dashed line shows the generating Σ values, where data were generated from a multivariate normal model. The colored dashed line shows the post hoc correlations estimated from PMwG-Diag. PMwG-Diag = Particle Metropolis within Gibbs-Diagonal. Here V1 to V10 are the simulated parameters. See the online article for the color version of this figure.

by a common phenomenon in model inversion, where simpler models are preferred in cases where there is not enough information in the data to recover the full model (for an in-depth discussion on model inference and model inversion, see Gronau et al., 2020; Lee, 2018). Given the low, yet representative, number of subjects used in the simulation study, it is, therefore, not surprising that a factor model is preferred by the marginal likelihood estimate. Note that some variance–covariance distributions generated using the inverse-Wishart distribution might resemble a factor structure more closely than others, and therefore results could vary between different random simulations.

Illustrative Example

In simulations reported so far, we used a normal model for the individual level, which provides a good estimate when assessing group-level recovery and identifiability, while being far less computationally burdensome than realistic psychological models. However, to show that we can also recover the factor structure with more complex psychological models we used our previously introduced illustrative example. In our illustrative example, 50 participants completed a simple decision-making task with two difficulty conditions while in an MRI scanner. For each participant, we simulated 200 trials and 200 time points of neural activity. For the behavioral data, the response times (RTs), and responses, we used a simplified version of the DDM:

$$\text{RTs, responses} \sim \text{DDM}(v, a, t_0), \tag{6}$$

where v is the drift rate (determining the rate of evidence accumulation, positive values favoring one response and negative values the other response), a is the boundary (determining the amount of evidence required to select a response), and t_0 the nondecision time (the combined time for stimulus encoding and response production). We used two drift rates, one for each difficulty condition. For the neural activity of each participant, we used a regression model:

$$y \sim \mathcal{N}(X\beta, \sigma), \tag{7}$$

where y is the neural activity, X is a design matrix for the time series and β is a set of regressors for each time series. We estimated two

Table 2
Model Recovery for the Four Generated Data Sets

Model	Parameters	FA_1	FA_2	FA_3	MVN
Two factors	20	−171,662.9 (0.7)	**−171,652.7 (1.0)**	−171,687.4 (2.6)	−171,869.6 (20.9)
Two factors	40	−343,797.2 (3.3)	**−343,678.9 (4.2)**	−343,753.6 (8.6)	−348,486.2 (146.4)
MVN	20	−171,846.2 (0.6)	−171,847.2 (1.8)	−171,825.9 (1.5)	**−171,787.7 (13.1)**
MVN	40	−343,979.3 (1.2)	−343,727.1 (2.3)	**−343,715.6 (3.6)**	−347,123.1 (173.1)

Note. Marginal likelihood values for each model as given by IS^2 are shown, with a bootstrap standard error shown in brackets. Boldface marginal likelihood values indicate the highest marginal likelihood for each exercise. All data sets included 30 simulated subjects, with 200 trials per condition, where each condition was a separate measure per subject (i.e., number of conditions = number of parameters).IS^2 = importance sampling squared. MVN = multivariate normal; FA = factor analysis.

regressors, one for each difficulty condition. For the group level factor model we constructed a factor structure with two factors based on a desirable result in a typical joint modeling study, such that drift rates and prefrontal activity loaded highly on the same factors across conditions, and the threshold parameter loaded highly onto a second factor. Similarly, we used plausible values for the DDM and the regression model. Individual level parameters were simulated from a multivariate normal with group-level mean and the variance–covariance matrix reconstructed from the factor structure.

Figure 8 demonstrates that the factor loadings were recovered well. Details of the illustrative example, examples of how to perform posterior inference, and how to apply different types of constraint are given at https://github.com/niekstevenson/hierarchical-factor-analysis.

Applications of PMwG-FA

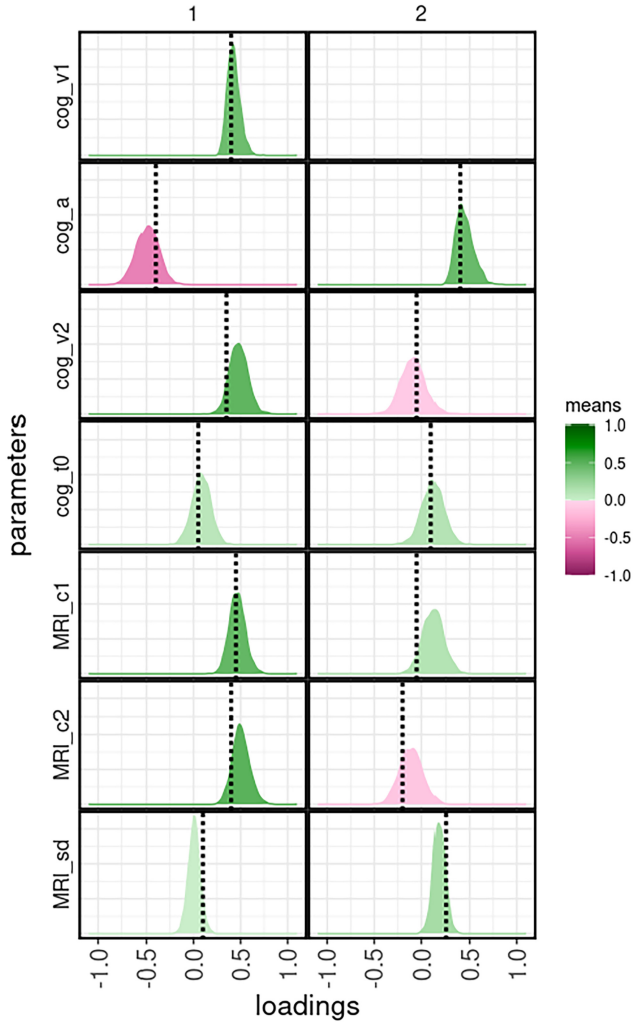
In this section, we detail three applications of PMwG-FA to archival data (Forstmann et al., 2008; van Maanen et al., 2011; Wall et al., 2021). The three data sets are from binary decision-making experiments that fit with an EAM, and include a "joint-modeling" component. In each case the EAM used is the linear ballistic accumulator (LBA; Brown & Heathcote, 2008), which shares conceptually similar parameters to the DDM used in our illustrative example, nondecision time (t_0), drift rate (v) and boundary (b). In contrast to the DDM, there are separate evidence-accumulation processes for each response, and so potentially different rates and boundaries for each. In our applications, we estimate different rates for each accumulator, with a larger rate for the correct accumulator (whose response matches the stimulus) than the error accumulator (whose response mismatches the stimulus). However, we estimate the same boundary for both accumulators, with larger boundary values producing more accurate but slower responses.

The LBA assumes noise in evidence comes from trial-to-trial variability in the starting point of evidence accumulation (A) and drift rates (s_v), but these can be challenging to estimate unless each participant performs many trials (Johansen-Berg et al., 2018). In all applications we fix $s_v = 1$ (which also makes the model identifiable, see Donkin et al., 2009). We estimate A in the first two applications, but assume $A = 0$ in the third, where participants performed fewer trials.

Application 1: In-Scanner Versus Out-of-Scanner Data

The first application used the data from Forstmann et al. (2008) (retrieved from https://osf.io/rf8nd). In this study, 19 participants completed a binary random-dot motion discrimination task both in and out of an fMRI scanner. We modeled the data from the two environments as two behavioral components of a joint model. The EAM was identical for the two components, with only the environment differing, giving us a component for each environment (i.e., model in-scanner and model out-of-scanner). The LBA model for each

Figure 8
Posterior Samples of Λ_{true} (Factor Loadings) for a Two-Factor Model of Our Illustrative Example



Note. Here, cog refers to our cognitive model parameters and MRI to our model for the simulated functional magnetic resonance imaging data. The dashed line shows the generating Λ_{true} values. See the online article for the color version of this figure.

component of the joint model has seven parameters for each subject *j*. Three boundary parameter boundaries differ between conditions with instructions that told participants to emphasize either decision speed or accuracy, or emphasized both (neutral). The correct and error accumulator drift rates are parameterized in terms of their mean, which indexes the overall urgency to respond (larger values leading to faster responses) and the correct minus error difference, which acts similarly to the DDM drift rate (larger values cause more accurate responding). Finally, there is a single nondecision time and start-point variability parameter for all accumulators and conditions;

$$\theta_j = A_j, b_{(\text{speed})_j}, b_{(\text{accuracy})_j}, b_{(\text{neutral})_j}, t_{0_j}, v_{(\text{mean})_j}, v_{(\text{difference})_j}.$$

In the following, the "in-scanner" and "out-of-scanner" components are labeled with either "In" or "Out," for example, "In-v_{mean}" or "Out-v_{mean}." Previously, Wall et al. (2021) had estimated this joint model for this data set using PMwG-MVN. In our application, the current method adds benefit by estimating latent factors, which can provide clear links between the models across environments. Furthermore, with PMwG-FA the parameter space is reduced compared to PMwG-MVN, leading to more informed group-level estimates, which is especially relevant for this application where there are relatively few subjects.

