Hierarchical Factor Models For Analysis in Experimental Designs

Jeffrey N. Rouder[1], Mahbod Mehrvarz[1], & Niek Stevenson[2]

[1] University of California, Irvine

[2] University of Amsterdam

Author Note

Correspondence concerning this article should be addressed to Jeffrey N. Rouder, Department of Cognitive Science, University of California, Irvine, CA, 92697. E-mail: jrouder@uci.edu

Abstract

It is challenging to study individual differences in experimental tasks because performance on these tasks is highly noise prone. This excessive noise if unaccounted results in attenuation of correlations, overstated confidence, and increased residuals in factor models. One advantage in using experimental tasks is that they are comprised of many repeated trials allowing for a separate modeling of trial noise and individual variation. Here, we develop a set of hierarchical factor models for individual tasks. We show that these models disattenuate correlations, provide for more clear assessments of factor structures, and have bona-fide measures of uncertainty for observed and latent variables. Attention is paid to computational issues including mixing and rotational ambiguity, and code for analysis in R using JAGS is available. The methods are illustrated with analysis of two individual-difference batteries in visual illusions and cognitive control.

Keywords: Factor Models, Hierarchical Models, Cognitive Control, Illusions, Mixed Models, Bayesian Analysis

Hierarchical Factor Models For Analysis in Experimental Designs

In his presidential address to the American Psychological Association, Lee Cronbach proposed a merger of sorts between experimental psychology and correlational psychology (Cronbach, 1957). Indeed, in many ways we have seen this merger as it is now popular to study individual differences in cognitive experimental tasks. By understanding how individuals' performance covaries across these tasks, it is perhaps possible to uncover an underlying latent structure of cognitive processing. A classic example in cognitive control comes from Miyake et al. (2000), who used latent-variable models to decompose individual differences in cognitive-control tasks into the three factors of inhibition, shifting, and updating. In today's psychological science, it is common to see latent-variable analyses of tasks that are comprised of many repeated trials. Examples come from fields as diverse as development (e.g., Snyder, Friedman, & Hankin, 2021), brain imaging (e.g., Bolt, Nomi, Yeo, & Uddin, 2017), aging (Kim-Spoon et al., 2021), and clinical pathology (Green, Hellemann, Horan, Lee, & Wynn, 2012).

Here is the usual data-analytic pipeline: **First**, individuals complete a battery of tasks. In the study of executive function, examples of tasks include the Stroop task (Stroop, 1935), the flanker task (Eriksen & Eriksen, 1974), and the antisaccade task (Kane, Bleckley, Conway, & Engle, 2001) among many others. These experimental tasks are comprised of many repeated trials per condition. **Second**, observations from repeated trials are aggregated into a single score per individual per task. The usual way of computing the score is to take either a sample mean or a sample contrast. For example, in a Stroop task, the score is the mean difference in response time between conflict (incongruent) and non-conflict (congruent) conditions. We will call this the *take-the-mean* approach, and it is prevalent, perhaps ubiquitous, in application. **Third**, The matrix of scores per individual across tasks serves as input for computing the correlation among the tasks. **Fourth,** Task covariation is decomposed into latent variables in a

structural-equation model (Bollen, 1989; Skrondal & Rabe-Hesketh, 2004). It is the relations among these latent variables that hopefully reveal the underlying structure of cognitive processes. The usual pipeline with take-the-mean scores is shown in Figure 1.

Attenuation of Correlations, Overconfidence in Bad Ansewrs

Although this usual approach is common, the results with experimental tasks have been beset with problems. First, experimental tasks are unreliable as measures of individual differences (Enkavi et al., 2019; Franco-Martínez, Vicente-Conesa, Shanks, & Vadillo, 2024; Hedge, Powell, & Sumner, 2018; Karr et al., 2018; Vadillo, Malejka, Lee, Dienes, & Shanks, 2022). Values of split-half reliability are often below .5 in value. Second, several tasks that purportedly measure the same construct do not correlate well. For example, the correlation between flanker and Stroop effects in large studies is often near .1 and rarely greater than .25 (Enkavi et al., 2019; Rey-Mermet, Gade, & Oberauer, 2018; Rouder, Kumar, & Haaf, 2023). Third, latent-variable decomposition in cognitive-control domains seems unreplicable. This lack of replicability is showcased by Karr et al. (2018) who found that latent-variable analysis with simulated data infrequently recovered the generating model when the sample sizes and parameter values used in the simulations came from extant studies.

These problems in practice reflect a statistical flaw in the usual approach with take-the-mean approach. The resulting estimates of correlation are statistically inconsistent. Figure 2 shows why. Figure 2A shows a scatter plot for 200 hypothetical individuals on two tasks. Plotted are *true scores.* True scores are the ideal and would be obtained if we had an unlimited number of trials per individual per task. There is a healthy correlation of .5 in value. Yet, we observe data from finite trials, and because of this, the sample scores are perturbed from the true scores. Figure 2B shows the scatter among the sample scores. The trial noise perturbs these scores in all directions, and the result is the well-known attenuation of correlation by measurement error (Spearman, 1904b).

One of the unappreciated difficulties with trial noise is that the resulting confidence intervals (CIs) on correlation coefficients are miscalibrated. Figure 2C shows CIs computed the conventional way with Fisher's z-transform method (Fisher, 1921). Here, we ran a small simulation where data were generated and analyzed several times. The first third of bars, colored red in Figure 2C, show the CIs from 100 runs for 20 individuals per run (ordered by sample correlation value). The width of these CIs reflect the number of individuals but not the amount of trial noise or the number of trials. For 20 individuals, they are quite wide, but even so, the coverage of the true correlation is not near the nominal .95 value. As the number of individuals is increases (blue and green bars), the CIs narrow and the coverage becomes even worse. Let's not gloss this over—coverage actually decreases with an increasing number of individuals! The analyst becomes more sure of a bad answer.

The main difficulty with these CIs is that they do not reflect the influence of trial noise. They are the CIs that reflect uncertainty from a finite number of individuals as if each had run an unlimited number of trials. The difficulties with these CIs is highlighted by comparing them to a more appropriate hierarchical treatment. A hierarchical linear model with separate variance components for trial noise and individuals was applied to the trial-level data underlying Figure 2B (we specify this model subsequently as the unconstrained-variance model). The resulting posterior on the correlation across two tasks is shown in Figure 2D. The true value of .5 is located in the middle of the posterior indicating that the model effectively disattenuates the correlation. The posterior, however, is wide. It is much wider than the corresponding CI (see the horizontal error bar) showing the needed inflation from trial noise. We have shown previously that these credible intervals have reasonable frequentist coverage properties (Mehrvarz & Rouder, 2025; Rouder et al., 2023)—95% credible intervals cover the true value about 95% of the time in realistic simulations.

Underestimating The Salience of Factors

These difficulties in estimating correlations extend to factor analyses of experimental-task data. The main problem is that trial noise inflates residual variation resulting in less clear and less salient factor structures. We illustrate this problem with the following synthetic example that will be used throughout. Consider a battery of eight experimental tasks. Each task has two conditions, congruent and incongruent, with the later requiring more executive control than the former. The difference between the two conditions forms the target of inquiry. Each of 200 individual performs 100 trials in each condition, and the total number of observations is 320000 ($200 \times 8 \times 2 \times 100$). With the take-the-mean approach, this total is reduced to a matrix that is 200 individual by 8 tasks or 1600 scores.

Suppose performance on the task is mediated by two factors. Table 1 shows ground-truth loadings for this case. Factor 1 loads most highly on Task 1, second-most highly on Task 2 and decreases subsequently. Factor 2 loads most highly on Task 8, second most highly on Task 7, and decreases subsequently. Factor models are a decomposition of variance into common factor and unique residual components (Bollen, 1989), and this decomposition is depicted in Figure 3. The top row, labeled "Population," shows the case for the ground truth at the population level. The total variance across tasks is shown on the left, and the influences of the factors are shown on the right. Because Factor 1 loads most highly on Task 1 and 2, the contribution to variance of these tasks is greatest. Factor models include a residual variance term, which is variability unique to a task. This residual is a diagonal matrix. It is helpful to track the contribution of each factor and the residual to the total, and a suitable measure is the trace of each matrix (the trace is the sum of the diagonal elements). We designed this example so that Factor 1, Factor 2, and the residual components each account for about 1/3rd of the total variance.

The second row shows the effect of sampling individuals from the population. The

variance matrix is that of true scores and is analogous to the scatter in Figure 2A. The factor decomposition of this matrix is shown across the second row. There is some distortion of the factor loadings as there is only a finite sample of 200 individuals. Importantly, the influence of the residual is not amplified. In fact, each component contributes about 1/3 of the variances as the fluctuations are not systematic. The third row shows what happens with observed data. The inputs are the trial-by-trial observations, and scores are computed with take-the-mean aggregation. Shown is the empirical variance matrices of the scores, and, as before, the effect is a notable attenuation of off-diagonal terms. This attenuation occurs because trial noise inflates the residual. The factor decomposition of this take-the-mean variance matrix is shown. The factors each account for about 18.5% of the variance while the residual accounts for 63% making the factors appear less salient and theoretically less important than they truly are.

The goal of this paper is to provide accurate factor models for experimental task data. The last row of Figure 2 foreshadows how we meet this goal. It is the results of the hierarchical factor model that is the main topic of this report. As can be seen, the model is highly successful—the salience of the factors is restored in the hierarchical model analysis. Each factor accounts for about 1/3rd of the variance, and the residual accounts for the remaining 1/3rd. It is this success that should change how researchers perform factor analysis with experimental data.

Overview of The Hierarchical Solution

Although researchers often incorporate tasks with repeated trials into their latent-variable analysis, they rarely account for trial noise. As a result, latent patterns are appear both less salient and more stable than they are. The solution to this problem is conceptually straightforward—trial noise should be modeled simultaneously with individual differences in a hierarchical setup. At the first level of the hierarchy—*the data level*—observables are modeled as a function of individual true scores in tasks and trial

noise. At the second level–*the latent level*—these individual true scores are modeled with latent-variable models. In this paper we combine linear data models with exploratory and confirmatory factor models. Our goal is to make this development convenient for routine use in psychological science. We develop in the Bayesian framework as it is computationally convenient for hierarchical models (Jackman, 2009; Lee, 2007) and associated tools such as Stan (Carpenter et al., 2017) and JAGS (Plummer, 2003) are powerful and user friendly.

Here is how to differentiate this hierarchical use of factor models from the more conventional use. In the more usual use, the input is a matrix of observed scores. These are manifest variables, and we can refer to the model here as a *manifest factor model.* In the current development the input to the factor models are not observables, they are themselves latent variables that must be modeled. We term these models as *hierarchical factor models* as they are hierarchical in the sense of Raudenbush and Bryk (2002). Unfortunately, the term *hierarchical factor models* has a different meaning in an older factor-analysis literature. It refers to placing latent factors themselves in a hierarchy (Schmid & Leiman, 1957; Tucker, 1940) rather than using a hierarchical structure to account for measurement error. In the modern literature, the older usage of hierarchical factors would fall under structural-equation modeling. Hence, we use the term *hierarchical factor models* in the more modern sense of Raudenbush and Bryk (2002) and not in the older sense of Tucker (1940). In the experimental-psychology context, they refer to models that explicitly account for trial noise along with factor structures on individual variation.

Factor models have a long and storied history in psychology (Caspi et al., 2014; McCrae & Costa Jr, 1997; Spearman, 1904a; Thurstone, 1938), and Bayesian analysis with factor models is well developed (Ando, 2009; Bhattacharya & Dunson, 2011; Lopes & West, 2004; Merkle, Fitzsimmons, Uanhoro, & Goodrich, 2021). The usual context is that factor models are manifest in that they are applied to a matrix of observed scores. A few authors have developed them in the hierarchical context promoted here. In many of the previous applications, the factor model serves as a regularization devise to help estimate

either individual scores or the variance matrix among these scores (Kang, Yi, & Turner, 2022; Mehrvarz & Rouder, 2025). Our development is most similar to Stevenson, Innes, Gronau, Miletić, et al. (2024) and Stevenson.etal.2024b inasmuch as the target of analysis is the estimation of a factor structure in a hierarchical setting.

As prologue, the hierarchical factor model has the following advantages that should promote their adoption:

1. **Accurate Estimates:** In Figures 2 and 3 we demonstrated that non-hierarchical methods dramatically underestimate correlations and the salience of factors. In hierarchical factor models, model-based estimates are accurate. The salience of the factor structure is not understated and resulting correlations are not attenuated in value. The disattenuation is similar in spirit to Spearman's disattenuation (Spearman, 1904b) with two notable advantages: First, the disattenuation is more general in that it applies to the salience of factors as well as correlations; second, it is more interpretable in that correlations and proportion-of-variance measures are constrained to the appropriate intervals.

2. **Calibrated Uncertainty:** In Figure 2 we showed that confidence intervals are miscalibrated—in fact too narrow—because they fail to account for trial noise. In the hierarchical factor model, uncertainty reflects the number of trials, the degree of trial noise relative to true variation across individuals, and the number of individuals (Rouder & Mehrvarz, 2024). When correlations are disattenuated for trial noise, there is uncertainty in the disattenuation process that must be reflected in the uncertainty about correlations and factor loadings. The hierarchical factor model provides an appropriate approach for understanding the limits of the resolution of the data for inference about correlations and the associated factor structure. This understanding assuredly leads to better science.

3. **No Heywood Cases.** Heywood cases are a well-known problem in factor analysis

where the resulting variances may be negative and resulting correlations may be out of the [-1,1] range. The Bayesian approach adopted here with proper priors is not prone to Heywood cases (Congdon, 2003; Lopes & West, 2004).

In the next section, we provide a quick review of the conventional orthogonal factor model on manifest variables. Following this brief review, we present the hierarchical factor model. In the remaining sections we provide a series of applications that highlight the usefulness, performance, and the unique insights afforded by hierarchical factor models.