We fit PMwG-FA to the data with 1,000 iterations of burn-in, up to 5,000 iterations to adapt sampling to be efficient (although it dropped out early as conditional proposal distributions could be constructed, see Gunawan et al., 2020), and 1,000 iterations to produce the samples that were subsequently analyzed, with 200 particles for each iteration. We constrained Λ by setting the diagonal = 1 and upper triangle = 0 to ensure identifiability. Therefore, we reordered the parameters along the diagonal to ensure that no meaningful parameter relations were fixed for the factor models with more than one factor. We estimated five different (group-level) models which were either PMwG-FA (with $K = 1, \ldots, 3$ factors), PMwG-MVN, or PMwG-Diag (see the "PMwG-Diagonal" section in Appendix A).

We compared the five models using the IS^2 method (Tran et al., 2021). We first constructed the proposal distribution (with five degrees of freedom) by fitting a students *t*-distribution to the 1,000 posterior samples, and then estimated the marginal likelihood with 5,000 importance samples and 250 particles. We report median log marginal likelihood estimates—where higher values represent the favored model—and Monte Carlo standard error (obtained by bootstrapping the importance samples; Tran et al., 2021). This reporting differs slightly from Tran et al. (2021) as we found the median measure was less likely to be biased by extreme marginal likelihood values.

Results

PMwG-FA produced results consistent with the original analysis (Forstmann et al., 2008). IS^2 results provided an interesting, but not unexpected, result, with the one-factor model outperforming the competing models. Median log marginal likelihood values (with standard errors) were: one factor = 8,471.19 (0.14), two factors = 8,455.74 (0.24), three factors = 8,385.32 (0.39), MVN = 8,429.42 (1.84), and Diag = 8,354.84 (0.86). That the single factor model outperformed the other models was unsurprising given that PMwG-FA significantly reduces the estimated parameter space compared to PMwG-MVN. That the single factor model outperformed PMwG-FA models with more factors could suggest consistent behavior by subjects in and out of the scanner, with the caveat that small subject numbers may be insufficient to identify more complex models. Lastly, that PMwG-FA was preferred over PMwG-Diag was unsurprising, given that PMwG-Diag ignores group-level interrelationships.

Parameter estimates from the winning model can be seen in Table 3. Similar to previous findings there was a difference observed between the estimated speed and accuracy boundary parameters, and the neutral condition resembled the accuracy condition. Furthermore, there were mean differences over subjects observed between scanning environments, with higher t_0 and lower $v_{(\text{mean})}$ for the in-scanner environment.

Adding to the factor analysis model being preferred (by marginal likelihood estimates), the factor loadings present a deeper layer of inference compared to differences between environments in mean parameter estimates. The factor loadings can be seen in Figure 9. In both environments boundary and mean rate parameters load positively whereas nondecision time and rate-difference parameters load negatively. This suggests potential compensatory tradeoffs that appear to be somewhat weaker in than out of the scanner. First, participants with smaller rate differences, indicating lesser ability to discriminate correct from incorrect choices, compensated by setting higher response boundaries. These higher boundaries, which slow responding, are countervailed by speeding due to smaller nondecision time and higher mean rates, indicating a greater sense of urgency. Given the small number of participants, these tradeoffs, and particularly the trend for them to be larger out than in scanner need to be treated with some caution, but do illustrate the type of inference enabled by PMwG-FA.

Application 2: Data From Multiple Tasks

In Wall et al. (2021), a much larger sample, 110 participants, completed a visual search task, a stop-signal task, and a match-to-memory task, where task difficulty was manipulated over three levels within each task. We excluded data from the stop-signal task due to poor estimation of the error drift rate across all group-level models (due to very low errors) and, because the data excludes the stopping component, so the model was uninformative.

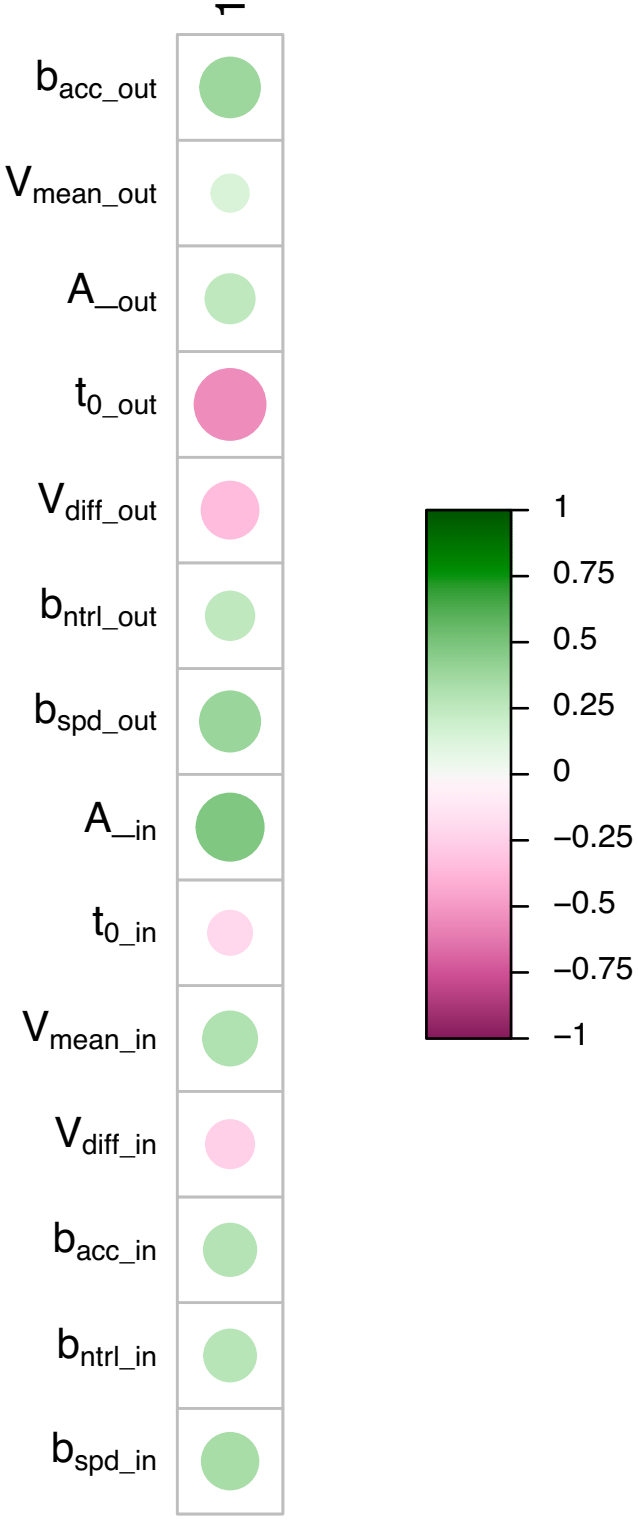
In their study, Wall et al. (2021) used the PMwG-MVN method to evaluate LBA parameter correlations, finding positive correlations between thresholds across tasks, as well as several other between-task correlations. In the following example, we highlight the added benefit of PMwG-FA, first because of the reduction in parameter space, and second, as researchers can use the latent factor

Table 3
Posterior Mean Estimates for LBA Model Parameters Across Environments, Shown on the Natural Scale

Parameter	In scanner	Out of scanner
$b_{(\text{accuracy})}$	0.58	0.92
$b_{(\text{neutral})}$	0.47	0.85
$b_{(\text{speed})}$	0.06	0.49
A	0.85	0.70
$v_{(\text{error})}$	0.43	1.73
$v_{(\text{correct})}$	2.91	3.33
t_0	0.21	0.11

Note. LBA = linear ballistic accumulator.

Figure 9
Posterior Samples of Λ_{true} for the Winning One Factor Model of the Forstmann et al. (2008) Data



Note. See the online article for the color version of this figure.

estimates in addition to the parameter correlations to make more nuanced inferences. Following Wall et al. (2021), our joint model used the same nine-parameter LBA model for each task:

$$\theta_j = A_j, b_j^{(1)}, b_j^{(2)}, b_j^{(3)}, t_{0_j}, v_{c_j}^{(1)}, v_{c_j}^{(2)}, v_{c_j}^{(3)}, v_{e_j}.$$

Each within-subject condition (denoted 1, 2, and 3) had a different boundary and a different rate for the correct accumulator. Finally, the rate for the error accumulator was assumed the same across the three conditions. For each condition, the same parameter value was estimated for start-point noise, nondecision time, and the rate of the error accumulator. The latter assumption was made because the error rate is difficult to estimate given it is mainly constrained by error responses, which were rare as accuracy was high.