The Manifest Factor-Model Specification

Factor models form a class of latent variable models that have proved to be fruitful in psychological science. We assume readers have a passing familiarity with the aims, uses, and interpretations of factor models. In the following, we review the conventional factor model for manifest or observed data. Our focus here is on issues pertinent to Bayesian analysis and hierarchical extensions. Readers seeking a systematic development are referred to Skrondal and Rabe-Hesketh (2004) and Gana and Broc (2019) for theoretical and applied treatments, respectively.

The input to a manifest factor model is a matrix of observations. The rows of the matrix denote individuals; the columns denote measures (or tasks in psychology batteries); the entries, Y_{ij} are the score for the ith individual, $i = 1, \ldots, I$ on the jth measure, $j = 1, \ldots, J$. In the previous example, there were $I = 200$ individuals and $J = 8$ tasks. The inputted matrix to the manifest model had 1600 observations.

Factor models describe a decomposition of the matrix of scores into a reduced number of factors, M, where $M < J$. An intuitive but incomplete expression of the model is:

$$Y_{ij} = \mu_j + \sum_{m=1}^{M} \lambda_{jm}\eta_{im} + \epsilon_{ij}.$$

Here, μ_j is a population mean score on the jth measure, λ_{jm} is the *factor loading* for the jth measure on the mth factor, η_{im} is the *factor score* for the ith individual on the mth factor, and $\epsilon_{ij} \sim \mathrm{N}(0, \delta_j^2)$ is normally-distributed, zero-centered, measure-specific residual error not accounted for by the factor structure. Each person may be described by the vector of factor scores: $\boldsymbol{\eta}_i = \eta_{i1}, \ldots, \eta_{iM}$. The scores are usually z-scores, so a value of $\eta_{im} = -2$ is a particularly low score. More importantly, these values are treated as random effect, e.g., $\eta_{im} \sim \mathrm{N}(0, 1)$. The more complete specification is:

$$Y_{ij} \mid \eta_{i1}, \ldots, \eta_{iM} = \mu_j + \sum_{m=1}^{M} \lambda_{jm} \eta_{im} + \epsilon_{ij}$$

$$\eta_{im} \sim \mathrm{N}(0, 1)$$

It is convenient to recast the model with matrix notation. Let $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{iJ})'$ be a column vector of scores for the ith individual. Then, the model may be rewritten as

$$\boldsymbol{Y}_i \mid \boldsymbol{\eta}_i \quad \sim \quad \mathrm{N}_J(\boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\eta}_i,\ D(\boldsymbol{\delta}^2)), \tag{1}$$

$$\boldsymbol{\eta}_i \quad \sim \quad \mathrm{N}_M(\boldsymbol{0}, \mathbf{I}).$$

Model parameters are as follows: $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_J)'$ is a vector of J true means, $\boldsymbol{\Lambda}$ is the $J \times M$ matrix of factor loadings, $\boldsymbol{\eta}_i = (\eta_{i1}, \ldots, \eta_{iM})'$ is a vector of M factor scores for the ith individual, and $D(\boldsymbol{\delta}^2)$ is a $J \times J$ diagonal matrix with the diagonal given by $(\delta_1^2, \ldots, \delta_J^2)$.

Equation 1 is conditional on factor scores. Whereas factor scores are fully specified, it is possible to marginalize with respect to them. The marginal form of the model is

$$\boldsymbol{Y}_i \sim \mathrm{N}_J(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{D}(\boldsymbol{\delta}^2). \tag{2}$$

The conditional and marginal forms are identical in content. That said, they have different computational properties. In the conditional form, the scores are conditionally independent and the residual variance has no covariance terms. Covariance from the factor structure enter through the location (mean) with additions of $\mathbf{\Lambda}\boldsymbol{\eta}_i$. In the marginal form, covariance from the factor structure enters the variance rather than the mean. Because the variance now captures the complete factor structure, it is useful for understanding model behavior.

Figure 3 shows the decomposition for two factors. The key is that the crossproduct $\mathbf{\Lambda}\mathbf{\Lambda}'$ may be written as the sum of the influence of each factor on variance. Let $\boldsymbol{\lambda}_m = (\lambda_{1m}, \ldots, \lambda_{Jm})'$ be the loadings of all tasks onto the mth factor. Then, the influence on variance of this factor is $\boldsymbol{\lambda}_m\boldsymbol{\lambda}'_m$. For example, the second column in Figure 3 shows $\boldsymbol{\lambda}_1\boldsymbol{\lambda}'_1$, the influence of the first factor. The third column in Figure 3 shows $\boldsymbol{\lambda}_2\boldsymbol{\lambda}'_2$, the influence of the second factor. The crossproduct $\mathbf{\Lambda}\mathbf{\Lambda}'$ is simply the sum of these influences from all factors, that is, $\mathbf{\Lambda}\mathbf{\Lambda}' = \sum_m \boldsymbol{\lambda}_m\boldsymbol{\lambda}'_m$. We find breaking up the crossproduct this way to be helpful in visualizing factor decompositions and, to our limited knowledge, the representation is novel in psychology.

Hierarchical Factor Model Specification

The main goal is to account for multiple trials and trial noise in experimental settings. We develop a hierarchical model with a data level and a latent level, where the latent level is a factor model. We take the development in two steps. Consider a battery composed of J experimental tasks. Each task has a control and treatment condition, and the main measure of interest is the difference between them. For example, in the Stroop task, incongruent stimuli serve as treatment and congruent ones serve as control. Each observation in this battery is denoted by $Y_{ijk\ell}$ where $i = 1, \ldots, I$ indexes individuals, $j = 1, \ldots, J$ indexes tasks, $k = 1, 2$ indexes conditions, and $\ell = 1, \ldots, L_{ijk}$ indexes the replicate trials in the conditions.

The input to the model is the full array of data. For example, if there are $I = 200$ individuals, $J = 8$ tasks, and $L_{ijk} = 100$ replicates per individual per task per condition, then the total number of observations is $N = I \times J \times 2 \times L = 320000$ observations. It is instructive to compare this input to that of the manifest model. To use the manifest model, we take means over replicates and conditions to arrive at a single score per person per task. For $I = 200$ and $J = 8$, there are 1600 scores that serve as input. In contrast, the hierarchical version is a much larger models encompassing a great many more observations.

The first step is the data model. Here is a simple data model for treatment-control tasks:

$$Y_{ijk\ell} \mid \theta_{ij} \sim \mathrm{N}(\alpha_{ij} + x_k\theta_{ij}, \ \tau_j^2)$$

In this setup, α_{ij} is the overall performance of the ith individual on the jth task, $x_k = -1/2, \ 1/2$ is a condition indicator, and θ_{ij} is the *effect* of the ith individual on the jth task. Parameter τ_j^2 is the trial noise on the jth task. The target of inquiry are the effects, θ_{ij}, which are modeled as random effects subsequently. Because these are random and the target of inquiry, we include them as a conditional in the above equaiton.

The data model may be generalized beyond a treatment-control task as follows. Let $\boldsymbol{Y}_{ij}$ be a vector of n_{ij} observations for the ith person on the jth task. Then,

$$\boldsymbol{Y}_{ij} \sim \mathrm{N}_{n_{ij}}(\boldsymbol{X}_{0ij}\boldsymbol{\alpha}_{ij} + \boldsymbol{X}_{1ij}\theta_{ij}, \boldsymbol{I}\tau_j^2), \tag{3}$$

where $\boldsymbol{X}_{0ij}$ is a $n_{ij} \times K$ design matrix, $\boldsymbol{\alpha}_{ij}$ is a vector of auxiliary parameters of length P, $\boldsymbol{X}_{1ij}$ is a design vector of length n_{ij}, $\boldsymbol{I}$ is the $n_{ij} \times n_{ij}$ identity matrix, and τ_j^2 describes the trial noise for the jth task. For the above example with an overall mean and difference score, $\boldsymbol{X}_{0ij}$ would be a vector of 1's; $\boldsymbol{\alpha}_{ij}$ would be a single parameter α_{ij}, $\boldsymbol{X}_{1ij}$ would be a vector with -1/2 and 1/2 entries for congruent and incongruent trials, respectively.

The data model is an ordinary linear model with the constraint that trial noise is

constant across people within a task. Although we develop the hierarchical model with this simple data model, it may be possible to use alternative data models depending on context. For example, if observations are dichotomous, then it may be desirable to use a logit-linked data model. Likewise, if observations are response times, it may be desirable to use a skewed distribution. It is conceptually straightforward to adopt these data models in the Bayesian framework (Kang et al., 2022; Rouder, 2025; Stevenson, Innes, Gronau, Miletić, et al., 2024; Stevenson, Innes, Gronau, Miletic, et al., 2024). In this publication, we develop with a linear data model; developments outside this model remain for future studies.

The next step is placing a factor model on target effects θ_{ij}. Let $\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{iJ})'$ be a column vector of effects:

$$\boldsymbol{\theta}_i \sim \mathrm{N}_J(\boldsymbol{\nu}\,\boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + D(\boldsymbol{\delta}^2). \tag{4}$$

The hierarchical factor model is comprised of data model (3) and factor model (4).

Analysis

Several researchers have proposed and developed Bayesian analysis of the manifest factor models (Ando, 2009; Bhattacharya & Dunson, 2011). We follow Merkle et al. (2021) who provide development on two separate tracks. One of these is based on the conditional form of the model with explicit parameters for factor scores and factor loadings (see Eq. 1). The other is based on the marginal form (see Eq. 2).

The conditional form is perhaps the simplest. The advantage of this approach is that conditional posterior distributions of factor scores and factor loadings are independent normal distributions, which are exceedingly quick to sample in Markov chain Monte Carlo (MCMC) analysis. The disadvantage may be poor mixing in some contexts where posterior chains are highly correlated and slow to converge (Ghosh & Dunson, 2009). This poor mixing is problematic when the chain relies on sequential sampling of each parameter in

Gibbs steps. The marginal form is recommended by Merkle et al. (2021) to avoid poor mixing in the conditional form. The drawback of using the marginal form is speed—it takes substantially longer to sample from correlated multivariate normal distributions. We have developed both forms. We find that for the hierarchical factor models presented here that the conditional form mixes well in JAGS and Stan and runs substantially faster than the marginal form.

Priors. Model specification is completed by providing priors on parameters. At the data level, priors are needed for the collection of auxiliary parameters $\boldsymbol{\alpha}_{ij}$ and trial variance τ_j^2. We choose conjugate forms that are weakly informed and have mass across a broad range of plausible values:

$$\boldsymbol{\alpha}_{ij} \sim \mathrm{N}_P(\boldsymbol{m}_\alpha, \boldsymbol{S}_\alpha),$$

$$\tau_j \sim \text{Inverse-}\chi^2(1, s_j^2),$$

where the inverse-$\chi^2(a, b)$ where parameters a and b denote the degrees-of-freedom and scale of the distribution, respectively. Values of $\boldsymbol{m}_\alpha$, $\boldsymbol{S}_\alpha$, and s_j must be chosen before hand. For example, in the treatment-control design with subsecond response times as data, we placed a normally-distributed prior on α_{ij} with means of 800 ms and standard deviations of 1000 ms. Likewise, the value of s_j reflects an anticipated standard deviation with repeated trials. We set it to 200 ms as this is fairly typical. The inverse-χ^2 with 1 degree of freedom is a fat-tailed distribution, and with this setting, there is sizable prior mass from 20 ms to 2000 ms. The choices here are benign in that they have no undue influence on posterior quantities.

At the latent level, priors are needed for for task mean, $\boldsymbol{\mu}$, factor scores, $\boldsymbol{\lambda}$, and

residual variances, $\boldsymbol{\delta}^2$. Again, conjugate forms are broad, flexible, and convenient:

$$\mu_j \sim \mathrm{N}(a_j, b_j^2)$$

$$\lambda_{jm} \sim \mathrm{N}(0, c_j^2)$$

$$\delta_j^2 \sim \text{Inverse-}\chi^2(1, d_j^2)$$

As before, values of constants must be chosen before hand. We used Haaf, Hoffstadt, and Lesche (2024), who surveyed 64 cognitive control tasks, as guidance. The priors on $\boldsymbol{\mu}$ may be diffuse. In the executive-function task battery example, typical differences between congruent and incongruent conditions are 60 ms, and we chose a_j to be this value and $b_j = 100$ ms such that a broad range of effects on the tens to hundreds of milliseconds are plausible. We used more informed priors for factor loadings and residuals. According to Haaf et al. (2024), we may expect that the total variation in $\boldsymbol{\theta}$ is about 35 ms in standard deviation. Guided by this value, we chose c_j and d_j to be 25 ms. We explore sensitivity to these choices subsequently.

Rotations. It is well known that factor models are identified only up to rotation. A matrix $\boldsymbol{A}$ is called a rotation matrix if $\boldsymbol{A}' = \boldsymbol{A}^{-1}$ and that the determinant of $\boldsymbol{A}$ is 1 in value. A good example of such a matrix is Euler's rotation matrix:

$$\boldsymbol{A}(\phi) = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix},$$

To confirm this is a rotation matrix, it is straightforward to show that for all ϕ, $\det(\boldsymbol{A}(\phi)) = 1$ and $\boldsymbol{A}' = \boldsymbol{A}^{-1}$. Rotations are problematic because different factor-loadings give rise to the same covariance matrix $\boldsymbol{\Sigma}$. Here is how. Note that in the factor model, $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + D(\boldsymbol{\delta}^2)$ for some $\boldsymbol{\Lambda}$. We can rotate $\boldsymbol{\Lambda}$ by $\boldsymbol{A}$ to yield a new factor loading matrix $\boldsymbol{\Lambda}_r = \boldsymbol{\Lambda}\boldsymbol{A}$. Note that $\boldsymbol{\Lambda}_r\boldsymbol{\Lambda}_r' = \boldsymbol{\Lambda}\boldsymbol{A}(\boldsymbol{\Lambda}\boldsymbol{A})' = \boldsymbol{\Lambda}\boldsymbol{A}\boldsymbol{A}'\boldsymbol{\Lambda}' = \boldsymbol{\Lambda}\boldsymbol{A}\boldsymbol{A}^{-1}\boldsymbol{\Lambda}' = \boldsymbol{\Lambda}\boldsymbol{I}\boldsymbol{\Lambda}' = \boldsymbol{\Lambda}\boldsymbol{\Lambda}'$. Hence, $\boldsymbol{\Sigma}$ is concordant with both $\boldsymbol{\Lambda}$ and $\boldsymbol{\Lambda}_r$; both are equally good factor-loading solutions for

any data set.