The joint model was estimated for the full vector of 18 parameters (for search and match tasks), where parameters were labeled with the relative task name (e.g., "match-t_0"). We fit PMwG-MVN and PMwG-FA with 1–3 factors with 1,000 iterations of burn-in, 5,000 iterations of adaptation, and 2,000 iterations of sampling. For PMwG-FA we again applied diagonal constraint to the loading matrices for identifiability of Λ_{true}. Furthermore, we specified that there were no relationships between v_e and the other parameters, given that v_e was based on little data (i.e., we constrained the factor loadings to 0), and to illustrate how PMwG-FA is more flexible in allowing constraints to be imposed compared to the approach used by Wall et al. (2021).

Results

PMwG-FA produced results consistent with those of Wall et al. (2021), with drift and threshold increasing across task difficulty. Similar to Application 1, the one-factor model outperformed the competing models. Median log marginal likelihood values (with standard errors) were: one factor = 22,119.3 (2.3), two factors = 22,101.5 (1.4), three factors = 22,001 (3.1), and MVN = 21,932.2 (25.2).

The factor loadings of the one factor model shown in Figure 10 are consistent with the original finding that only boundaries correlate across tasks, as the factor loads mainly on boundary parameters, although more so for the search than match task. The fact that a one-factor model again provided the best account of the data with a relatively large number of participants promotes confidence in the conclusion that participants display consistent individual differences in the strategic processes they use to set boundaries.

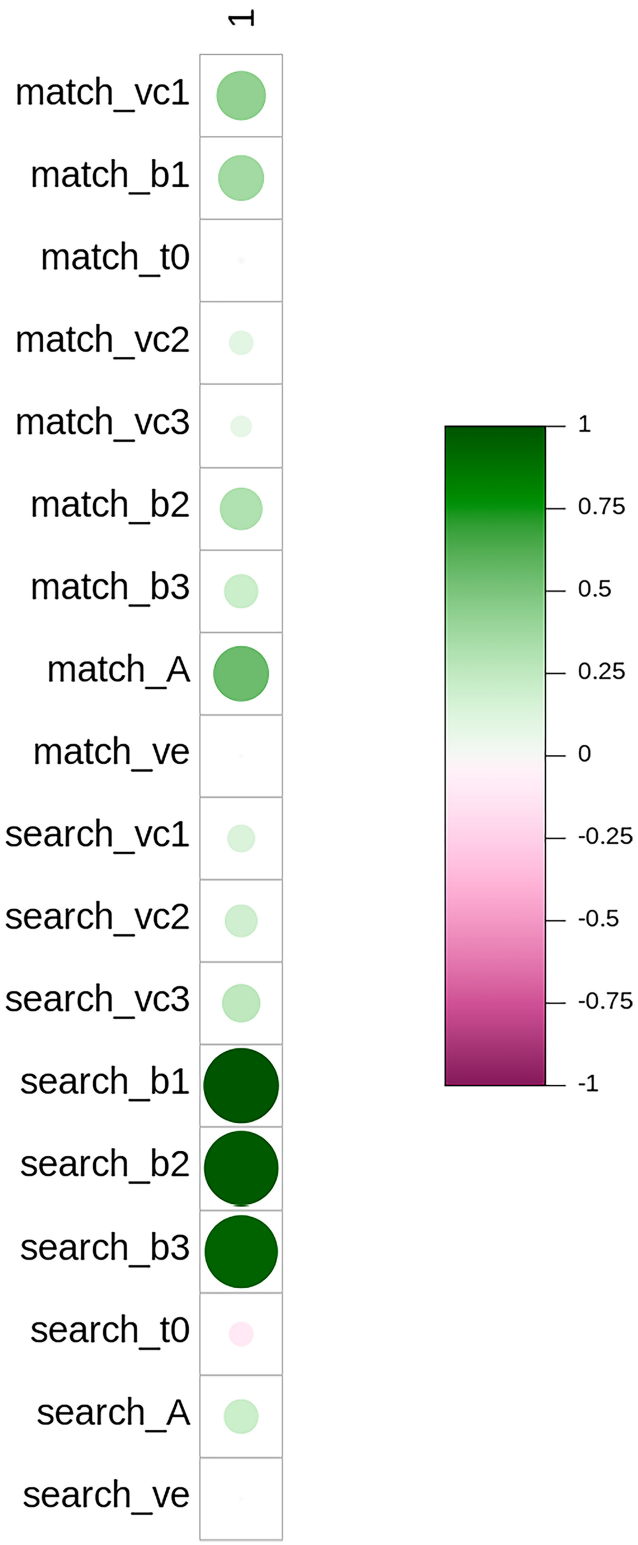
Application 3: Neural and Behavioral Data

In the third application of PMwG-FA, we constructed a joint model of fMRI scans and behavioral data from 17 participants reported in van Maanen et al. (2011). The participants completed a binary random-dot motion task with a speed versus accuracy manipulation similar to Application 1 but without the neutral condition. For this example, the key aspects of interest are the commonalities observed in functional neuroimaging data across the speed and accuracy emphasis conditions.

For the fMRI data, we constructed general linear models for the observed activity of five ROIs during speed or accuracy trials; the anterior cingulate cortex (ACC), left and right striatum (l/r STR), and left and right presupplementary motor area (l/r preSMA). The

Figure 10

Posterior Samples of Λ_{true} for the Winning One-Factor Model of the Wall et al. (2021) Data



Note. Here the model parameters were labeled with the relative task name (e.g., "match-t0"). See the online article for the color version of this figure.

masks of STR were obtained from the ATAG atlas (Young Adults; Keuken & Forstmann, 2015). The mask of the ACC was extracted from the Harvard–Oxford cortical atlas (Desikan et al., 2006). Finally, a discrete preSMA mask was based on coordinates provided by Johansen-Berg et al. (2004). These ROIs were selected based on the results of van Maanen et al. (2011).

Preprocessing steps were done using fMRIprep (see Appendix C; Esteban et al., 2019). For each region we first regressed out any nuisance parameters, such as scanner drift and motion correction parameters, using ordinary least squares methods (Satterthwaite et al., 2013a). We then estimated regression coefficients for the intercept, the BOLD response during speed cues, the derivative of the speed BOLD response, the BOLD response during accuracy responses, and the derivative of the accuracy BOLD response, and lastly the standard deviation, using the joint model for each of those regions (Poldrack, 2007).

Thus, our neural model comprised six parameters for each ROI j for a total of 30 neural parameters:

$$\theta_{\text{ROI}_j} = \beta_{(0)_j}, \beta_{(\text{spd})_j}, \beta_{(\text{spd,deriv})_j}, \beta_{(\text{acc})_j}, \beta_{(\text{acc,deriv})_j}, \sigma_j.$$

The regression model for a ROI for participant i can be written as:

$$y_i \sim \mathcal{N}(X_i\beta_i, \sigma_i).$$

with y_i a vector of size N, with N the number of time points. β_i a vector of P, with P the number of regressors of interest and X_i is a design matrix of N x P.

The behavioral model consisted of only 80 trials per participant, and hence we estimated a simplified LBA model that included no between-trial variation in the start point to improve estimation precision. We kept boundaries the same for each accumulator but allowed them to vary between the speed and accuracy conditions, and we let drift rates vary between the correct and error accumulators;

$$\theta_{\text{behavioral}} = b_{(\text{speed})}, b_{(\text{accuracy})}, t_0, v_{(c)}, v_{(e)}.$$

We estimated PMwG-FA group-level models with one to three factors, as well as PMwG-MVN. We used 5,000 iterations of burn-in, 5,000 iterations of adaptation, and 20,000 iterations of sampling. More samples were necessary compared to earlier applications since the likelihood was mostly determined by the neural component, causing slower convergence of the parameters corresponding to the behavioral component of the joint model.[5]

We constrained the Λ diagonal equal to 1 and the upper triangle fixed at 0 to ensure identifiability. We also fixed factor loadings of all ROI intercept and standard deviation parameters to 0, as there is no psychological interpretation of these loadings (i.e., they are generally considered nuisance parameters in fMRI analyses). We compared all models using IS^2 with the same settings as described in Application 1.