In frequentist analysis, this rotational indeterminancy is benign. Analysts rotate their solutions to aid interpretation much like a hiker might rotate a topographic map to match the view of the landscape. A good example of rotation-to-aid-intepretation is the varimax rotation (Kaiser, 1958) where factors are rotated so that tasks load highly on a minimal number of factors. Even though rotational indeterminacy is benign in frequentist analysis, it is far more difficult in Bayesian analysis. The problem comes about in sampling with MCMC. Different iterations in the MCMC chain often correspond to different rotations.

Figure 4A shows this complication. Plotted in red are factor loadings for Task 1 for different iterations. They sweep a large arc, with the arc representing different rotations for different iterations. Plotted in blue are the loading for Task 4, and they sweep too large of an arc for the same reason. The two solid lines correspond to a single iteration, and they show the relationship between Task 1 and Task 4. The cosine of the angle is the correlation between the tasks on the iteration, and it is about .62 in value. The dashed lines correspond to a different iteration, and the cosine of the angle is also about .62 even though the iteration is rotated compared to the first. The stability of the correlation is expected as different rotations on different iterations are samples from the same covariance matrix.

The problem remains that the posteriors for factor loadings are artifactually too diffuse. Solutions to this rotation problem with MCMC are two-fold: One solution is to use priors to set a specific rotation where the first task loads only onto the first factor, the second task loads only onto the first two factors and so on. The requirement is that $\boldsymbol{\lambda}$ forms a lower triangular matrix with positive diagonal elements, and examples of this approach are in Ghosh and Dunson (2009), Geweke and Zhou (1996) and Vandekerckhove (2014). The drawback of this approach is that the choice of which loadings to set to zero is arbitrary yet impactful (Chen, Li, & Zhang, 2020). An alternative solution is to use post-sampling alignment algorithms (Papastamoulis & Ntzoufras, 2022; Poworoznek,

Ferrari, & Dunson, 2021; Ročková & George, 2016). In these approaches, chains are unrestricted and each may correspond to different rotations as they do in Figure 4A. Afterwards, loadings on each iterations are aligned to a common rotation. Figure 4B shows the results after alignment with the Poworoznek et al.'s `jointrot` algorithm. As can be seen, this algorithm works exceedingly well and the excessive diffusion from rotation indeterminancy is eliminated.

Implementation

The above development may seem daunting especially to psychologists who are used to working in `Lavaan` and other extant factor analysis programs. There may be much new here including Bayesian models, hierarchical models, and post-sampling alignment. Analysis is conveniently performed in R with JAGS. We provide a supplement at xxxx that provides the code and step-by-step instructions for using it.

Performance

Now that the model is specified and analyzed, the next step is assessing the quality of the analysis.

Mixing

The first task is assessing the mixing in the chains. Mixing may be problematic in Bayesian factor models, especially when the conditional model is used with Gibbs steps (Ghosh & Dunson, 2009). The main culprit in the conditional model is the trade-off between factor loadings λ_{jm} and factor scores, η_{im}. These enter into the model as products, $\lambda_{jm}\eta_{im}$, which means that the product is preserved if loadings are multiplied by a constant while scores are divided by the same constant. Indeed, we do find poor mixing when we derive the full-conditional posteriors and sample with Gibbs sampling. Fortunately, these

models are easily implemented in JAGS and Stan, which do not rely on Gibbs sampling. We have implemented the conditional and marginal hierarchical models in both JAGS and Stan, and highlight the case here for the JAGS implementation of the conditional hierarchical model as this runs most quickly. Figure 5 shows the evidence for good mixing. The first two panels show the case for selected factor loading and factor score parameters. The last panel shows a boxplot of effective sample sizes (Kong, Liu, & Wong, 1994) for all parameters. The chain was run for 5000 iterations, and as can be seen, the vast majority of effective sample sizes are within 80% of this value. Hence, we can be assured that posterior quantities are based on at least 4000 effective samples. The bottom line is that mixing is not problematic in modern general-purpose samplers.

Parameter Recovery Parameter recovery is straightforward to assess as true values are known. Figure 6 shows the quality of recovery. Panel A shows the case for factor loadings. The true values are shown as a point enumerated by task; the 95% credible ellipses on posterior estimates of factor loadings are shown in matching color. Overall, given the high-noise levels in the simulared data, the recovery is as expected. Panel B shows factor loadings, and there is regularization or shrinkage as expected for high-noise data. Panel C is perhaps the most important. It shows the posterior mean residual standard deviation δ_j along with 95% credible intervals. The true value, 20 ms, is shown as a horizontal line. Key here is that the the posterior mean is **not** inflated as would be the case for take-the-mean scores. Panel D shows the posterior distribution of the salience of the factors and the residual. The true value is 1/3 of the variance for each, and the recovery is quite reasonable for a high-noise data. Panel E shows the uncertainty in correlations between tasks. There is much uncertainty, but that too is expected.

Standardizing Factor Loadings For Decomposing Correlation Matrices

Consider the manifest factor model $\boldsymbol{Y}_i \sim N_J(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + D(\boldsymbol{\delta}^2)$. One issue in this model is that $\boldsymbol{\Sigma}$ contains physical units. For example if one measure is the individual's

height in inches and the other is the individual's weight in pounds, then elements of $\boldsymbol{\Sigma}$ are in inches^2, inches$\times$pounds, and pounds^2. The presence of different units for different measures makes visualizing and interpreting variances and loadings challenging. The same problem holds for the hierarchical version where latent scores, $\boldsymbol{\theta}$, have physical units as well.

To ease this interpretation burden, analysts often work in correlations rather than variances and in standardized unit-free loadings rather than loadings with units. One approach is to standardize model parameters (Anderson, 2003; Bartholomew, Knott, & Moustaki, 2011). The loadings standardized by total variance are:

$$\lambda_{jm}^* = \frac{\lambda_{jm}}{\sqrt{\sum_{m'} \lambda_{jm'}^2 + \delta_j^2}}.$$

When factor loadings are standardized this way, they decompose the correlation matrix rather than the variance. Here is how it works: Recall that $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + D(\boldsymbol{\delta}^2)$. Any variance matrix may be re-expressed as a correlation matrix as follows: $\boldsymbol{\Sigma} = D(\boldsymbol{\sigma})\boldsymbol{\rho}D(\boldsymbol{\sigma})$ where $D(\boldsymbol{\sigma})$ is a diagonal matrix with $\boldsymbol{\sigma}$, a vector of standard deviations, on the diagonal, and $\boldsymbol{\rho}$ is the correlation matrix. For the factor model, $\sigma_j = \sqrt{\sum_{m'} \lambda_{jm'}^2 + \delta_j^2}$. Therefore, standardized factor loadings are the factor decomposition of correlation matrices given by:

$$\boldsymbol{\rho} = \boldsymbol{\Lambda}^*(\boldsymbol{\Lambda}^*)' + D((\boldsymbol{\delta}^*)^2),$$

where $\boldsymbol{\Lambda}^*$ is the factor loading matrix of standardized loadings and $\delta_j^* = \delta_j / \sqrt{\sum_{m'} \lambda_{jm'}^2 + \delta_j^2}$ is the standardized residual.

We will continue to place priors on unstandardized loadings and residuals. For convenience, we will report correlations, standardized loadings and standardized residuals in the following applications.

Application: Individual differences in visual illusions

One main question in the vision literature is whether people who are susceptible to one type of visual illusion are susceptible to other types as well. There is a division in the literate were some authors advocate for a general-susceptibility factor (Makowski, Te, Kirk, Liang, & Chen, 2023) while others advocate that performance across illusions is at best weakly correlated (Cretenoud et al., 2019). We analyze data from Mehrvarz, Popat, and Rouder (2025) on individual differences in perceiving visual illusions. An example of one of these—the Brentano illusion—is shown in Figure 7A. There are two demonstration examples, and consider for a moment just the upper one with the outer arrows pointing rightward and the inner arrow pointing leftward. Individuals had to adjust the location of the inner arrow so that the red and blue lines equaled each other in length. Though it may not appear it, they are equal in the figure. Individuals tend to set the inner arrow too far to the left lengthening the red line at the expense of the blue line. The lower demonstration has the arrows pointing in the reverse direction, and the participants here tend to set the inner arrow too far to the right lengthening the blue line at the expense of the red line. Figure 7B shows a Zöllner Illusion; the individual adjusted the angle of the near horizontal lines until they were perfectly parallel. In the figure, the lines are indeed perfectly parallel, but because of the illusion, individual tend to apply a counter clockwise rotation to the 1rst, 3rd, and 5th line and a clockwise rotation to the 2nd and 4th line.

The structure of the battery informs the subsequent modeling. In the battery, there where were two versions of each tasks. The two versions of the Brentano illusion are shown in Figure 7A. Mehrvarz, Popat, and Rouder (2024) included two versions to control for response bias. For example, if a participant was biased toward setting the chevron to the left, it would enhance the illusion in one version and attenuate it in the other. Mehrvarz et al. (2024) ran a total of five illusions, the Brentano Illusion (both versions shown), the Zöllner Illusion (one version shown), the Ebbinghaus Illusion (not shown), the Poggendorf

Illusion (not shown), and the Ponzo Illusion (not shown). There were two versions of each illusion to control for response bias. The two versions of each of five illusions results in a total of 10 tasks.

Conventional Analysis

Data for conventional analysis were derived by taking the mean across replicate trials. A single score, the strength of the illusion, was computed per individual per task. Figure 7C shows the sample correlations of these scores across tasks. Inspection of the sample correlations shows that for three illusions, there is a high within-illusion correlations. Other than that, correlations are modest.

Mehrvarz et al. (2025) report a conventional exploratory factor analysis on take-the-mean scores. Yet, the results were not so insightful. Factors rotated by varimax largely assigned factors to tasks, with the first three factors isolating the three tasks with high within-illusion correlations. Here, the design choice to control for possible response bias is what is most recovered—when there is little response bias there is a high degree of uninteresting correlation across the different versions of the same task.

To gain more insight and answer the theoretical questions about whether there is a common susceptibility factor, Mehrvarz et al. constructed a conventional bifactor model. There were five illusion-specific factors that were used to account for correlations within illusions. The remaining factor—the factor of interest—potentially loaded on all tasks. The authors, however, were unable to fit this model (or similar ones) in a conventional framework because resulting variances were negative indicating Heywood cases (Cooperman & Waller, 2022; Rindskopf, 1984).

Hierarchical Analysis

Bayesian analysis with proper priors is not hobbled by Heywood cases (Lopes & West, 2004). We fit a hierarchical Bayesian bifactor model with five specific factors (one for each illusion) and two general factors. The structure of the model is shown in Table 2, which shows standardized loadings Λ^*. Where there are "-", the loading was fixed to zero. The first factor therefore is a Zöllner factor as it loads only on the two versions of the Zöllner illusion. The first five factors are illusion-specific and capture within-illusion correlations. These five factors are treated as nuisance as the main substantive questions are not about correlations within illusions but across them. The substantively important factors were the sixth and seventh. These factors are exploratory in that they may load on all 10 tasks.

Data: Let X_{ijk} be the kth observation for the ith individual in the jth task ($i = 1, \ldots, 138$, $j = 1, \ldots, 10$, $k = 1, \ldots, K_{ij}$) in natural units such as pixels or angular displacement. Observations were scaled by the original authors as follows: Let v_{ij} be the sample variance of observations for the ith person and jth task. Effect-size-scaled observations Y_{ijk} were defined as $Y_{ijk} = X_{ijk}/s_j$, where $s_j = \sqrt{(\sum_i v_{ij})/I}$ is a task-averaged standard deviation. The original authors performed this scaling to obviate the need for units in all of there analyses and we follow them here. These effect-size-scaled observations served as input to the hierarchical factor model.

Data Model. The data model is $Y_{ijk} \sim \mathrm{N}(\theta_{ij}, \tau_j^2)$.

Factor Model. The factor model is given in (4) where $\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{i,10})'$ for the 10 tasks.

Priors. The priors were as follows: $\mu_j \sim \mathrm{N}(2, 2^2)$, $\lambda_{jm} \sim \mathrm{N}(0, 1)$, $\tau_j^2 \sim \text{Inverse-}\chi^2(1, 1)$, and $\delta_j^2 \sim \text{Inverse-}\chi^2(1, 1)$.

Implementation. Chains were run for 5000 iterations with the first 500 serving as burn in. To align differing rotations within the MCMC chain, Factors 6 and 7 were aligned with the Poworoznek et al. (2021) post-sampling alignment algorithm. Nuisance factors do

not enter into rotational alignment as rotations are fixed by the zero loadings. These factors still may have a sign indeterminacy across iterations, and to address this issue, we fixed the loadings to be positive for the first version of each task for these nuisance factors. The resulting chains mixed well, and the 4500 iterations were more than sufficient for consideration of posterior quantities. A supplementary document at xxxx provides a tutorial with code on fitting the model to the data set.

Results

We report the standardized version of factor-model parameters. The main output is the matrix $\boldsymbol{\rho}$, the correlation matrix, and it's factor-model decomposition into loadings, and residuals. This decomposition is shown in Figure 8. First, note that the model based correlation matrix on the left is highly similar to the sample-based correlation in Figure 7. This similarity means that there was not much disattenuation and that trial noise is not problematic in this application. Second, the decomposition is into nuisance factors (Factors 1 through 5), the two general factors (Factors 6 and 7), and the residual. Factor 6 seems quite robust in accounting for variation; Factor 7 does not. These result indicate that there is a single factor innterpretable as a general susceptibility to visual illusions. The conclusion is reinforced by inspecting the standardized factor loadings in Table 2. Importantly, Factor 6 loads on all tasks; that is, it is indeed general.

Figure 9 shows the uncertainty in critical parameters. The proportional of variance accounted for by the 6th factor—susceptibility to illusions—centers at 23% with plausible values ranging from 15% to 30%. This value is substantially more than the 6% accounted for by the seventh factor confirming the salience of a one-factor solution after accounting for within-illusion correlations. Across illusion correlations are fairly small and fairly well localized. And although the one-factor structure is clear, the associated amount of variance predicted from one illusion to another is a rather small 5% in value. In this sense, the results are in line too with researchers who stress weak associations across different visual

illusions (Cottier, Turner, Holcombe, & Hogendoorn, 2023; Cretenoud, Grzeczkowski, Kunchulia, & Herzog, 2021; Grzeczkowski, Clarke, Francis, Mast, & Herzog, 2017; Mazuz, Kessler, & Ganel, 2023).

Application: Individual Differences in Cognitive Control

We analyze data from Enkavi et al. (2019) to assess the dimensionality of individual differences in cognitive control. Included are six behavioral experimental tasks: a flanker interference task (Eriksen & Eriksen, 1974), a global-to-local interference task (Navon, 1977), a control-of-negative-priming task (Egner & Hirsch, 2005), a Simon interference task (Simon, 1969), and a Stroop interference task (Stroop, 1935). Each task consisted of a congruent and an incongruent condition with the incongruent condition requiring more control to ignore the interfering information. In each task, the dependent measure was response time, and the critical contrast to measure cognitive control was the difference in response time between the incongruent and congruent condition. Response times in all tasks were about 700 ms and critical contrasts were about 70 ms. The main theoretical question was whether there was a single factor corresponding to an ability to ignore interfering information.

Figure 10A shows the observed correlations among the critical contrast across tasks where these scores come from the take-the-mean approach. None of these correlations are too large (maximum is .24), and several are near zero. One question is whether this lack of structure reflects attenuation from excessive trial noise, or, alternatively, whether it reflects a lack of true correlation across the tasks.

Hierarchical Analysis

We fit a hierarchical Bayesian factor model with three general factors.

Data Model. Let $Y_{ijk\ell}$ be a response time for the ith individual, $i = 1, \ldots, 203$, for

the jth task, $j = 1, \ldots, 6$, in the kth congruency condition ($k = 1, 2$), for the ℓth replicate trial, $\ell = 1, \ldots, L_{ijk}$. The data model is $Y_{ijk} \sim \mathrm{N}(\alpha_{ij} + x_k\theta_{ij}, \tau_j^2)$, where $x_k = -1/2,\ 1/2$ and θ_{ij}, the congruency effect, is the target of inquiry.

Factor Model. The factor model is given in (4) where $\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{i,6})'$ for the 6 tasks.

Priors. The priors were as follows: $\mu_j \sim \mathrm{N}(70, 100^2)$, $\lambda_{jm} \sim \mathrm{N}(0, 25^2)$, $\tau_j^2 \sim \text{Inverse-}\chi^2(1, 200^2)$, $\delta_j^2 \sim \text{Inverse-}\chi^2(1, 25^2)$, and $\alpha_{ij} \sim \mathrm{N}(1000, 1000^2)$.

Implementation. Chains were run for 5000 iterations with the first 500 serving as burn in. Factors were aligned with the Poworoznek et al. (2021) post-sampling alignment algorithm. Chains mixed well as in the preceding example, and the 4500 iterations were more than sufficient for consideration of posterior quantities.

Results

Figure 10B shows the disattenuated correlation matrix from the model. The degree of disattenuation is sizable because there is a high degree of trial noise and relatively few trials (Rouder & Mehrvarz, 2024). The largest disattenuated correlation value is .36, which is 50% larger than the observed value.

Figure 10C-F show the factor decomposition of this correlation matrix, and Figure 10G shows the proportion of variance accounted for by each component. Loadings are shown in Table 3. With a few exceptions, factor loadings are fairly small in value. Much of the variability is unique, and the factors themselves are difficult to interpret. Factor 1 reflects the correlation between the Global-Local interference task and the Stroop task; Factor 2 reflects the correlation between Stroop and task-switching tasks, and Factor 3 does not have an obvious role. In summary, the factor structure simply highlights a few idiosyncratic correlations that are greater in magnitude than the others. There is no evidence for a general inhibition or cognitive control factor. Figure 10H shows that

correlation coefficients are not well localized. In summary, there is little structure that can be identified perhaps because there is so much trial noise. The hierarchical approach does well here not overstating confidence when there is much trial noise, and this facet is in contrast to the take-the-mean approach which results in high confidence on low values.

Hierarchical Unconstrained-Variance Model

Factor models provide constraint on covariance. The smaller the number of factors, the larger the amount of constraint. Figure 11 shows how this constraint works. Figure 11A is a one-factor constraint, that is, it the correlation when $\mathbf{\Lambda}$, factor loadings, are restricted to a single column. The one-factor constraint is fairly stringent, and the resulting correlation matrix is limited in the number of patterns. Figure 11B is the two-factor constraint, and because the model is more complex, more diversity in the correlation matrix may be seen. In particular, stronger positive correlations and negative correlations are possible. Figure 11C shows the three-factor constraint, and increasing dimensionality allows for more subtlety in the resulting correlation matrix.

In conventional manifest factor models, this series of increasing factor complexity has an obvious endpoint—the sample correlation or sample variance matrix. Factor modeling, like PCA, is a way of approximating this sample correlation matrix with reduced dimensionality. In the hierarchical factor model, the endpoint is $\mathbf{\Sigma}$, the covariance of $\boldsymbol{\theta}_i$ across individuals: $\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}' + D(\boldsymbol{\delta}^2)$. The right-hand side serves as constraint, especially if the number of columns in $\mathbf{\Lambda}$ is much smaller than the number of tasks J. The problem is that $\mathbf{\Sigma}$ is not observed, in fact, it may not be obvious how to estimate it is without the factor constraint. In the manifest version, we may check how well our factor solution approximates the sample variance matrix. In the hierarchical model, the factor model itself does not provide this check.

Needed is a hierarchical model without the factor constraint so that we may visualize

$\boldsymbol{\Sigma}$. This need is met by placing diffuse priors on $\boldsymbol{\Sigma}$ that impart a minimal prior information. Suitable choices are the inverse Wishart (O'Hagan & Forster, 2004), the LKJ prior (Lewandowski, Kurowicka, & Joe, 2009), and the scaled inverse Wishart (Huang & Wand, 2013). We have found that inverse Wishart is computationally convenient but must be tuned carefully imposing a burden on the analyst; the LKJ is computationally slow though more robust to prior settings (see Yang & Rouder, 2025); and the scaled inverse Wishart is both computation convenient and robust to prior settings. We specify the following hierarchical unconstrainted variance model:

$$\boldsymbol{\theta}_i \mid \boldsymbol{\Sigma} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$\boldsymbol{\Sigma} \sim \text{Scaled Inverse Wishart}(\boldsymbol{S}, v).$$

The scaled inverse Wishart has two settings, a scale vector $\boldsymbol{S}$, and a degrees-of-freedom setting v. It is easiest to understand the role of these two settings through the decomposition $\boldsymbol{\Sigma} = D(\boldsymbol{\sigma})\boldsymbol{\rho}D(\boldsymbol{\sigma})$. The degrees-of-freedom parameter v is a prior on correlation matrix $\boldsymbol{\rho}$, and greater values of v correspond to more mass near zero. At $v = 2$, the marginal prior on each correlation is uniformly distributed, and this is why we used this value throughout. The scale parameter S affects only $\boldsymbol{\sigma}$, the standard deviation, and not $\boldsymbol{\rho}$, the correlation matrix of interest. Hence, it's setting is far less influential on the posterior values of correlation. For the cognitive control data set, we set each element S_j to 40 ms, a reasonable standard deviation on θ_{ij} for this set.

Figure 11D shows the correlation matrix for the unconstrained-variance model. It differs modestly from the three factor model indicating that there is still a small degree of outstanding variance. Given that the factors account for a small degree of variance and are essentially uninterpretable, we are not so concerned about additional variation beyond the three-factor solution. In applications where correlations are more salient and factor structures are more interpretable, it may be substantively important to use the

unconstrained-variance model as a catch-all to understand the factor-model approximation.

Influence of Prior Specification

Bayesian analysts need be aware of the role of arbitrary choices in model and prior specifications. In the hierarchical set up, the factor model (4) serves as a prior on $\boldsymbol{\theta}$ in the data model (3). The factor model regularizes estimation of the correlations of these latent scores across tasks. For example, the data model itself could be easily estimated with noninformative priors on $\boldsymbol{\theta}_{ij}$, albeit the posterior means would exactly match the take-the-mean aggregate, correlations would be dramatically attenuated; and factors would be less salient. The factor structure on $\boldsymbol{\theta}_i$ is a form of regularization to a specific structure that is interpretable. In this sense, posteriors need to be *appropriately* sensitive to prior specification rather than completely insensitive. Regularization almost always involves making a few choices and tuning a few parameters, and a good example of this tuning is in Lasso and ridge regression. Here, prior specification is a way of tuning regularization.

We have explored the influence of various prior settings. The most important element of prior setting is the scale of factor loadings relative to the scale of residual variation. In the illusions application, we had scaled the data by the trial noise, and used scales of 1.0 on factor loadings and residual variances, that is $c_j = d_j = 1$. In the cognitive control application, we used the meta-analysis by Haaf et al. (2024) to set $c_j = d_j = 25$ ms. In both cases, the ratio between c_j and d_j is 1-to-1. As it turns out, it is this ratio that is important in controlling regularization as it is the ratio of structure to residual. To understand the influence of this ratio, we created alternative prior ratios of 4-to-1 and 1-to-4 in addition to the 1-to-1 ratio used earlier. Figure 12 shows the resulting posteriors on the proportion of variance accounted for by the factor structure $\boldsymbol{\Lambda\Lambda}'$ relative to $\boldsymbol{\Sigma}$. As can be seen, the posterior is dependent on this prior ratio.

Two points emerge: First, the influence of prior settings on the ratio is a function of

the quality of the data. The data are of higher quality in the visual illusion set than in the cognitive control set inasmuch as there is far less trial noise and latent correlations are better localized. Because the data have higher quality in the visual illusion set, the posterior is less sensitive to the prior ratio than it is in the cognitive control set. Second, the ratio tunes the degree of regularization—the greater the ratio the greater the salience of the factor structure. Tuning parameters are part of most regularization approaches, and these parameters must be set not only in hierarchical modeling but in other methods as well such as lasso and ridge regression. The Bayesian hierarchical factor model is no different than other methods in this regard.

Overall, we think the 1-to-1 ratio serves as a reasonable default for many applications. That said, researchers are encouraged to explore how reasonable variations in this ratio affect their results. Understanding this sensitivity provides researchers with context for drawing conclusions.

Discussion

Factor models were developed in psychology to explore the covariation in survey data. Each individual provides responses to a number of items, and the covariation in the matrix of responses is decomposed into factors. Data from experiments however differ from survey data in two important regards: First, data from experiments are comprised of repeated trials nested within people, tasks, and conditions. This is an advantage as the variation across replicate trials may be accounted, which in turn, perhaps allows for more precise understanding of the covariation across individuals, conditions, and tasks. Second, data from experiments is often far more noise prone than survey data. In a typical response-time task, for example, one is trying to localize some 25 ms of variation across individuals in an effect whereas each trial adds some 200 ms of residual variability. In realistic designs, this degree of trial noise tends to lead to attenuation of correlations—observed correlations are about one-half of true values (Haaf et al., 2024;

Rouder et al., 2023). Moreover, subsequent latent-variable analyses, trial noise diminishes the salience of the latent variables themselves. In turn, this diminished saliency makes it more difficult to understand the structure in these latent variables. Several authors have noted the appeal of Bayesian hierarchical models to disattenuate these correlations and recover patterns of covariation (Haines et al., 2025; Matzke et al., 2017; Rouder & Haaf, 2019). Here we show that a simple data model of trial noise may be yoked with factor models to reveal the salience of underlying latent variables. There are a few technical challenges; fortunately, new computationally-convenient tools now exist to meet them.

Limitations and Future Directions

Model Comparison. Perhaps the most obvious limitation of the hierarchical factor models developed here is the lack of a formal mechanism for inference on structure. Needed is a way of comparing different models and stating evidence for various dimensional constraints. In frequentist factor-model analysis, model comparison formally occurs through AIC and BIC fit statistics (Vrieze, 2012) as well as through considerations of eigenvalues (Kaiser, 1960; Preacher & MacCallum, 2003). Model comparison in Bayesian analysis remains a contested topic with strong opinions about the utility of Bayes factors and posterior-predictive methods (Morey, Wagenmakers, & Rouder, 2016; Vanpaemel, 2010; Vehtari, Gelman, & Gabry, 2017). The Bayesian model comparison development from statistics has traditionally centered on Bayes factors (Jeffreys, 1961; Kass & Raftery, 1995), though we suspect it may not be too difficult to implement WAIC comparisons (Vehtari et al., 2017). Here, we discuss current developments with Bayes factors.

Bayes factor computations with nonhierarchical factor models have followed two related paths called here the *separate specification* approach and the *encompassing* approach. The separate specification approach is to (i) specify separate, competing models each with a different dimensionality, and (ii) compare them through marginal likelihood computations. Traditionally, this approach has not proven popular for factor models. We

suspect it is because much development of factor models in the statistics and computer science literature is designed to address high-dimensional problems such as those in genetics and image processing. In high-dimensional spaces, the number of possible models and possible model comparisons is unwieldy. Instead, the encompassing approach is used. Here there is a single, maximal model. Dimensions are then adaptively eliminated in analysis if they are not supported by the data. The encompassing approach is in the spirit of Lasso regularization where penalties are used to eliminate dimensions that have little support from the data (e.g., Jiang, He, & Zhang, 2016). Examples of this encompassing approach for latent variables include automatic-relevance-detection priors (MacKay, 1992), spike-and-slab priors (Oh & Kim, 2010), and multiplicative gamma priors (Bhattacharya & Dunson, 2011).