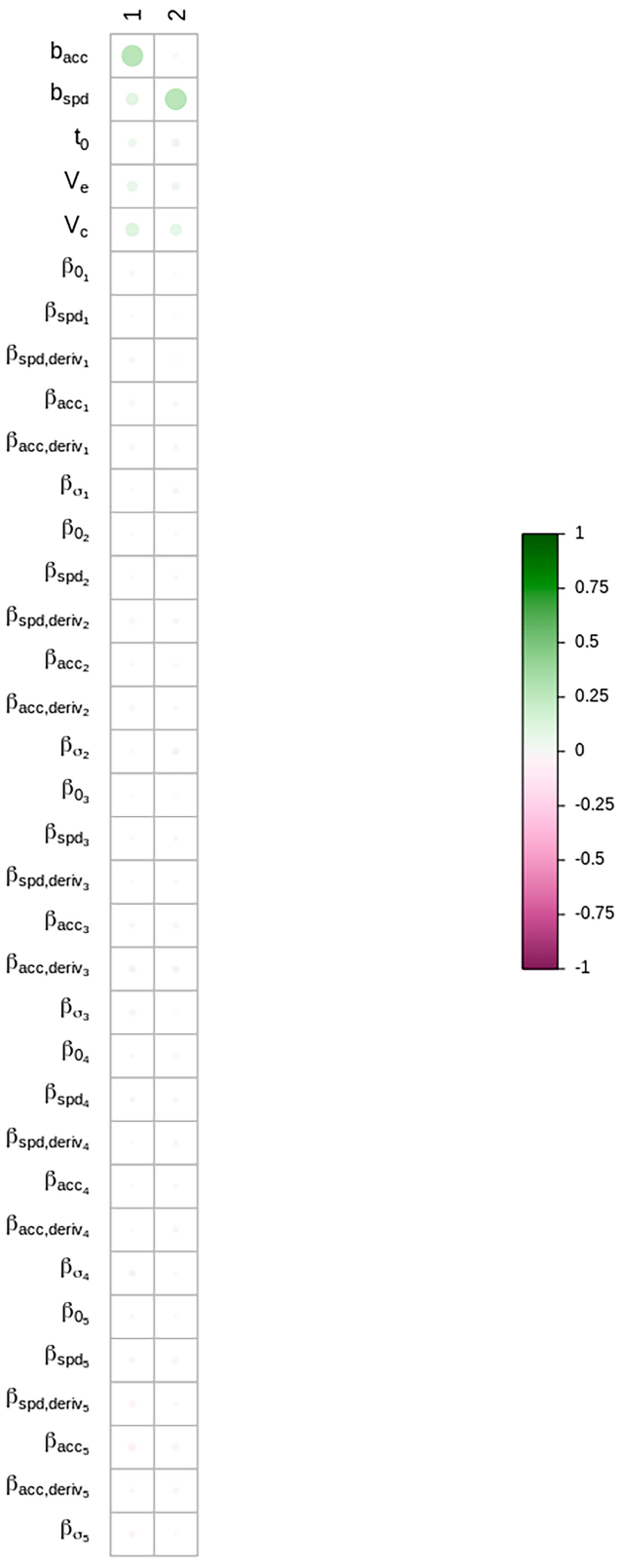
Results

The marginal log-likelihood estimate (with standard error) as calculated using IS^2 suggested that the two-factor model better accounted for the behavioral and neural data compared to the other models:

[5] We also estimated separate models for speed and accuracy conditions and compared these results to the joint model to convergence. Furthermore, we sampled three parameter chains and compared them to ensure stationarity and mixing in the sampling stage.

Figure 11

Posterior Samples of Λ_{true} for the Two Factor Joint Model of the Behavioral and Neural Data From van Maanen et al. (2011)



(*figure continue*)

one factor = −23,599.62(10.40),two factors = −23,566.58(7.58), three factors = −23,612.63(8.46),and MVN = −27,533.91(11.04). Mean factor loading estimates for the joint two-factor model can be seen in Figure 11.

The two factors were dominated by loadings for the accuracy and speed boundary respectively, suggesting that they correspond to individual differences in the strategies adopted by participants in these two conditions. There is a striking lack of any relationship between the behavioral and neural parameters, which is at odds with results reported by Turner, Wang, et al. (2017) and Kang et al. (2021) in their pooled analysis of the same data set that ignores differences between participants. A potential reason is that the small number of participants in this data set is insufficient to detect reliable brain–behavior relationships. In light of these findings, we recommend caution be exercised in using the pooling method. We also recommend that data sets with larger numbers of trials per participant and larger samples of participants be collected and analyzed with an approach like PMwG-FA so that brain–behavior relationships can be tested in a way that takes account of individual differences.

Discussion

In the current article, we propose a flexible Bayesian hierarchical framework that utilizes factor analysis to test inter-individual relationships among the parameters of different models. In this framework, the researcher can specify multiple behavioral models (e.g., for different tasks, or the same task performed in different settings) and/or neuroscience models (e.g., for different physiological measures) for each individual. Across individuals, the model parameters are reciprocally linked to a group-level factor structure, which describes the relationships between the individual-level parameter estimates. Factor analysis methods have a long tradition in psychological sciences, since the latent factors can provide a deeper layer of inference to interrelationships between sources of data or model parameters (Gorsuch, 2013; Schmiedek et al., 2007; Schmitz & Wilhelm, 2016). However, in contrast to previous model-based factor analyses, the relationships between the individual level parameters are estimated in one hierarchical Bayesian model, which reduces attenuation of the true relationships (Matzke et al., 2017).

Our work builds on previous methods that used a covariance matrix at the group level to incorporate relationships between model parameters (Gunawan et al., 2020; Turner, Forstmann, et al., 2013). This covariance approach is suited to the field of "joint modeling," where researchers are interested in the relationships between two or more different models. However, especially in models of neural activity which often examine many ROIs, the number of estimated covariances can quickly grow to be unmanageable. This increase in dimensionality leads to lower precision in the estimation of the relationships unless the number of participants is increased to compensate (Haines et al., 2020; Rouder et al., 2023). Throughout the article, we showcase the

Figure 11 (*continued*)

Note. Subscripts 1–5 correspond to the following brain regions: 1. ACC, 2. lPreSMA, 3. lSTN, 4. rPreSMA, 5. rSTN. ACC = anterior cingulate cortex; lSTN = left striatum; rSTN = right striatum; lPreSMA = left presupplementary motor area; rPreSMA = right presupplementary motor area. See the online article for the color version of this figure.

benefits of the dimensionality reduction obtained by the factor structure compared to the covariance approach with a particular focus on joint modeling applications. Furthermore, by reducing the dimensionality and providing a deeper layer on which to make inferences (i.e., latent factors), the hierarchical factor structure is not only useful for joint modeling, but more widely applicable to a variety of applications.

As shown by simulation exercises, the hierarchical factor model shows good recovery properties across various parameter, factors, and subject combinations. This holds for the factor loadings, the number of factors, and the covariance matrix which can be reconstructed through the factor structure. Additionally, by using the PMwG sampler (Gunawan et al., 2020), we maintain efficient estimation properties for subject and group-level effects through a fully hierarchical Bayesian sampling scheme.

Furthermore, we outlined methods to estimate the marginal likelihood and associated Bayes Factors (Kass & Raftery, 1995) to select among different factor-analysis models using the IS^2 algorithm (Tran et al., 2021). Throughout the applications we use the marginal likelihood estimates to showcase the benefits of the proposed factor models compared to the full covariance approach, and to determine the number of latent factors that best accommodate the data.

Through three applications of the hierarchical factor framework to previously collected data, we demonstrate its versatility for estimating joint models of different behavioral tasks, and joint models of neural and behavioral data. The joint models of Applications 1 (the same task in two different settings) and 2 (two different tasks) were both best described with a one-factor solution. Parameter loadings on this factor suggested similar conclusions to the original papers based on correlations among parameters, either estimated after the task models were fit (Forstmann et al., 2008) or simultaneously in a joint model (Wall et al., 2021). Furthermore, the fact that a one-factor solution was preferred over models with multiple factors suggested consistent individual differences across different contexts (Application 1) and different tasks (Application 2). Such conclusions exemplify the deeper layer of inference that the hierarchical factor structure provides.

The neural and behavioral joint model of Application 3 was best described with two factors that corresponded to individual differences in response caution between the speed and accuracy conditions (Turner, Wang, et al., 2017). However, we found a lack of relationships between the neural and behavioral parameters. The fact that a two-factor solution was preferred despite the small factor loadings is potential because the number of identified factors is jointly determined by the amount of information in the data (e.g., trials and subjects), and the number of relationships estimated. The larger the number of relationships estimated, albeit poorly informed, the larger the chance of needing multiple factors to describe the data. The small factors loadings in Application 3 are likely caused by the limited data from the task (only around 80 trials per subject) leading to poorly informed subject-level estimates, which leads to poor estimation at the group level. This application shows the importance of increasing subject and group-level precision, such that we avoid inflating parameter relationships.

One benefit observed across these applications is the dimensionality reduction afforded by the factor structure. The reduced parameter space compared to the covariance approach leads to more informed estimates of the relationships between the parameters, because the ratio of estimated parameters to available data points is smaller. This reduced parameter space can be particularly useful in experiments with small amounts of participants, such as in Applications 1 and 3. Additionally, the flexibility of the framework compared to the covariance approach is exemplified in Applications 2 and 3 where we were able to specify additional constraints a priori to avoid estimating relationships between nuisance parameters and the parameters of interest.

In order to promote the use of these new methods, the code for the analyses of the applications, and the simulation study results can be found at https://osf.io/pn3ca/. Furthermore, a tutorial on how to apply the code to new research can be found at https://github.com/niekstevenson/hierarchical-factor-analysis. There, we also describe how we adjusted the PMwG (and when appropriate IS^2) algorithm to improve its efficiency and to give it greater flexibility so that it can perform single-subject particle metropolis sampling (i.e., without the hierarchical Gibbs step), PMwG sampling with parameter independence assumed at the group level, and PMwG with a blocked covariance structure.

Evidently, the method shares many parallels with joint-modeling work of Turner, Wang, et al. (2017) and the follow-up work by Kang et al. (2021) that include Lasso priors on the factor loadings. However, in contrast to their approach, we do not make restrictive assumptions about the factors loading, and factor analysis is used in a more exploratory manner. Furthermore, our proposed framework allows a fully Bayesian hierarchical method of estimation, with subject-level and group-level models estimated that are reciprocally related, hence respecting the individual differences present in the data. Lastly, our framework allows the researcher to specify their own model(s) for the data of each individual and to estimate the marginal likelihood of different group-level models.

Despite the favorable properties of the hierarchical factor framework, we sound a number of notes of caution. First, it is difficult to determine the number of participants that are required to recover different numbers of parameters and factors included in a model. However, it is clear that collecting few data points per subject will lead to low individual-level certainty, which can impact group-level recovery. To test whether the appropriate number of subjects and data points per subject are obtained to identify a certain number of factors, we provide the file test-identifiability.R at https://github.com/niekstevenson/hierarchical-factor-analysis. This file contains code with which the researcher can set the number of parameters, subjects, and data points equal to their own design and test the recovery of the model with a different number of factors.

Furthermore, prior knowledge of (hierarchical) Bayesian methodology and the programming language R will significantly reduce implementation burdens. To facilitate researchers interested in applying the joint modeling framework to their own data started, we also provide code for our illustrative example (toy-example.R at https://github.com/niekstevenson/hierarchical-factor-analysis), which showcases how the joint models are set-up, common estimation checks, and how posterior inference is drawn.

In summary, we hope that the work presented here will help researchers move away from separate, post hoc, nonhierarchical and/or purely correlational approaches toward methods using generative hierarchical joint models. We believe this approach is especially applicable in neuroscience, where researchers have access to multiple modalities of data (often behavioral and neural), which allows for more refined links to be drawn between brain and behavior. Previously available estimation methods made joint modeling

statistically and computationally difficult. Fortunately, recent advances in model estimation, combined with our methods, make using joint modeling achievable for many more researchers. The methods developed here are general and so extend beyond the EAMs used in our examples, meaning this new approach is applicable to many other fields within psychology and the neurosciences.

References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., & Gramfort, A. (2014). Machine learning for neuroimaging with Scikit-learn. *Frontiers in Neuroinformatics*, *8*(1), Article 14. https://doi.org/10.3389/fninf.2014.00014

Aguilar, O., & West, M. (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business & Economic Statistics*, *18*(3), 338–357. https://doi.org/10.1080/07350015.2000.10524875

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267–281). Akademiai Kiado.

Avants, B., Epstein, C., Grossman, M., & Gee, J. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, *12*(1), 26–41. https://doi.org/10.1016/j.media.2007.06.004

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, *37*(1), 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042

Bhattacharya, A., & Dunson, D. B. (2011). Sparse Bayesian infinite factor models. *Biometrika*, *98*(2), 291–306. https://doi.org/10.1093/biomet/asr013

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*(3), 153–178. https://doi.org/10.1016/j.cogpsych.2007.12.002

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I segmentation and surface reconstruction. *NeuroImage*, *9*(2), 179–194. https://doi.org/10.1006/nimg.1998.0395

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., & Albert, M. S. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021

Donkin, C., & Brown, S. D. (2018). Response times and decision-making. In Wagenmakers, E.-J. (Ed.), *Stevens' handbook of experimental psychology and cognitive neuroscience* (Vol. 5, pp. 1–33). John Wiley & Sons.

Donkin, C., Brown, S. D., & Heathcote, A. J. (2009). The over-constraint of response time models: Rethinking the scaling problem. *Psychonomic Bulletin & Review*, *16*(6), 1129–1135. https://doi.org/10.3758/PBR.16.6.1129

Esteban, O., Blair, R., Markiewicz, C. J., Berleant, S. L., Moodie, C., Ma, F., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Sitek, K. R., Gomez, D. E. P., Lurie, D. J., Ye, Z., Salo, T., Poldrack, R. A., Valabregue, R., & Gorgolewski, K. J. (2018). *fMRIPrep* [Software]. https://doi.org/10.5281/zenodo.852659

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Fonov, V., Evans, A., McKinstry, R., Almli, C., & Collins, D. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, *47*(Suppl. 1), Article S102. https://doi.org/10.1016/S1053-8119(09)70884-5

Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., Von Cramon, D. Y., Ridderinkhof, K. R., & Wagenmakers, E. J. (2008). Striatum and pre-SMA facilitate decision-making under time pressure. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(45), 17538–17542. https://doi.org/10.1073/pnas.0805903105

Forstmann, B. U., & Wagenmakers, E. J. (2015). *An introduction to model-based cognitive neuroscience*. Springer.

Gelman, A. (2006). Multilevel (hierarchical) modeling: What it can and cannot do. *Technometrics*, *48*(3), 432–435. https://doi.org/10.1198/004017005000000661

Geweke, J., & Zhou, G. (1996). Measuring the pricing error of the arbitrage pricing theory. *The Review of Financial Studies*, *9*(2), 557–587. https://doi.org/10.1093/rfs/9.2.557

Ghosh, J., & Dunson, D. B. (2009). Default prior distributions and efficient posterior computation in Bayesian factor analysis. *Journal of Computational and Graphical Statistics*, *18*(2), 306–320. https://doi.org/10.1198/jcgs.2009.07145

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in Python. *Frontiers in Neuroinformatics*, *5*, Article 13. https://doi.org/10.3389/fninf.2011.00013

Gorgolewski, K. J., Esteban, O., Markiewicz, C. J., Ziegler, E., Ellis, D. G., Jarecka, D., Notter, M. P., Johnson, H., Burns, C., Manhães-Savio, A., Hamalainen, C., Yvernault, B., Salo, T., Goncalves, M., Jordan, K., Waskom, M., Wong, J., Modat, M., Loney, F., … Dewey, B. (2018). *Nipype* [Software]. https://doi.org/10.5281/zenodo.596855

Gorsuch, R. L. (2013). *Factor analysis*. Psychology Press.

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060

Gronau, Q. F., Heathcote, A., & Matzke, D. (2020). Computing Bayes factors for evidence-accumulation models using Warp-III bridge sampling. *Behavior Research Methods*, *52*(2), 918–937. https://doi.org/10.3758/s13428-019-01290-6

Gronau, Q. F., Sarafoglou, A., Matzke, D., Ly, A., Boehm, U., Marsman, M., Leslie, D. S., Forster, J. J., Wagenmakers, E. J., & Steingroever, H. (2017). A tutorial on bridge sampling. *Journal of Mathematical Psychology*, *81*, 80–97. https://doi.org/10.1016/j.jmp.2017.09.005

Guan, M., Vandekerckhove, J., & Lee, M. D. (2020). A cognitive modeling analysis of risk in sequential choice tasks. *Judgment and Decision Making*, *15*(5), 823–850. https://doi.org/10.1017/S1930297500007956

Gunawan, D., Carter, C., Fiebig, D. G., & Kohn, R. (2017). *Efficient Bayesian estimation for flexible panel models for multivariate outcomes: Impact of life events on mental health and excessive alcohol consumption*. arXiv preprint arXiv:1706.03953v1.