Factor model analysis in psychology often comprises just a few dimensions or factors. In the current applications in cognition, one hopes to discriminate among a handful of possible factors. Even in the most advanced applications in say personality and intelligence, the dimensionality of the problem is limited. For these applications, we suspect the separate-specification approach will be appropriate and convenient. Computing marginal likelihoods (Bayes factors) is often computationally involved (Kass, 1993). Fortunately, Bayesian analysts have a well-stocked toolbox of computational approaches including Lapace approximation (Shun & McCullagh, 1995), importance sampling (Sinharay & Stern, 2005), and bridge sampling (Gronau et al., 2017; Meng & Wong, 1996). And many authors have used a variety of techniques to estimate Bayes factors in factor models including Geweke and Zhou (1996), Lopes and West (2004), and Pitt, Chan, and Kohn (2006). Most recently, Schnuerch, Mehrvarz, and Rouder (2025) provide an easy-to-use bridge-sampling Bayes-factor solution for manifest factor models. How well any of these approaches work in hierarchical-factor-model settings is a topic for future research.

Realistic Psychological-Process Models. The data model here are linear models. Although linear models are a workhorse of statistical analysis, they rarely serve as

insightful models of mental processes. Over the last 60 years, researchers have proposed nonlinear accounts of mental processes, and popular model classes include multinomial tree models (Riefer & Batchelder, 1988), evidence accumulation models (Ratcliff & Rouder, 1998; Vickers & Smith, 1985), race models (Brown & Heathcote, 2008; Logan, 1988), and information gating models (Frank, 2005; Townsend & Ashby, 1982). And in the last twenty years, it has become increasingly common to develop hierarchical versions of these process models that account for trial noise as well as variation across individuals (Heck, Arnold, & Arnold, 2018; Rouder, Province, Morey, Gomez, & Heathcote, 2015; Vandekerckhove, Tuerlinckx, & Lee, 2011).

There has been notable progress in combining latent variable models with cognitive process models. One of the early demonstrations comes from Vandekerckhove (2014) who used a factor decomposition of accumulation-rate parameters in the diffusion model of perception (Ratcliff & McKoon, 2008). To simplify analysis, Vandekerckhove used an algebraic decomposition rather than a stochastic one—drift rates were assumed to be the product of factor loadings and factor scores without any additional residual variation. Turner and colleagues (Kang et al., 2022; Turner, Wang, & Merkle, 2017) develop hierarchical factor models much like those presented here for diffusion model parameters as a way of linking fMRI activation with behavior. The factor models in their work serve as a regularization device in localizing brain-behavior links. Most recently, Stevenson, Innes, Gronau, Miletic, et al. (2024) provide a general hierarchical factor model solution for process models such as the diffusion model. Stevenson et al. focus on the factor structure in drift rates. Rouder (2025) provides a different example—he integrates a factor model with a psychophysical model to assess the factor structure of thresholds. It is a matter for future research to compare the pragmatic differences in choosing a process model, as done by Stevenson, Innes, Gronau, Miletic, et al. (2024) and Rouder (2025), to the more generic and easier-to-analyze linear models used here.

References

Anderson, T. W. (2003). *An introduction to multivariate statistical analysis* (Vol. 3). New York: Wiley.

Ando, T. (2009). Bayesian factor analysis with fat-tailed factors and its exact marginal likelihood. *Journal of Multivariate Analysis*, *100*(8), 1717–1726. Retrieved from https://www.sciencedirect.com/science/article/pii/S0047259X09000396

Bartholomew, D. J., Knott, M., & Moustaki, I. (2011). *Latent variable models and factor analysis: A unified approach.* John Wiley & Sons. Retrieved from https://books.google.com/books?hl=en&lr=&id=5YxvE1S56DwC&oi=fnd&pg=PT6& dq=Bartholomew,+D.+J.,+Knott,+M.,+%26+Moustaki,+I.+(2011).+Latent+ Variable+Models+and+Factor+Analysis:+A+Unified+Approach+(3rd+ed.).+Wiley. &ots=0IGoBPPlel&sig=Vd91yDtodeH3DSVioIqTiKsfss0

Bhattacharya, A., & Dunson, D. B. (2011). Sparse Bayesian infinite factor models. *Biometrika*, *98*(2), 291–306. Retrieved from https://academic.oup.com/biomet/article-pdf/doi/10.1093/biomet/asr013/46695653/asr013.pdf

Bollen, K. A. (1989). *Structural equations with latent variables.* Wiley.

Bolt, T., Nomi, J. S., Yeo, B. T. T., & Uddin, L. Q. (2017). Data-Driven Extraction of a Nested Model of Human Brain Function. *The Journal of Neuroscience*, *37*(30), 7263–7277. doi:10.1523/JNEUROSCI.0323-17.2017

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice reaction time: Linear ballistic accumulation. *Cognitive Psychology*, *57*, 153–178.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Bettencourt, M., . . . Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*.

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., . . . Moffitt, T. E. (2014). The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders? *Clinical Psychological Science*, *2*(2), 119–137.

doi:10.1177/2167702613497473

Chen, Y., Li, X., & Zhang, S. (2020). Structured latent factor analysis for large-scale data: Identifiability, estimability, and their implications. *Journal of the American Statistical Association.* Retrieved from 10.1080/01621459.2019.1635485

Congdon, P. (2003). *Applied Bayesian Modelling.* John Wiley & Sons. Retrieved from https://books.google.com?id=ImejAwAAQBAJ

Cooperman, A. W., & Waller, N. G. (2022). Heywood you go away! Examining causes, effects, and treatments for Heywood cases in exploratory factor analysis. *Psychological Methods*, *27*(2), 156–176. doi:10.1037/met0000384

Cottier, T. V., Turner, W., Holcombe, A. O., & Hogendoorn, H. (2023). Exploring the extent to which shared mechanisms contribute to motion-position illusions. *Journal of Vision*, *23*(10), 8. doi:10.1167/jov.23.10.8

Cretenoud, A. F., Grzeczkowski, L., Kunchulia, M., & Herzog, M. H. (2021). Individual differences in the perception of visual illusions are stable across eyes, time, and measurement methods. *Journal of Vision*, *21*(5), 26–26. doi:10.1167/jov.21.5.26

Cretenoud, A. F., Karimpur, H., Grzeczkowski, L., Francis, G., Hamburger, K., & Herzog, M. H. (2019). Factors underlying visual illusions are illusion-specific but not feature-specific. *Journal of Vision*, *19*(14), 12. doi:10.1167/19.14.12

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*(11), 671–684. doi:10.1037/h0043943

Egner, T., & Hirsch, J. (2005). Where Memory Meets Attention: Neural Substrates of Negative Priming. *Journal of Cognitive Neuroscience*, *17*(11), 1774–1784. doi:10.1162/089892905774589226

Enkavi, A. Z., Eisenberg, I. W., Bissett, P. G., Mazza, G. L., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Large-scale analysis of test–retest reliabilities of self-regulation measures. *Proceedings of the National Academy of Sciences*, *116*(12), 5472–5477.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a
    target letter in a nonsearch task. *Perception & Psychophysics*, *16*, 143–149.

Fisher, R. (1921). On the "Probable Error" of a Coefficient of Correlation Deduced from a
    Small Sample. *Metron*, *1*, 3–32.

Franco-Martínez, A., Vicente-Conesa, F., Shanks, D., & Vadillo, M. A. (2024, March 5).
    Measurement and sampling noise undermine inferences about awareness in location
    probability learning: A modeling approach. doi:10.31234/osf.io/grkya

Frank, M. J. (2005). Dynamic Dopamine Modulation in the Basal Ganglia: A
    Neurocomputational Account of Cognitive Deficits in Medicated and Nonmedicated
    Parkinsonism. *Journal of Cognitive Neuroscience*, *17*(1), 51–72.
    doi:10.1162/0898929052880093

Gana, K., & Broc, G. (2019). *Structural equation modeling with lavaan.* John Wiley &
    Sons. Retrieved from https://books.google.com/books?hl=en&lr=&id=
    QMOCDwAAQBAJ&oi=fnd&pg=PP2&dq=Structural+Equation+Modeling+with+
    lavaan&ots=VmnKfYew6K&sig=3wM4yrQd4mkQjOWRopNaDH0C0IE

Geweke, J., & Zhou, G. (1996). Measuring the pricing error of the arbitrage pricing theory.
    The Review of Financial Studies, *9*(2), 557–587. Retrieved from
    https://academic.oup.com/rfs/article-abstract/9/2/557/1631132

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.

Green, M. F., Hellemann, G., Horan, W. P., Lee, J., & Wynn, J. K. (2012). From
    Perception to Functional Outcome in Schizophrenia: Modeling the Role of Ability and
    Motivation. *Archives of General Psychiatry*, *69*(12), 1216–1224.
    doi:10.1001/archgenpsychiatry.2012.652

Gronau, Q. F., Sarafoglou, A., Matzke, D., Ly, A., Boehm, U., Marsman, M., . . .
    Steingroever, H. (2017). A tutorial on bridge sampling. *Journal of Mathematical*

Psychology, *81*, 80–97. Retrieved from

http://www.sciencedirect.com/science/article/pii/S0022249617300640

Grzeczkowski, L., Clarke, A. M., Francis, G., Mast, F. W., & Herzog, M. H. (2017). About

individual differences in vision. *Vision Research*, *141*, 282–292.

doi:10.1016/j.visres.2016.10.006

Haaf, J. M., Hoffstadt, M., & Lesche, S. (2024). Attentional Control Data Collection: A

Resource for Efficient Data Reuse. doi:10.31234/osf.io/4evy6

Haines, N., Kvam, P. D., Irving, L., Smith, C. T., Beauchaine, T. P., Pitt, M. A., . . .

Turner, B. M. (2025). A tutorial on using generative models to advance psychological

science: Lessons from the reliability paradox. *Psychological Methods*.

doi:10.1037/met0000674

Heck, D. W., Arnold, N. R., & Arnold, D. (2018). TreeBUGS: An R package for

hierarchical multinomial-processing-tree modeling. *Behavior Research Methods*, *50*(1),

264–284. doi:10.3758/s13428-017-0869-7

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive

tasks do not produce reliable individual differences. *Behavioral Research Methods*.

Huang, A., & Wand, M. P. (2013). Simple Marginally Noninformative Prior Distributions

for Covariance Matrices. *Bayesian Analysis*, *8*(2), 439–452. doi:10.1214/13-BA815

Jackman, S. (2009). *Bayesian Analysis for the Social Sciences*. Chichester, United

Kingdom: John Wiley & Sons.

Jeffreys, H. (1961). *Theory of Probability (3rd Edition)*. New York: Oxford University

Press.

Jiang, Y., He, Y., & Zhang, H. (2016). Variable Selection With Prior Information for

Generalized Linear Models via the Prior LASSO Method. *Journal of the American

Statistical Association*, *111*(513), 355–376. doi:10.1080/01621459.2015.1008363

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis.

Psychometrika, *23*(3), 187–200. doi:10.1007/BF02289233

Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analysis.
    Educational and Psychological Measurement, *20*(1), 141–151.
    doi:10.1177/001316446002000116

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A
    controlled-attention view of working-memory capacity. *Journal of Experimental
    Psychology: General*, *130*(2), 169–183. Retrieved from
    http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2001-17501-
    002&loginpage=Login.asp&site=ehost-live&scope=site

Kang, I., Yi, W., & Turner, B. M. (2022). A regularization method for linking brain and
    behavior. *Psychological Methods*, *27*(3), 400. Retrieved from 10.1037/met0000387

Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera,
    M. A. (2018). The unity and diversity of executive functions: A systematic review and
    re-analysis of latent variable studies. *Psychological Bulletin*, *144*(11), 1147.

Kass, R. E. (1993). Bayes Factors in Practice. *The Statistician*, *42*, 551–560.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical
    Association*, *90*, 773–795. Retrieved from
    http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1995.10476572

Kim-Spoon, J., Herd, T., Brieant, A., Elder, J., Lee, J., Deater-Deckard, K., & King-Casas,
    B. (2021). A 4-year longitudinal neuroimaging study of cognitive control using latent
    growth modeling: Developmental changes and brain-behavior associations.
    NeuroImage, *237*, 118134. doi:10.1016/j.neuroimage.2021.118134

Kong, A., Liu, J. S., & Wong, W. H. (1994). Sequential Imputations and Bayesian Missing
    Data Problems. *Journal of the American Statistical Association*, *89*(425), 278–288.
    doi:10.2307/2291224

Lee, S.-Y. (2007). *Structural equation modelling: A Bayesian approach.* New York: Wiley.

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation
    matrices based on vines and extended onion method. *Journal of Multivariate Analysis*,

100(9), 1989–2001.

Logan, G. D. (1988). Towards an instance theory of automization. *Psychological Review,* *95*, 492–527.

Lopes, H. F., & West, M. (2004). Bayesian model assessment in factor analysis. *Statistica Sinica*, 41–67. Retrieved from https://www.jstor.org/stable/24307179?casa_token=gHhCc_ZZIL0AAAAA: 8pd7RR7uhvuygJjQ0hyo-aT9ghFSaW-Aiu2w6L6dBt0lCSHMqI1k7_ QDSgooVSi7hJFT2N5MVLq6Gm0X3PGKW92kxsN-UiEhZTrv1Rt0aC4-oq6wGyS6

MacKay, D. J. (1992). Bayesian interpolation. *Neural Computation*, *4*(3), 415–447. Retrieved from https://direct.mit.edu/neco/article-abstract/4/3/415/5639

Makowski, D., Te, A. S., Kirk, S., Liang, N. Z., & Chen, S. H. A. (2023). A novel visual illusion paradigm provides evidence for a general factor of illusion sensitivity and personality correlates. *Scientific Reports*, *13*(1), 6594. doi:10.1038/s41598-023-33148-5

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).

Mazuz, Y., Kessler, Y., & Ganel, T. (2023). The BTPI: An online battery for measuring susceptibility to visual illusions. *Journal of Vision*, *23*(10), 2. doi:10.1167/jov.23.10.2

McCrae, R. R., & Costa Jr, P. T. (1997). Personality trait structure as a human universal. *American Psychologist*, *52*(5), 509.

Mehrvarz, M., Popat, H., & Rouder, J. (2025, June 25). *Localizing Structure in Individual Differences: A Visual Illusion Case Study.* doi:10.31234/osf.io/6z92y_v2

Mehrvarz, M., Popat, H., & Rouder, J. N. (2024, May 2). *Individual Differences in Visual Illusions: Graphical and Analytic Approaches For Finding Structure in Real-World Cases.* doi:10.31234/osf.io/6z92y

Mehrvarz, M., & Rouder, J. N. (2025). Estimating correlations in low-reliability settings with constrained hierarchical models. *Behavior Research Methods*, *57*(2), 59.

doi:10.3758/s13428-024-02568-0

Meng, X., & Wong, W. H. (1996). Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statistica Sinica*, *6*, 831–860.

Merkle, E. C., Fitzsimmons, E., Uanhoro, J., & Goodrich, B. (2021). Efficient Bayesian Structural Equation Modeling in Stan. *Journal of Statistical Software*, *100*(6), 1–22. doi:10.18637/jss.v100.i06

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100.

Morey, R. D., Wagenmakers, E.-J., & Rouder, J. N. (2016). Calibrated Bayes factors should not be used: A reply to Hoijtink, van Kooten, and Hulsker. *Multivariate Behavioral Research*, *51*(1), 11–19.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognition*, *9*, 353–383.

O'Hagan, A., & Forster, J. J. (2004). *Kendall's advanced theory of statistics, volume 2B: Bayesian inference* (Vol. 2). Arnold.

Oh, H. S., & Kim, D.-G. (2010). Bayesian principal component analysis with mixture priors. *Journal of the Korean Statistical Society*, *39*(3), 387–396.

Papastamoulis, P., & Ntzoufras, I. (2022). On the identifiability of Bayesian factor analytic models. *Statistical Computing*, *32*(2), 1–29. Retrieved from 10.1007/s11222-022-10084-4

Pitt, M., Chan, D., & Kohn, R. (2006). Efficient Bayesian inference for Gaussian copula regression models. *Biometrika*, *93*(3), 537–554. Retrieved from https://academic.oup.com/biomet/article-abstract/93/3/537/380692

Plummer, M. (2003). JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*.

Poworoznek, E., Ferrari, F., & Dunson, D. (2021, July 29). *Efficiently resolving rotational ambiguity in Bayesian matrix sampling with matching.* Retrieved November 21, 2023, from http://arxiv.org/abs/2107.13783

Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's Electric Factor Analysis Machine. *Understanding Statistics*, *2*(1), 13–43. doi:10.1207/S15328031US0201_02

Ratcliff, R., & McKoon, G. M. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*, 873–922.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for decisions between two choices. *Psychological Science*, *9*, 347–356.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods. Second Edition.* Thousand Oaks, CA: Sage.

Rey-Mermet, A., Gade, M., & Oberauer, K. (2018). Should We Stop Thinking About Inhibition? Searching for Individual and Age Differences in Inhibition Ability. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* Retrieved from http://dx.doi.org/10.1037/xlm0000450

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

Rindskopf, D. (1984). Structural equation models: Empirical identification, Heywood cases, and related problems. *Sociological Methods & Research*, *13*(1), 109–119.

Ročková, V., & George, E. I. (2016). Fast Bayesian Factor Analysis via Automatic Rotations to Sparsity. *Journal of the American Statistical Association*, *111*(516), 1608–1622. doi:10.1080/01621459.2015.1100620

Rouder, J. (2025, June 21). *Random-Effects Psychophysics For Studying Individual Differences in Perception.* Retrieved June 21, 2025, from https://osf.io/7zb6p_v1

Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin and Review*, *26*(2), 452–467. Retrieved from

https://doi.org/10.3758/s13423-018-1558-y

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*, *30*(6), 2049–2066. doi:10.3758/s13423-023-02293-3

Rouder, J. N., & Mehrvarz, M. (2024). Hierarchical-Model Insights for Planning and Interpreting Individual-Difference Studies of Cognitive Abilities. *Current Directions in Psychological Science*, *33*(2), 128–135. doi:10.1177/09637214231220923

Rouder, J. N., Province, J. M., Morey, R. D., Gomez, P., & Heathcote, A. (2015). The lognormal race: A cognitive-process model of choice and latency with desirable psychometric properties. *Psychometrika*, *80*, 491–513.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, *22*(1), 53–61. doi:10.1007/BF02289209

Schnuerch, M., Mehrvarz, M., & Rouder, J. (2025, May 16). *Model comparison for factor models with Bayes factors through bridge sampling.* doi:10.31234/osf.io/4a965_v1

Shun, Z., & McCullagh, P. (1995). Laplace Approximation of High Dimensional Integrals. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(4), 749–760. doi:10.1111/j.2517-6161.1995.tb02060.x

Simon, J. R. (1969). Reactions toward the source of stimulation. *Journal of Experimental Psychology*, *81*, 174–176.

Sinharay, S., & Stern, H. S. (2005). An Empirical Comparison of Methods for Computing Bayes Factors in Generalized Linear Mixed Models. *Journal of Computational and Graphical Statistics*, *14*(2), 415–435. doi:10.1198/106186005X47471

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, longitudinal, and structural equation models.* Boca Raton: CRC Press.

Snyder, H. R., Friedman, N. P., & Hankin, B. L. (2021). Associations Between Task Performance and Self-Report Measures of Cognitive Control: Shared Versus Distinct Abilities. *Assessment*, *28*(4), 1080–1096. doi:10.1177/1073191120965694

Spearman, C. (1904a). 'General Intelligence,' obectively determined and measured. *American Journal of Psychology*, *15*, 201–293.

Spearman, C. (1904b). 'General intelligence,' objectively determined and measured. *The American Journal of Psychology*, *15*(2), 201–293. doi:10.2307/1412107

Stevenson, N., Innes, R. J., Gronau, Q. F., Miletic, S., Heathcote, A., Forstmann, B., & Brown, S. (2024). Using group level factor models to resolve high dimensionality in model-based sampling. *PsyArXiv*. Retrieved from 10.31234/osf.io/pn3wv

Stevenson, N., Innes, R. J., Gronau, Q. F., Miletić, S., Heathcote, A., Forstmann, B. U., & Brown, S. D. (2024). Using group level factor models to resolve high dimensionality in model-based sampling. *Psychological Methods*. doi:10.1037/met0000618

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662.

Thurstone, L. L. (1938). *Primary Mental Abilities*. University of Chicago Press.

Townsend, J. T., & Ashby, F. G. (1982). Experimental test of contemporary mathematical models of visual letter recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 834–864.

Tucker, L. R. (1940). The role of correlated factors in factor analysis. *Psychometrika*, *5*(2), 141–152. doi:10.1007/BF02287872

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

Vadillo, M. A., Malejka, S., Lee, D. Y. H., Dienes, Z., & Shanks, D. R. (2022). Raising awareness about measurement error in research on unconscious mental processes. *Psychonomic Bulletin & Review*, *29*(1), 21–43. doi:10.3758/s13423-021-01923-y

Vandekerckhove, J. (2014). A cognitive latent variable model for the simultaneous analysis of behavioral and personality data. *Journal of Mathematical Psychology*, *60*, 58–71.

Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2011). Hierarchical diffusion models for

two-choice response time. *Psychological Methods*, *16*, 44–62.

Vanpaemel, W. (2010). Prior sensitivity in theory testing: An apologia for the Bayes factor. *Journal of Mathematical Psychology*, *54*, 491–498.

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.

Vickers, D., & Smith, P. (1985). Accumulator and random-walk models of psychophysical discrimination: A counter-evaluation. *Perception*, *14*, 471–497.

Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, *17*(2), 228–243. doi:10.1037/a0027127

Yang, S., & Rouder, J. N. (2025, April 4). *Assessing Two Common Priors of Covariance in Hierarchical Designs.* doi:10.31234/osf.io/jen65_v1

Table 1

Ground Truth Factor Loadings (ms)

	Factor 1	Factor 2
Task 1	30.00	7.00
Task 2	26.71	10.29
Task 3	23.43	13.57
Task 4	20.14	16.86
Task 5	16.86	20.14
Task 6	13.57	23.43
Task 7	10.29	26.71
Task 8	7.00	30.00

Table 2

Standardized Factor Loadings For Visual Illusions

Task	Nuisance Factors					Exploratory Factors	
	1	2	3	4	5	6	7
Brentano 1	-	-	-	-	0.242	0.572	-0.11
Brentano 2	-	-	-	-	-0.04	0.489	0.061
Ebbinghaus 1	-	-	-	0.632	-	0.239	0.253
Ebbinghaus 2	-	-	-	0.617	-	0.377	-0.19
Poggendorf 1	-	-	0.641	-	-	0.512	0.043
Poggendorf 2	-	-	0.589	-	-	0.482	0.185
Ponzo 1	-	0.835	-	-	-	0.363	-0.131
Ponzo 2	-	0.845	-	-	-	0.265	0.139
Zöllner 1	0.585	-	-	-	-	0.573	-0.005
Zöllner 2	0.548	-	-	-	-	0.619	-0.057

Table 3

Factor Loadings For EFA Model of Cognitive Control

Task	Factor 1 (ms)	Factor 2 (ms)	Factor 3 (ms)
Flanker	4	-8	8
Global/Local	36	-5	1
Negative Priming	0	5	0
Simon	1	-1	2
Stroop	31	8	-1
Task Switching	5	15	4

A

Participant	Task 1 Score	Task 2 Score	Task 3 Score	Task 4 Score
1	75	120	60	70
2	100	150	30	20
3	20	10	40	50
…	…	…	…	…

B

	Task 1	Task 2	Task 3	Task4
Task 1	1	.3	.2	.1
Task 2		1	.6	-.1
Task 3			1	.3
Task 4				1