Gunawan, D., Hawkins, G. E., Tran, M. N., Kohn, R., & Brown, S. D. (2020). New estimation approaches for the hierarchical linear ballistic accumulator model. *Journal of Mathematical Psychology*, *96*, Article 102368. https://doi.org/10.1016/j.jmp.2020.102368

Haaf, J. M., & Rouder, J. N. (2019). Some do and some don't? Accounting for variability of individual difference structures. *Psychonomic Bulletin & Review*, *26*(3), 772–789. https://doi.org/10.3758/s13423-018-1522-x

Haines, N., Kvam, P. D., Irving, L., Smith, C., Beauchaine, T. P., Pitt, M. A., & Turner, B. (2020). *Theoretically informed generative models can advance the psychological and brain sciences: Lessons from the reliability paradox*. Psyarxiv. https://doi.org/10.31234/osf.io/xr7y3

Heathcote, A., Brown, S. D., & Wagenmakers, E. J. (2015). An introduction to good practices in cognitive modeling. In B. U. Forstmann & E. J. Wagenmakers (Eds.), *An introduction to model-based cognitive neuroscience*. Springer.

Huang, A., & Wand, M. P. (2013). Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Analysis*, *8*(2), 439–452. https://doi.org/10.1214/13-BA815

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1006/nimg.2002.1132

Jiang, J., & Nguyen, T. (2007). *Linear and generalized linear mixed models and their applications* (Vol. 1). Springer.

Johansen-Berg, H., Behrens, T. E. J., Robson, M. D., Drobnjak, I., Rushworth, M. F. S., Brady, J. M., Smith, S. M., Higham, D. J., & Matthews, P. M. (2004). Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(36), 13335–13340. https://doi.org/10.1073/pnas.0403743101

Johansen-Berg, H., Behrens, T. E. J., Robson, M. D., Drobnjak, I., Rushworth, M. F. S., Brady, J. M., Smith, S. M., Higham, D. J., & Matthews, P. M. (2018). Estimating across-trial variability parameters of the diffusion decision model: Expert advice and recommendations. *Journal of Mathematical Psychology*, *87*, 46–75. https://doi.org/10.1016/j.jmp.2018.09.004

Kang, I., Yi, W., & Turner, B. M. (2021). A regularization method for linking brain and behavior. *Psychological Methods*, *27*(3), 400–425. https://doi.org/10.1037/met0000387

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572

Keuken, M., Alkemade, A., Stevenson, N., Innes, R. J., & Forstmann, B. U. (2021). Structure-function similarities in deep brain stimulation targets cross-species. *Neuroscience & Biobehavioral Reviews*, *131*, 1127–1135. https://doi.org/10.1016/j.neubiorev.2021.10.029

Keuken, M. C., & Forstmann, B. U. (2015). A probabilistic atlas of the basal ganglia using 7 T MRI. *Data in Brief*, *4*, 577–582. https://doi.org/10.1016/j.dib.2015.07.028

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Chaibub Neto, E., & Keshavan, A. (2017). Mindboggling morphometry of human brains. *PLoS Computational Biology*, *13*(2), Article e1005350. https://doi.org/10.1371/journal.pcbi.1005350

Lanczos, C. (1964). Evaluation of noisy data. *Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis*, *1*(1), 76–85. https://doi.org/10.1137/0701007

Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as statistical method* (Technical Report). https://catalogue.nla.gov.au/catalog/474398

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

Lee, M. D. (2018). Bayesian methods in cognitive modeling. In Wagenmakers, E.-J. (Ed.), *The Stevens' handbook of experimental psychology and cognitive neuroscience* (Vol. 5, pp. 37–84). John Wiley & Sons.

Lee, M. D., Criss, A. H., Devezer, B., Donkin, C., Etz, A., Leite, F. P., Matzke, D., Rouder, J. N., Trueblood, J. S., White, C. N., & Vandekerckhove, J. (2019). Robust modeling in cognitive science. *Computational Brain & Behavior*, *2*(3), 141–153. https://doi.org/10.1007/s42113-019-00029-y

Lerche, V., von Krause, M., Voss, A., Frischkorn, G. T., Schubert, A. L., & Hagemann, D. (2020). Diffusion modeling and intelligence: Drift rates show both domain-general and domain-specific relations with intelligence. *Journal of Experimental Psychology: General*, *149*(12), 2207–2249. https://doi.org/10.1037/xge0000774

Lopes, H. F. (2014). Modern Bayesian factor analysis. In I. Jeliazkov & X.-S. Yang (Eds.), *Bayesian inference in the social sciences* (pp. 115–153). John Wiley & Sons. https://doi.org/10.1002/9781118771051

Lopes, H. F., & West, M. (2004). Bayesian model assessment in factor analysis. *Statistica Sinica*, *14*(1), 41–67. https://www.jstor.org/stable/24307179

Matzke, D., Ly, A., Selker, R., Weeda, W. D., Scheibehenne, B., Lee, M. D., & Wagenmakers, E. J. (2017). Bayesian inference for correlations in the presence of measurement error and estimation uncertainty. *Collabra: Psychology*, *3*(1), 25. https://doi.org/10.1525/collabra.78

Myung, I. J., & Pitt, M. A. (1997). Applying Occam's Razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin & Review*, *4*(1), 79–95. https://doi.org/10.3758/BF03210778

Palestro, J. J., Bahg, G., Sederberg, P. B., Lu, Z. L., Steyvers, M., & Turner, B. M. (2018). A tutorial on joint models of neural and behavioral measures of cognition. *Journal of Mathematical Psychology*, *84*, 20–48. https://doi.org/10.1016/j.jmp.2018.03.003

Poldrack, R. A. (2007). Region of interest analysis for fMRI. *Social Cognitive and Affective Neuroscience*, *2*(1), 67–70. https://doi.org/10.1093/scan/nsm006

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*, *84*(Suppl. C), 320–341. https://doi.org/10.1016/j.neuroimage.2013.08.048

Ratcliff, R., & Childers, R. (2015). Individual differences and fitting methods for the two-choice diffusion model of decision making. *Decision*, *2*(4), 237–279. https://doi.org/10.1037/dec0000030

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review*, *111*(1), 159–182. https://doi.org/10.1037/0033-295X.111.1.159

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, *20*(4), 260–281. https://doi.org/10.1016/j.tics.2016.01.007

Rouder, J. N., Kumar, A., & Haaf, J. M. (2023). Why many studies of individual differences with inhibition tasks may not localize correlations. *Psychonomic Bulletin & Review*. Advance online publication. https://doi.org/10.3758/s13423-023-02293-3

Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, *12*(4), 573–604. https://doi.org/10.3758/BF03196750

Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., Eickhoff, S. B., Hakonarson, H., Gur, R. C., Gur, R. E., & Wolf, D. H. (2013a). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*, *64*, 240–256. https://doi.org/10.1016/j.neuroimage.2012.08.052

Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., Eickhoff, S. B., Hakonarson, H., Gur, R. C., Gur, R. E., & Wolf, D. H. (2013b). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*, *64*(1), 240–256. https://doi.org/10.1016/j.neuroimage.2012.08.052

Scheibehenne, B., & Pachur, T. (2015). Using Bayesian hierarchical parameter estimation to assess the generalizability of cognitive models of choice. *Psychonomic Bulletin and Review*, *22*(2), 391–407. https://doi.org/10.3758/s13423-014-0684-4

Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H. M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General*, *136*(3), 414–429. https://doi.org/10.1037/0096-3445.136.3.414

Schmitz, F., & Wilhelm, O. (2016). Modeling mental speed: Decomposing response time distributions in elementary cognitive tasks and correlations with working memory capacity and fluid intelligence. *Journal of Intelligence*, *4*(4), 13. https://doi.org/10.3390/jintelligence4040013

Seber, G. A. (2009). *Multivariate observations* (Vol. 252). John Wiley & Sons.

Smith, A. F., & Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Methodological)*, *55*(1), 3–23. https://doi.org/10.1111/j.2517-6161.1993.tb01466.x

Song, X. Y., & Lee, S. Y. (2001). Bayesian estimation and test for factor analysis model with continuous and polytomous data in several populations.

British Journal of Mathematical and Statistical Psychology, *54*(2), 237–263. https://doi.org/10.1348/000711001159546

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(4), 583–639. https://doi.org/10.1111/1467-9868.00353

Steyvers, M., & Benjamin, A. S. (2019). The joint contribution of participation and performance to learning functions: Exploring the effects of age in large-scale data sets. *Behavior Research Methods*, *51*(4), 1531–1543. https://doi.org/10.3758/s13428-018-1128-2

Teller, D. (1984). Linking propositions. *Vision Research*, *24*(10), 1233–1246. https://doi.org/10.1016/0042-6989(84)90178-0

Tran, M. N., Scharth, M., Gunawan, D., Kohn, R., Brown, S. D., & Hawkins, G. E. (2021). Robustly estimating the marginal likelihood for cognitive models via importance sampling. *Behavior Research Methods*, *53*(3), 1148–1165. https://doi.org/10.3758/s13428-020-01348-w

Turner, B. M., Forstmann, B. U., Love, B. C., Palmeri, T. J., & Van Maanen, L. (2017). Approaches to analysis in model-based cognitive neuroscience. *Journal of Mathematical Psychology*, *76*, 65–79. https://doi.org/10.1016/j.jmp.2016.01.001

Turner, B. M., Forstmann, B. U., & Steyvers, M. (2019). *Joint models of neural and behavioral data*. Springer.