C
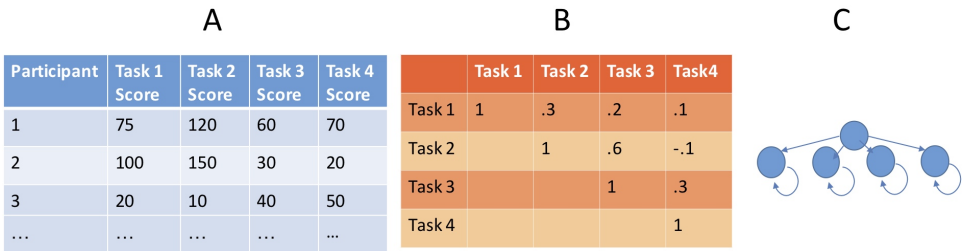
Figure 1. Usual data-analytic pipeline: A. The raw data are tabulated into individual scores by taking the mean across repeated obsercations. B. The correlation among these individual scores may be computed C. These correlations may be decomposed with structural equation models.
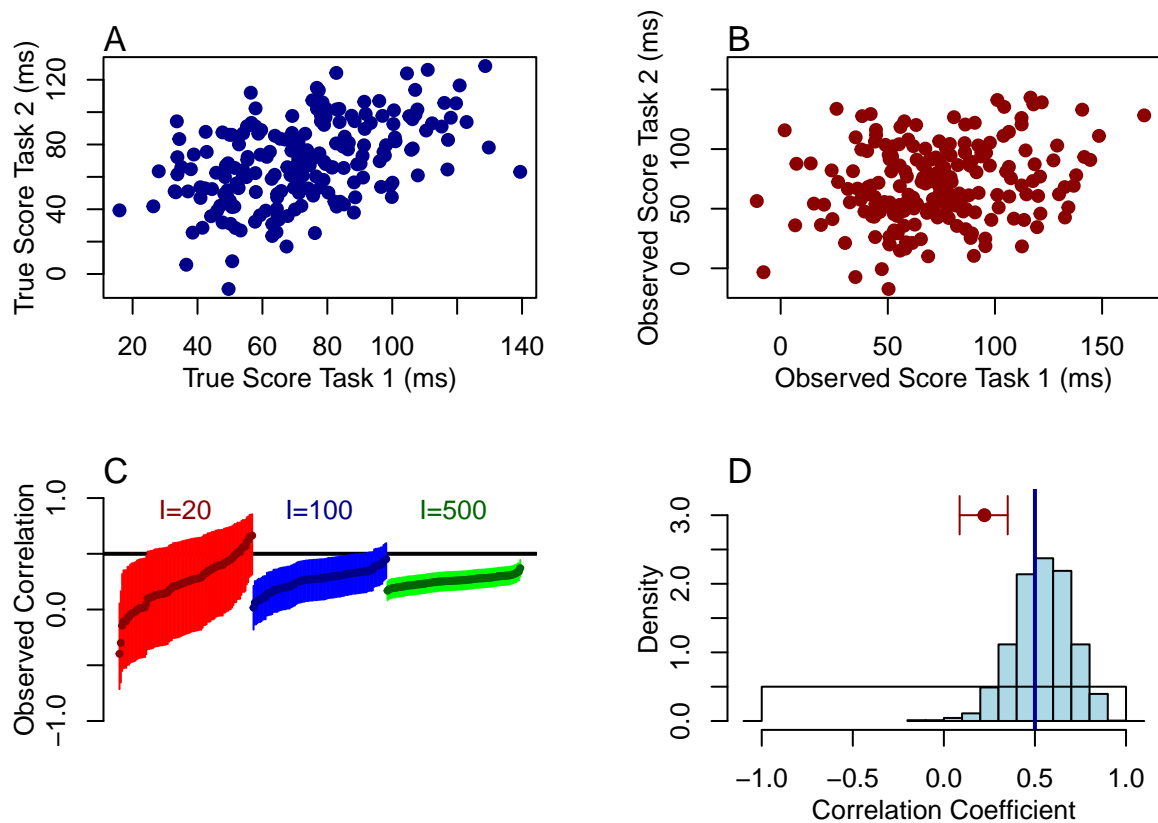
Figure 2. Effects of trial noise on correlations. A. Hypothetical true values are correlated at .5 in value. B. Adding trial noise perturbs points in the scatter plot resulting in a substantially lower correlation. C. Sample correlations with 95% confidence intervals across repeated simulations. As the number of individuals increase, the analyst becomes more sure in an attenuated answer. D. Posterior of correlation coefficient from a hierarchical model. The overall correlation values do not suffer from attenuation. Moreover, the uncertainty is large because it reflects trial noise as well as finite samples of individuals.
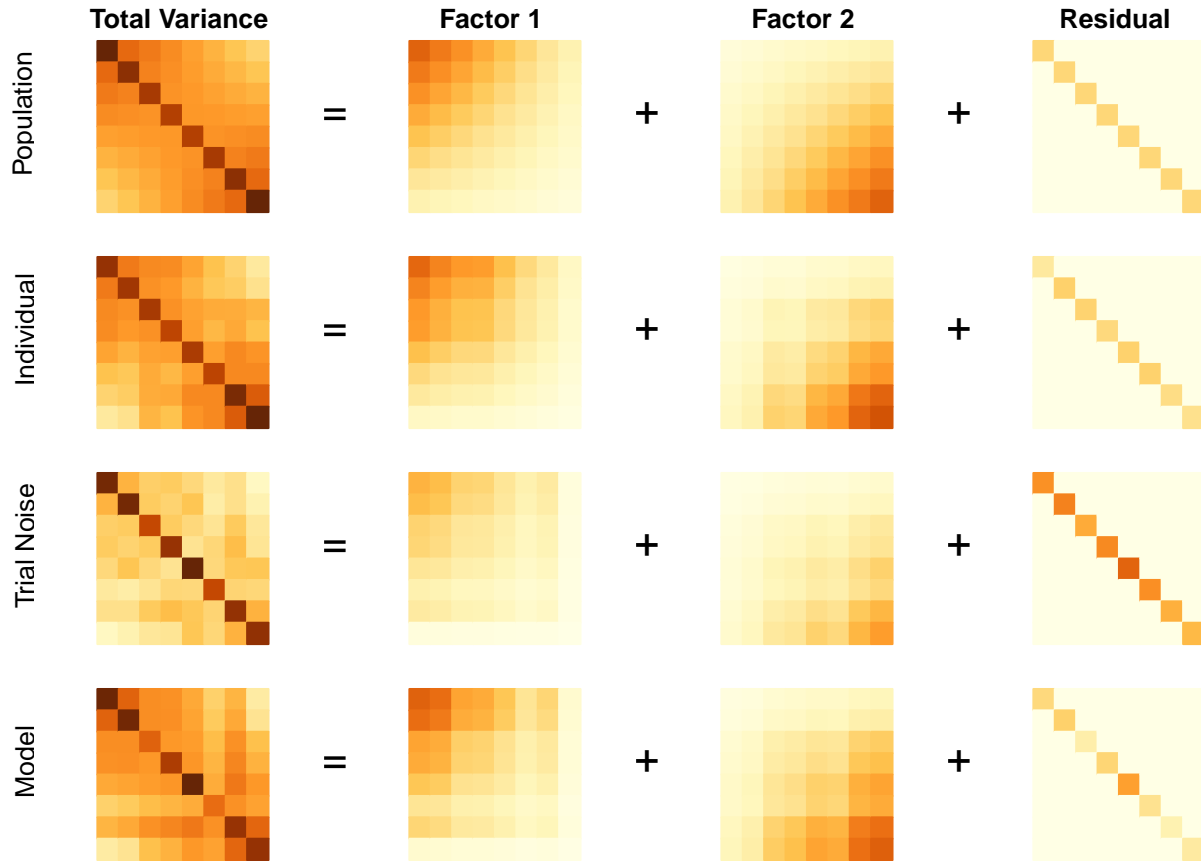
Figure 3. Factor decomposition of variance. The first column shows the overall variance, which is the sum of influences from two factors (second and third columns) and a residual (fourth column). The top row, "Population", shows the decomposition for true values across the population. The next row, "Individual" shows the effects of sampling 200 individuals. True values unperturbed by trial noise are decomposed. The next row, "Trial Noise" shows the decomposition from take-the-mean scores on observations. The difficulty here is that trial noise is manifest in the residual making factors appear less salient. The last row, "Model" shows the decomposition within the hierarchical model. Trial noise no longer inflates the residuals making factors appropriately salient.
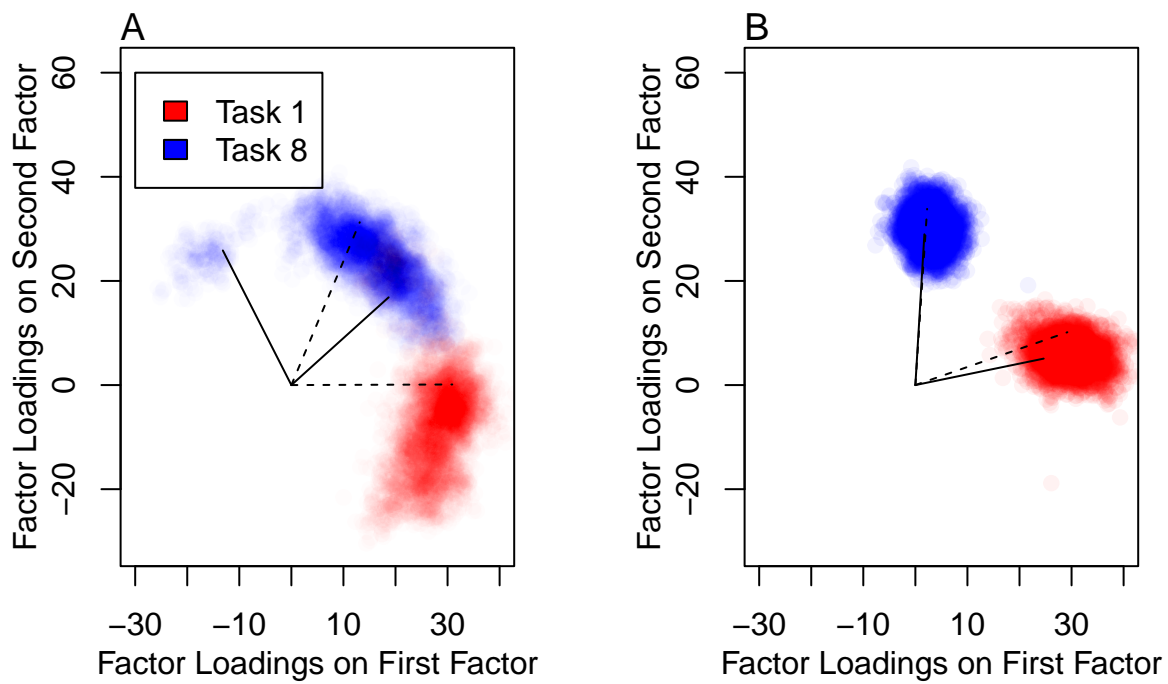
Figure 4. Rotations may be problematic in Bayesian MCMC analysis without post-rotational alignment. A. Factor loadings on two dimensions for Tasks 1 and 5. Multimodal posteriors result because each iteration corresponds to a different rotation.
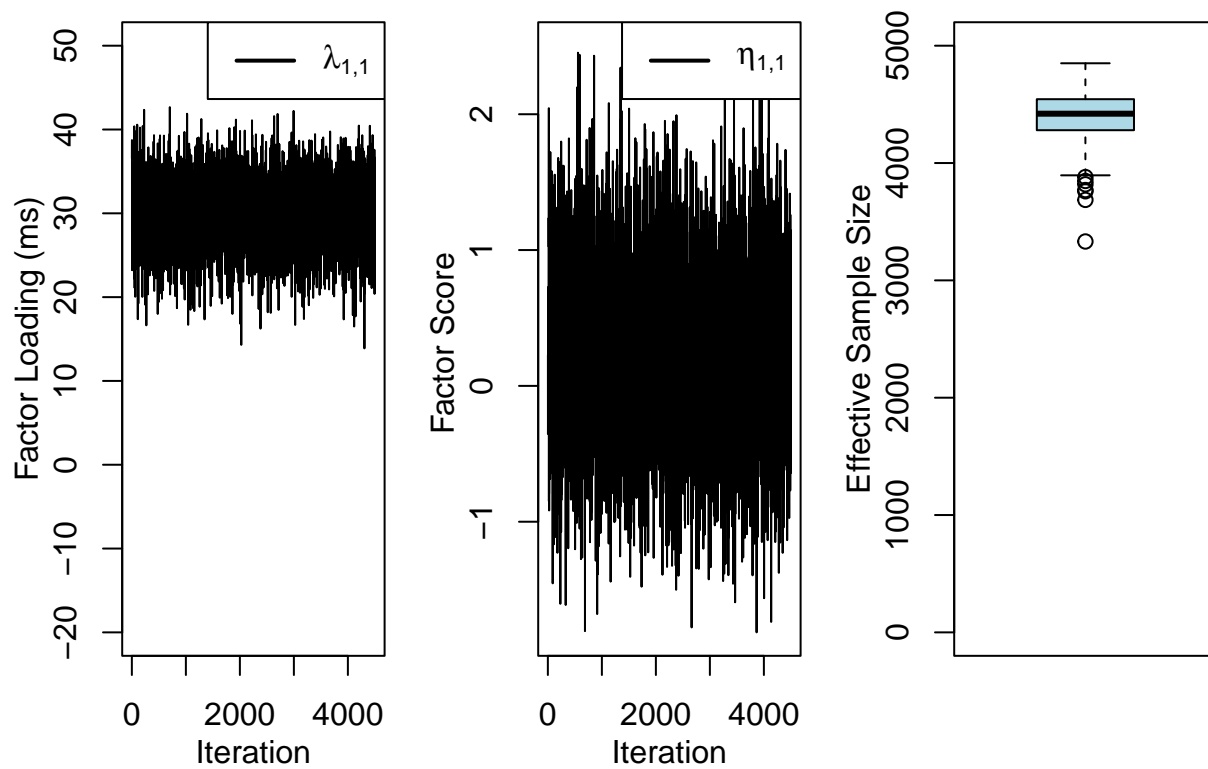
Figure 5. Hierarchical model mixes well when implemented in JAGS. Left and Center: Chains for select parameters. Right: Boxplot of effective sample sizes for all parameters for a chain with 5000 iterations.
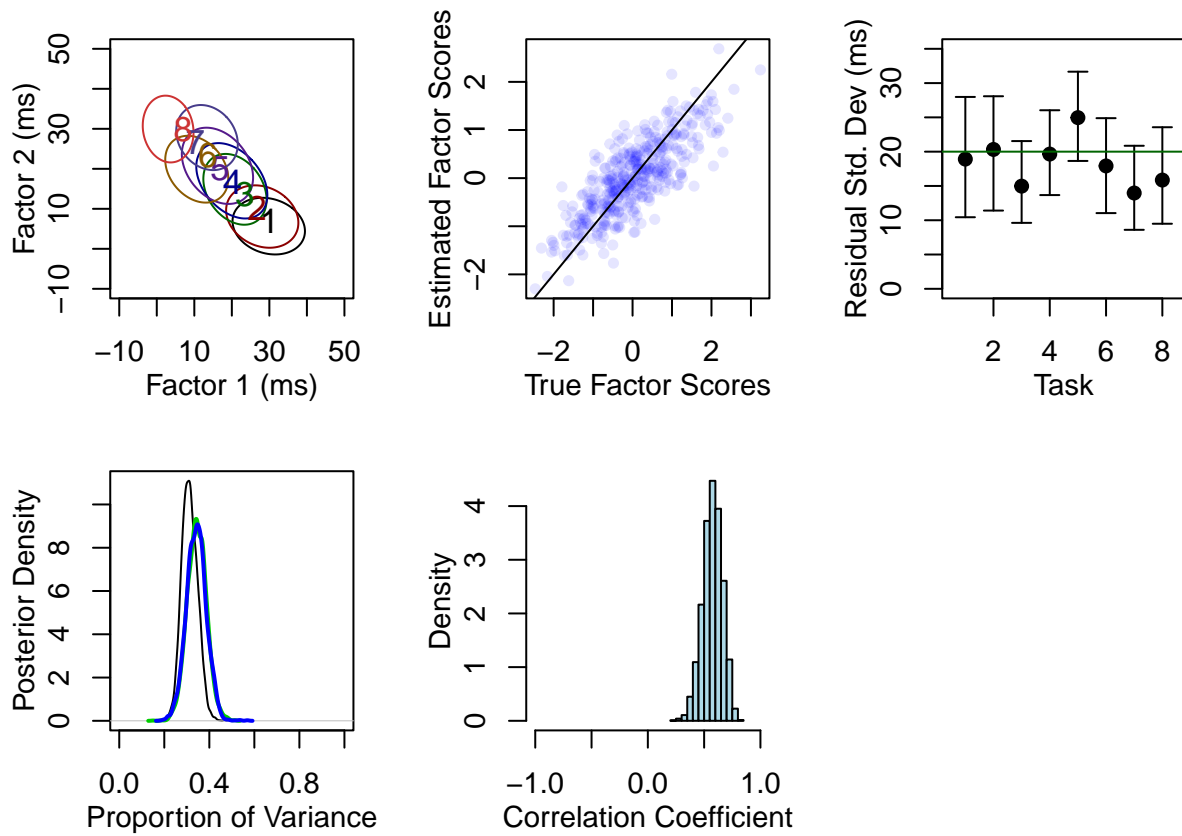
Figure 6. Parameter recovery. A. Factor loadings show the gradual tradeoff in Table ef{tab.exampleTrueLoads}. Ellipses, the 95% credible intervals, show uncertainty in the loadings. B. Recovery of factor scores shows an expected degree of regularization. C. Recovery of residual variation. D. Posterior uncertainty on proportion of variance accounted for by each factor and the residual. E. Posterior on a correlation coefficient reflects uncertainty from individuals and trials.
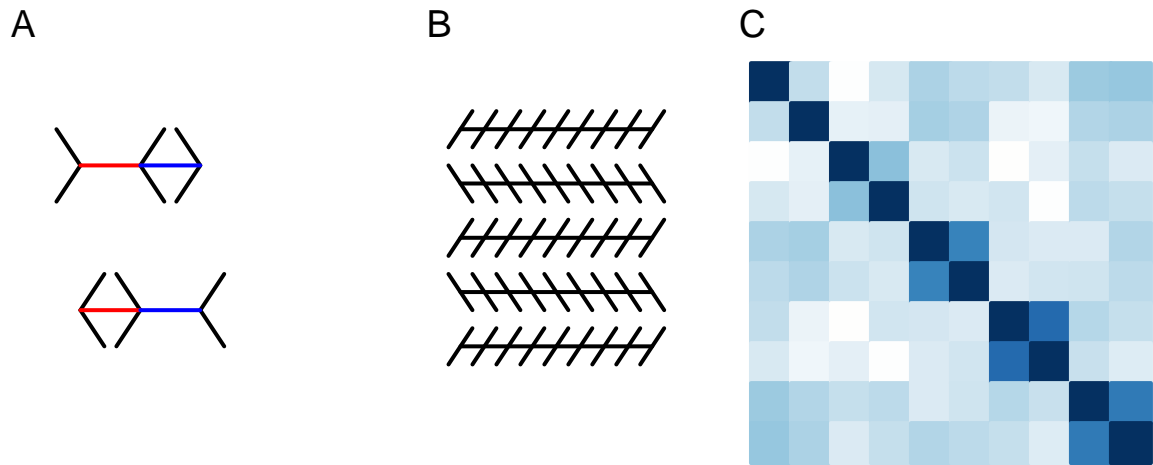
Figure 7. A. Two versions of the Brentano Illusion. The blue and red segments are of equal length in each version. B. The Zöllner Illusion. The horizontal lines are parallel. C. Observed correlation matrix for 2 versions of 5 illusions. The on-diagonal blocks correspond to two versions of the same illusion.
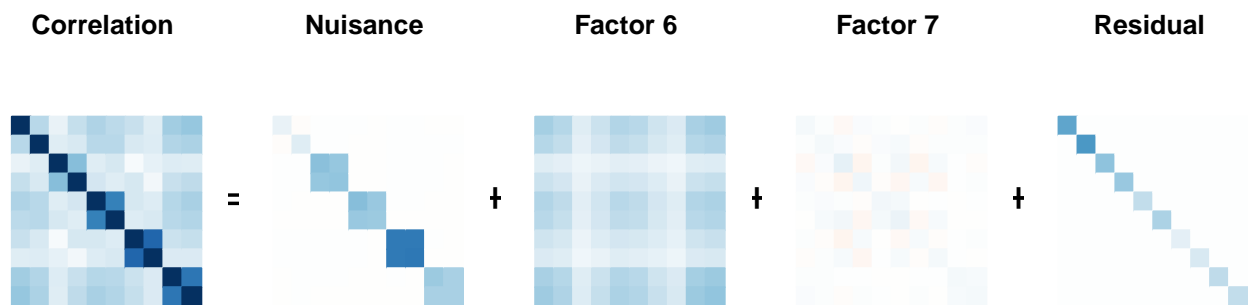
Figure 8. Factor decomposition of modeled correlation for Mehrvarz et al's (2025) visual-illusion battery.
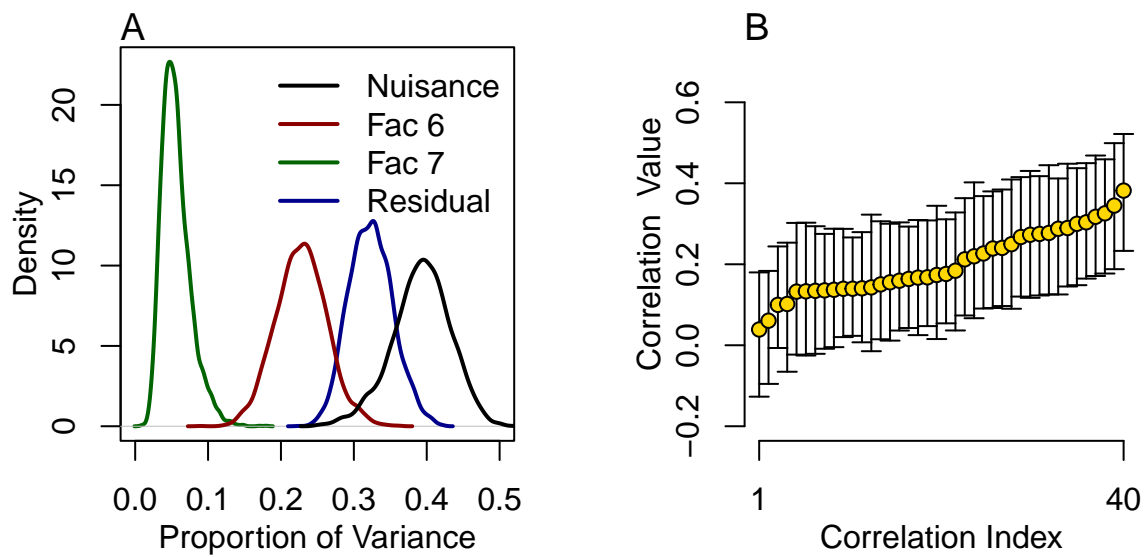
Figure 9. A. Proportion of variance accounted for components of the factor decomposition. B. Between-illusion correlations with 95% credible intervals. These values are small and well localized.
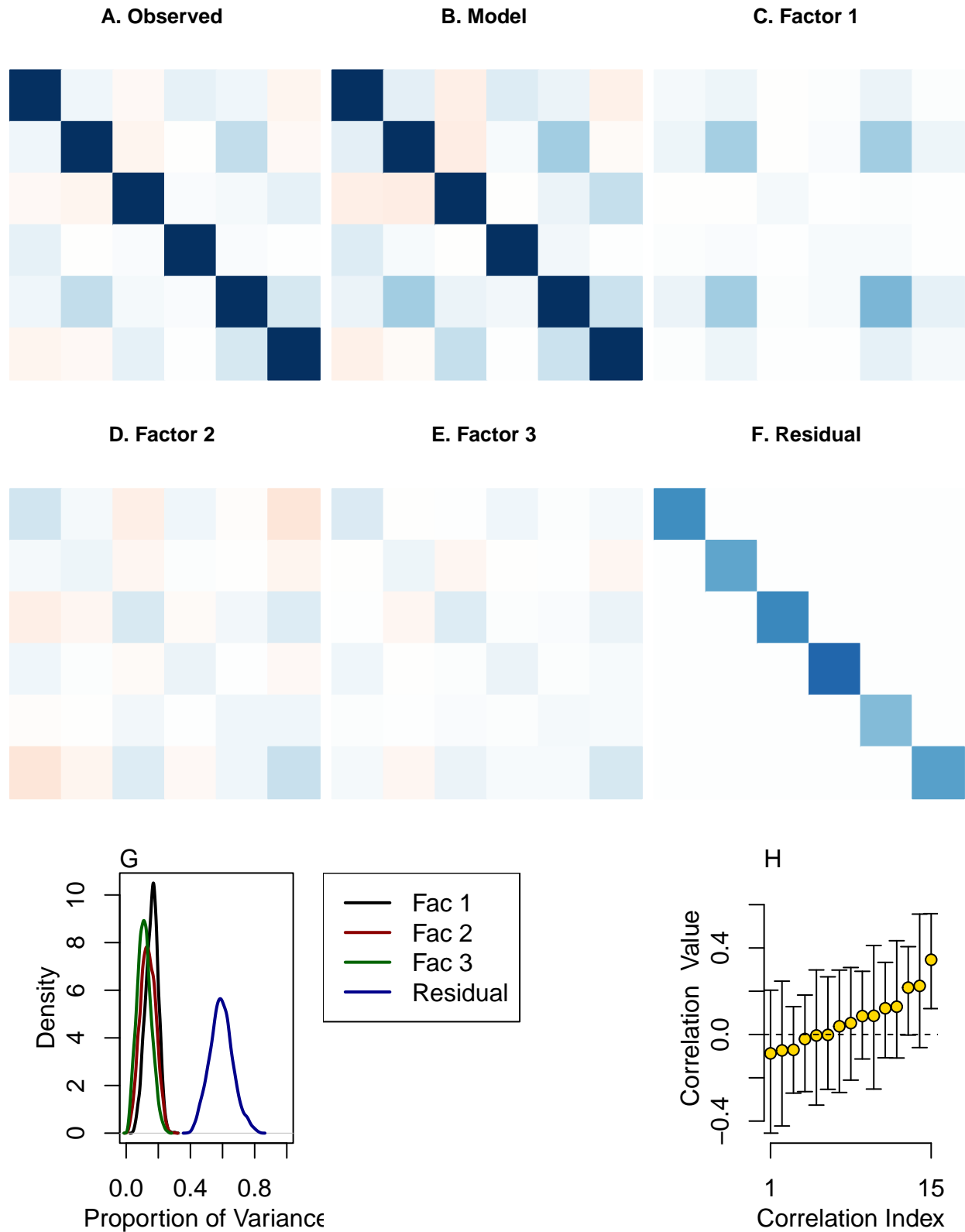
Figure 10. Results from cognitive control example. A. Observed correlation matrix. B. Model-based correlation matrix. C-F. Factor decomposition of the model-based correlation. G. Posterior proportion of variance accounted for by each component. H. Model-based correlations with 95% CIs. Correlations are difficult to localize because of excessive trial noise.
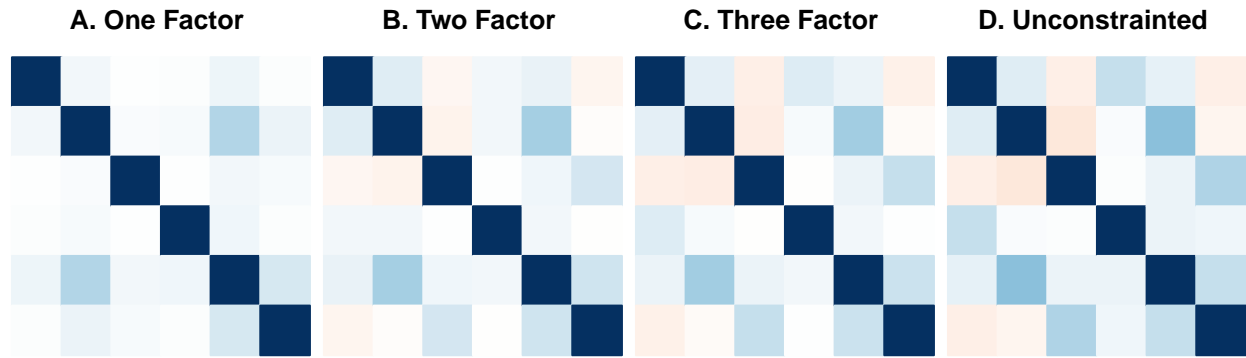
Figure 11. Factor approximation of a correlation matrix. In the hierarchical model, the matrix being approximated may be estimated with an unconstrained prior on covariance.
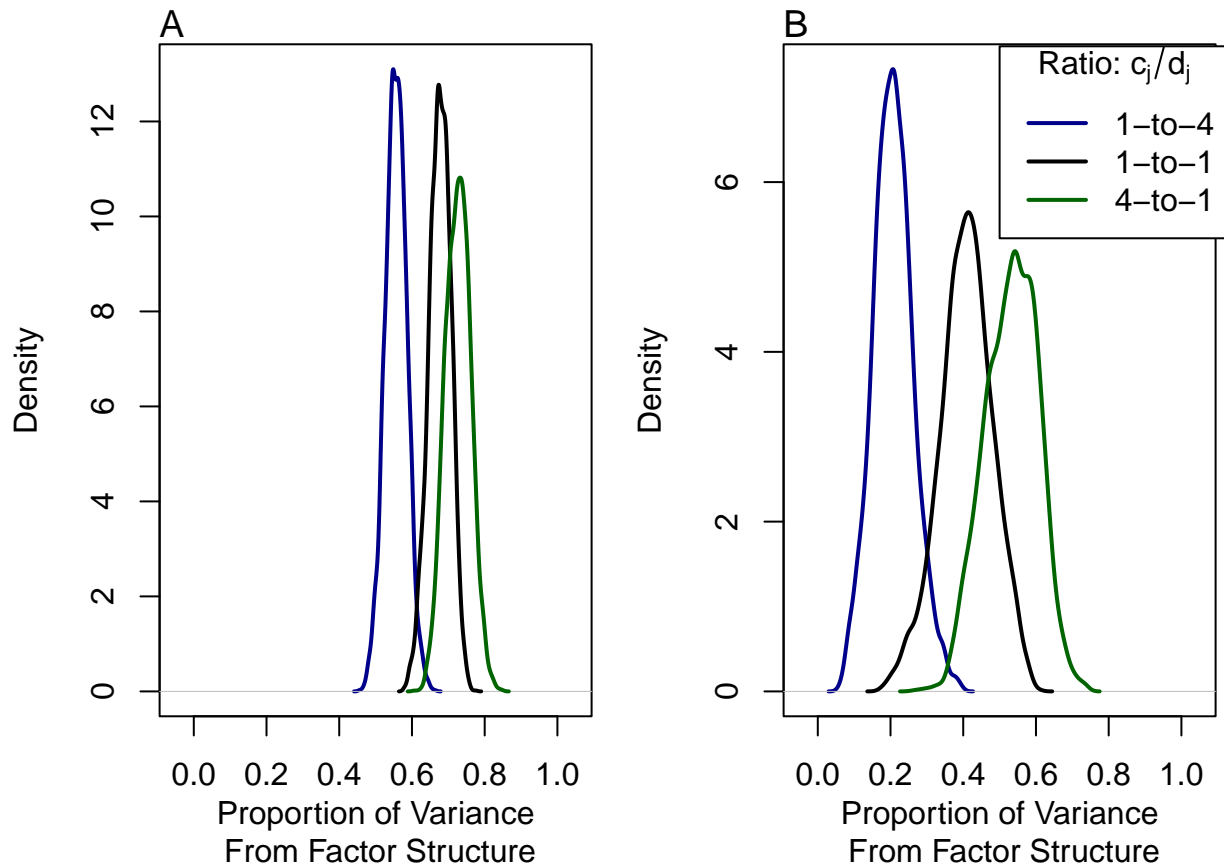

Figure 12. Effect of prior settings on loadings (c_j) and residuals (d_j) on the proportion of variance accounted for by the factor structure. The effect enters as the ratio c_j/d_j. A. Visual illusion data set. The three curves correspond to ratios 1-to-4, 1-to-1, and 4-to-1. B. Cognitive control data set. The three curves correspond to the same three ratios.