Turner, B. M., Forstmann, B. U., Wagenmakers, E. J., Brown, S. D., Sederberg, P. B., & Steyvers, M. (2013). A Bayesian framework for simultaneously modeling neural and behavioral data. *NeuroImage*, *72*, 193–206. https://doi.org/10.1016/j.neuroimage.2013.01.048

Turner, B. M., Sederberg, P. B., Brown, S. D., & Steyvers, M. (2013). A method for efficiently sampling from distributions with correlated dimensions. *Psychological Methods*, *18*(3), 368–384. https://doi.org/10.1037/a0032222

Turner, B. M., Wang, T., & Merkle, E. C. (2017). Factor analysis linking functions for simultaneously modeling neural and behavioral data. *NeuroImage*, *153*, 28–48. https://doi.org/10.1016/j.neuroimage.2017.03.044

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 Bias Correction. *IEEE Transactions on Medical Imaging*, *29*(6), 1310–1320. https://doi.org/10.1109/TMI.2010.2046908.

van Maanen, L., Brown, S. D., Eichele, T., Wagenmakers, E. J., Ho, T., Serences, J., & Forstmann, B. U. (2011). Neural correlates of trial-to-trial fluctuations in response caution. *Journal of Neuroscience*, *31*(48), 17488–17495. https://doi.org/10.1523/JNEUROSCI.2924-11.2011

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. https://doi.org/10.1007/s11222-016-9696-4

Wall, L., Gunawan, D., Brown, S. D., Tran, M. N., Kohn, R., & Hawkins, G. E. (2021). Identifying relationships between cognitive processes across tasks, contexts, and time. *Behavior Research Methods*, *53*(1), 78–95. https://doi.org/10.3758/s13428-020-01405-4

Weigard, A., Clark, D. A., & Sripada, C. (2021). Cognitive efficiency beats top-down control as a reliable individual difference dimension relevant to self-control. *Cognition*, *215*, Article 104818. https://doi.org/10.1016/j.cognition.2021.104818

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the drift-diffusion model in Python. *Frontiers in Neuroinformatics*, *7*, Article 14. https://doi.org/10.3389/fninf.2013.00014

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation–maximization algorithm. *IEEE Transactions on Medical Imaging*, *20*(1), 45–57. https://doi.org/10.1109/42.906424

(*Appendices follow*)

Appendix A

Further Methods

Single-Subject Estimation

Many methods of model estimation avoid complications of hierarchical sampling approaches by instead estimating only random effects (per subject). These methods generally rely on evolutionary algorithms or the SIMPLEX algorithm to estimate the parameters using maximum likelihood per subject. In an extension to PMwG, we removed the group level hierarchy and Gibbs step, so that the particle step is used for single-subject estimation. For the particle step, a mix of proposal particles is still used as in PMwG, however, rather than using the group level variance and a proposal distribution from the group level, now the prior variance and prior distribution are used. This means that the particle sampling will be slightly less efficient and rely more heavily on the prior, however, this Particle Metropolis method does represent a single-subject estimation approach, allowing users to test for shrinkage and other factors of interest related to parameter recovery outside of the hierarchical framework.

PMwG-Diagonal

Previous hierarchical methods have avoided the extreme high dimensionality problem faced by PMwG-MVN (to some extent), by assuming only a mean and variance prior for each parameter at the group level. For many models, this is a reasonable assumption, where parameters are assumed to have no covariance. As a means of dimensionality reduction, one could also make this assumption for the group level variance, where, rather than estimating the full covariance structure, only the diagonal is estimated (where the diagonal is the within parameter variance), with zero off-diagonal. The comparison between PMwG-MVN and PMwG-Diag for the LBA is shown in Gunawan et al. (2020), where PMwG-MVN has strength in that it accounts for this in the full covariance matrix. This means that the method is not useful for inference on joint modeling components, as there are no assumed relationships between parameters (or components). Regardless, the method is useful for other models where covariance is not apparent and dimensionality reduction is necessary. For this reason, we have included a code for PMwG-Diag within the online code.

To estimate the diagonal-only covariance matrix, the changes to the method of Gunawan et al. (2020) occur within the Gibbs step. In the original version of the sampler, the Gibbs step draws samples from a multivariate normal distribution with mean Θ_{μ} and variance Θ_{Σ}, where the prior used is taken from Huang and Wand (2013). For the diagonal-only version, only the variance is estimated and not the covariances. The required prior structure is also shown in Huang and Wand (2013), where the prior on the variance is a mixture of inverse-Gamma distributions (as opposed to a mixture of inverse-Wishart and inverse-Gamma distributions). Consequently, the Gibbs step is simplified and dimensions reduced (Figure A1).

Blocked Covariance Structure

Similar to only estimating the diagonal of the covariance matrix, blocked covariance structures assume that a limited number of components of the covariance matrix do not need to be estimated in order to produce the full covariance matrix. Here, only specified "blocks" of the matrix are estimated, as exemplified in Figure A2. For example, a joint model of two tasks could make the assumption of within-task variance, but no between-task variance. From this assumption, we get;

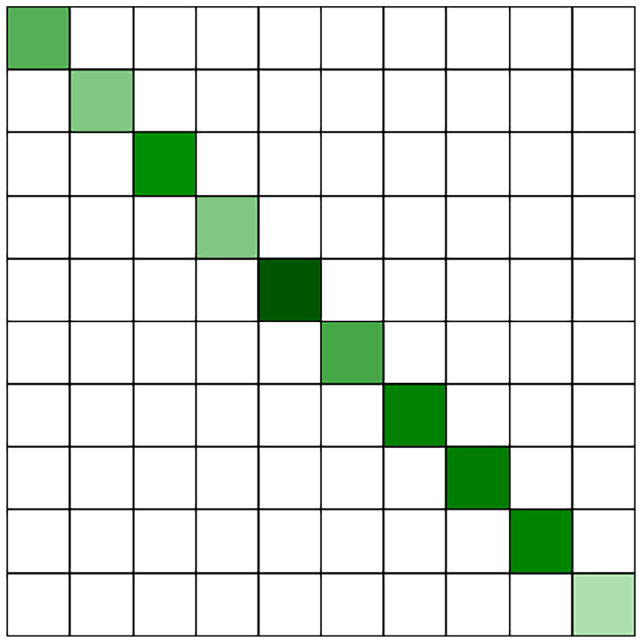
$$\Omega = \begin{bmatrix} \Omega_{T1} & 0 \\ 0 & \Omega_{T2} \end{bmatrix},$$

where the lower (and upper) rectangles are assumed zero. By setting the covariance matrix up in this way, all covariances are restricted to the diagonal blocks. This method is also useful to test specific hypotheses or could be combined with other methods (such as PMwG-Diag, PMwG-MVN, or PMwG-FA) to account for "nuisance" parameters that have no bearing on the covariance matrix. The prior on the blocked covariance matrix is simple, with a prior specified for each multivariate normal included in the structure (or the relevant prior given the group-level model of that component).

For a joint modeling approach, a blocked covariance matrix where the blocks are the two components is not useful for assessing the between component relationships, as the off-diagonal rectangles are the result that provides the greatest detail to answer joint modeling questions. However, the joint model could be parameterized such that all x parameters from Components A and B are grouped in the upper diagonal rectangle, and all y parameters of Components A and B are grouped in the lower diagonal rectangle. For example, in

Figure A1

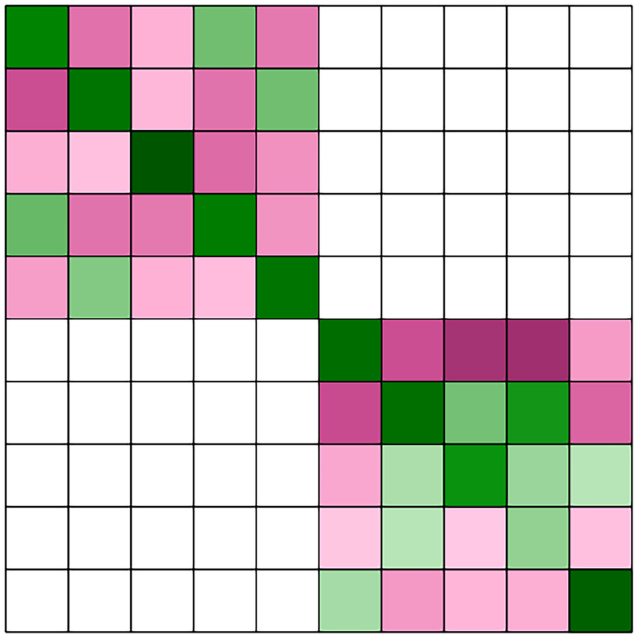
Graphical Example of the Covariance Matrix of the PMwG-Diag Approach



Note. PMwG-Diag = Particle Metropolis within Gibbs-Diagonal. See the online article for the color version of this figure.

(*Appendices continue*)

Figure A2
Graphical Example of the Covariance Matrix of A Blocked PMwG Approach



Note. PMwG = Particle Metropolis within Gibbs. See the online article for the color version of this figure.

a joint behavioral model of two environments (such as Application 1), we could block all of the threshold parameters together, and all of the drift rate parameters together, provided we had the assumption of no relationships between these parameters. This blocked covariance approach is flexible in that manner, however, does rely on specific hypotheses and assumptions made by the researcher.

Appendix B

Conjugate Priors and Gibbs Steps

Here we outline the Gibbs steps for $N_{1\ldots n}$ subjects, $P_{1\ldots p}$ parameters and $K_{1\ldots k}$ factors. The factor structure can be described using:

1. Factor scores matrix H, with dimensions $N \times K$ and individual row vectors $\eta_1 \ldots \eta_n$.
2. Factor loading matrix Λ with dimensions $P \times K$ and individual row vectors $\lambda_1 \ldots \lambda_p$
3. Diagonal residual matrix Σ with dimensions $P \times P$ and individual diagonal elements $\sigma_1^2 \ldots \sigma_p^2$
4. Diagonal factor variance covariance matrix Ψ with dimensions $K \times K$ and individual diagonal elements $\psi_1 \ldots \psi_k$
5. Mean vector μ, with individual elements $\mu_{\ 1} \ldots \mu_{\ p}$

The conditional posterior distributions used in Gibbs sampling as derived in Ghosh and Dunson (2009) are:

$$
\begin{aligned}
&\pi(\lambda_p \mid H, \Psi, \Sigma, \mu, y) \sim \mathcal{N}_p((\Sigma_{0\lambda}^{-1} + \sigma_j \eta_p^{\mathrm{T}} \eta_p)^{-1} (\Sigma_{0\lambda}^{-1} \mu_{0\lambda} + \sigma_j \eta_p^{\mathrm{T}} (Y_p - \mu_p), (\Sigma_{0\lambda}^{-1} + \sigma_j \eta_p^{\mathrm{T}} \eta_p)^{-1}),\\
&\pi(\eta_n \mid \lambda, \Psi, \Sigma, \mu, y) \sim \mathcal{N}_p((\Psi^{-1} + \Lambda^{\mathrm{T}} \Sigma^{-1} \Lambda)^{-1} (\Lambda^{\mathrm{T}} \Sigma^{-1} (y_i - \mu)), (\Psi^{-1} + \Lambda^{\mathrm{T}} \Sigma^{-1} \Lambda)^{-1}),\\
&\pi(\sigma_p^{-2} \mid H, \lambda, \Psi, \mu, y) \sim G\left(a_l + \frac{n}{2}, b_l + \frac{1}{2} \sum_{n=1}^{N} (y_{np} - \mu_p - \eta_{np} \lambda_p)^2\right),\\
&\pi(\psi_k^{-1} \mid H, \lambda, \Sigma, \mu, y) \sim G\left(c_k + \frac{n}{2}, d_k + \frac{1}{2} \sum_{n=1}^{N} \eta_{nk}^2\right),\\
&\pi(\mu_p \mid H, \lambda, \Sigma, \Psi, y) \sim \mathcal{N}_p((N\Sigma^{-1} + \Sigma_{0\mu})^{-1} (\Sigma^{-1} (Y - H\Lambda) + \Sigma_{0\mu}^{-1} \mu_{0\mu}), (N\Sigma^{-1} + \Sigma_{0\mu})^{-1}).
\end{aligned}
\tag{A1}
$$

For more details, see the parameter-expanded Gibbs sampler from Ghosh and Dunson (2009).

(*Appendices continue*)

Appendix C

fMRIprep Boilerplate

fMRIprep Boilerplate uses fMRIPrep 20.2.6 (Esteban, Blair, et al., 2018; Esteban et al., 2019; RRID:SCR_016216), which is based on Nipype 1.7.0 (K. Gorgolewski et al., 2011; K. J. Gorgolewski et al., 2018; RRID:SCR_002502).

Anatomical Data Preprocessing

A total of one T1-weighted (T1w) image was found within the input BIDS data set. The T1w image was corrected for intensity nonuniformity with `N4BiasFieldCorrection` (Tustison et al., 2010), distributed with ANTs 2.3.3 (Avants et al., 2008, RRID: SCR_004757), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a Nipype implementation of the `antsBrainExtraction.sh` workflow (from ANTs), using OASIS30ANTs as the target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white matter (WM) and gray matter (GM) was performed on the brain-extracted T1w using `fast` (FSL 5.0.9, RRID:SCR_002823, Zhang et al., 2001). Brain surfaces were reconstructed using `recon-all` (FreeSurfer 6.0.1, RRID:SCR_001847, Dale et al., 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical GM of Mindboggle (RRID:SCR_002438, Klein et al., 2017). Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with `antsRegistration` (ANTs 2.3.3), using brain-extracted versions of both T1w-reference and the T1w template. The following template was selected for spatial normalization: ICBM 152 Nonlinear Asymmetrical template version 2009c [Fonov et al., 2009, RRID: SCR_008796; TemplateFlow ID: MNI152NLin2009cAsym],

Functional Data Preprocessing

For each of the two BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. Susceptibility distortion correction was omitted. The BOLD reference was then coregistered to the T1w-reference using `bbregister` (FreeSurfer) which implements boundary-based registration (Greve & Fischl, 2009). Coregistration was configured with six degrees of freedom. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using `mcflirt` (FSL 5.0.9, Jenkinson et al., 2002). The BOLD time series (including slice-timing correction when applied) were resampled onto their original, native space by applying the transforms to correct for head motion. These resampled BOLD time series will be referred to as preprocessed BOLD in original space, or just preprocessed BOLD. The BOLD time series were resampled into standard space, generating a preprocessed BOLD run in MNI152NLin2009cAsym space. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. Several confounding time series were calculated based on the preprocessed BOLD: framewise displacement (FD), derivative of root mean square error over voxels (DVARS) and three region-wise global signals. FD was computed using two formulations following Power (absolute sum of relative motions; Power et al. (2014)) and Jenkinson (relative root mean square displacement between affines; Jenkinson et al. (2002)). FD and DVARS are calculated for each functional run, both using their implementations in Nipype (following the definitions by Power et al., 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (CompCor, Behzadi et al., 2007). Principal components are estimated after high-pass filtering the preprocessed BOLD time series (using a discrete cosine filter with 128s cutoff) for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM, and combined CSF + WM) are generated in anatomical space. The implementation differs from that of Behzadi et al. in that instead of eroding the masks by 2 pixels on BOLD space, the aCompCor masks subtract a mask of pixels that likely contain a volume fraction of GM. This mask is obtained by dilating a GM mask extracted from the FreeSurfer's aseg segmentation, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the k components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50% of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head-motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al., 2013b). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardized DVARS were annotated as motion outliers. All resamplings can be performed with a single interpolation step by composing all the pertinent transformations (i.e., head-motion transform matrices, susceptibility distortion correction when available, and coregistrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using `antsApplyTransforms` (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964). Nongridded (surface) resamplings were performed using `mri_vol2surf` (FreeSurfer).

Many internal operations of fMRIPrep use Nilearn 0.6.2 (Abraham et al., 2014, RRID:SCR_001362), mostly within the functional processing workflow. For more details of the pipeline, see the section

(*Appendices continue*)

corresponding to workflows in fMRIPrep's documentation (https://fmriprep.readthedocs.io/en/latest/workflows.html).

Copyright Waiver

The above boilerplate text was automatically generated by fMRIPrep with the express intention that users should copy and paste this text into their manuscripts unchanged. It is released under the CC0 license (https://creativecommons.org/publicdomain/zero/1.0/).

Received July 31, 2022
Revision received April 11, 2023
Accepted September 28, 2023 ■