

# A cognitive latent variable model for the simultaneous analysis of behavioral and personality data

Joachim Vandekerckhove\*  
University of California, Irvine

I describe a *cognitive latent variable model*, a combination of a cognitive model and a latent variable model that can be used to aggregate information regarding cognitive parameters across participants and tasks. The model is ideally suited for uncovering relationships between latent task abilities as they are expressed in experimental paradigms, but can also be used as data fusion tools to connect latent abilities with external covariates from entirely different data sources. An example application deals with the structure of cognitive abilities underlying an executive functioning task and its relation to personality traits.

**Keywords:** individual differences; cognitive model; latent variable; factor analysis; data fusion; diffusion model

## Introduction

### Cognitive psychometrics

*Cognitive psychometrics* is the term coined by Batchelder (1998) to describe the application of cognitive process models as assessment tools, or, more fundamentally, to apply the psychometrics of individual differences to cognitive process parameters. The practice of combining cognitive measurement models with individual variability, implemented as statistical random effects, serves in the first place to adapt cognitive models to the reality of randomly sampled, noninterchangeable participants (e.g., Batchelder, 2007). As has been pointed out by Estes (1956, 2002), Hamaker (2012), and Heathcote, Brown, and Mewhort (2000), averaging artefacts can lead to biased estimates and errors in inference. More

than that, however, the assumption that an individual's process parameters are in fact a random draw from some superordinate population distribution introduces a crucial new aspect to cognitive modeling: The idea that there might be formal structure to be derived from the individual differences researchers often observe among participants' cognitive model parameters.

Structured individual differences are a critical concept in certain fields of cognitive science. For example, intelligence research is dominated by studies in which individuals are assessed on a variety of tasks, and it is typically observed that participants who score high on one task also score high on other tasks (e.g., Kamphaus, Petoskey, & Morgan, 1997). This covariance is taken to imply that there exists a small set of person-specific abilities that jointly give rise to correlated behavior on the larger set of tasks (a "positive manifold"). An identical approach is often taken in fields such as working memory (e.g., Oberauer, Süß, Schulze, Wilhelm, & Wittmann, 2000) or executive functioning (e.g., Miyake et al., 2000), where unobserved factors supporting stable differences across individuals are inferred from the correlational pattern between multiple basic tasks. This type of data analysis is widely known as latent variable modeling (Bartholomew, Knott, & Moustaki, 2011; Skrondal & Rabe-Hesketh, 2004).

Importantly, the interpretability and usefulness of the results of such analyses depend on the interpretability of the quantities measured in the basic tasks. If each score in a given set of tasks can reasonably be thought to tap intelli-

---

This project is supported by NSF grant #1230118 from the Methods, Measurements, and Statistics panel. The author is grateful to Roger Ratcliff, Francis Tuerlinckx, and Michael Lee for fruitful discussions in earlier stages of this project; to Klaus Oberauer, Yves Rosseel, Michael Nunez, Zita Oravecz, Michael Pratte, and an anonymous referee for helpful comments on an earlier draft; to Dominik Wabersich for assistance in the implementation of the analyses; and especially to Madeline Pe, Peter Koval, and Peter Kuppens for providing the data set used in the demonstration.

\* Corresponding author. E-mail: joachim@uci.edu. Phone: (949) 824-5958. Address: University of California, Irvine; 2324 SBSG; Irvine, CA 92617-5100.

gence, then it is valid to conclude that the inferred latent factors relate to intelligence as well. If, on the other hand, scores in the basic tasks are nonlinear amalgams of more elementary variables, interpretation of the latent factors is complicated. Cognitive models serve to decompose such complex data into interpretable parameters. The modeling strategy proposed in this paper involves—within a single model—a latent variable structure built on top of a cognitive process model, to allow inference of latent variables that have cognitive interpretations.

### A qualitatively different type of conclusion

When latent variable models are combined with cognitive models to form a *cognitive latent variable model* (CLVM), this affords a qualitatively different type of conclusion from either classical psychometrics or classical cognitive modeling. For example, using a cognitive model with a parameter interpreted as *speed of information processing* (e.g., the drift rate in a diffusion model Ratcliff, 1978), a CLVM permits inferences about unobserved variables that contribute to the total rate of information processing in a particular task. A conventional psychometric model would not permit such process-based conclusions, whereas a conventional cognitive model would not be equipped to infer higher-order latent properties.

Combining cognitive models with latent variable models allows us to bridge the gap between experimental and individual-differences research—a long-standing issue in psychology since Cronbach’s (1957) lament that the science is split across two disparate disciplines, reiterated more recently by Borsboom (2006). It is the aim of the present paper to present an example of a CLVM, a formal model that extends the logic of cognitive psychometrics to include latent variable structures.

The structure of the paper is as follows. The next section will introduce two components of the CLVM: the diffusion model as a cognitive model of choice response time data and the factor analysis model as a measurement model to tie multiple tasks together. This section will also introduce some required notation. The section after that will focus on properties of the integrative CLVM. After that, a short section will be devoted to the relevant details of Bayesian inference and model selection. Finally, a section will provide detail regarding the application of the CLVM in the field of emotion psychology.

### Diffusion models for two-choice RT

The data level of this CLVM consists of a probabilistic representation of data as they are predicted by a particular cognitive model—the sampling scheme of the data. The cognitive model used here is a simplified diffusion model for two-choice RT (Stone, 1960), which has been very popular in cognitive science (see Wagenmakers, 2009, for an overview of recent applications and advances), with applications ranging from memory (Ratcliff, 1978) and low-level perception (Ratcliff & Rouder, 1998) to semantic cognition (Vandekerckhove, Verheyen, & Tuerlinckx, 2010) and emotion psychology (Pe, Vandekerckhove, & Kuppens, 2013; White, Ratcliff, Vasey, & McKoon, 2009). The diffusion model is based on the principle of *sequential accumulation of information*—it assumes that a decision making system samples small units of information, sequentially over time, from whatever stimulus to which it was exposed. These sampled units of evidence are aggregated with information already accumulated. After each accretion step, the system evaluates whether the total amount of evidence warrants the making of a decision. If so, the process ends and a response is executed. This accumulation process is the fundamental assumption—the “central dogma”—of a broad and highly successful class of sequential sampling models for RT.

More specifically, the process assumptions of the diffusion model are that a single evidence counter accumulates towards one of two decision boundaries, with a starting point that may be closer to one boundary than the other. Figure 1 illustrates the process. Given the freedom of two decision bounds, the model can account for two distinct types of bias in the response process. In addition to biased processing of information (which is reflected in the average rate of evidence accumulation, a parameter called the *drift rate*,  $\delta$ ), the diffusion model allows for an a-priori bias that is prior to and independent of the information accumulation process (here parameterized as a proportion, so that a *bias*  $\beta = 0.5$  implies a-priori indifference). The distance between the decision bounds (known as the *boundary separation*  $\alpha$ ) performs a separate, interesting task in the diffusion process. Bounds that are close together lead to fast decisions that are largely independent from the information contained in the stimulus (i.e., close to chance level), whereas distant bounds lead to slow response processes whose outcome is mostly determined by the direction of the accumulation process (i.e., if  $\delta$  is positive and  $\alpha$  is high, the upper boundary is likely to be hit). This parameter hence captures the well-known *speed-accuracy trade-off*. The fourth and final parameter of the diffusion model is the *nondecision time*  $\tau$ . This shift parameter

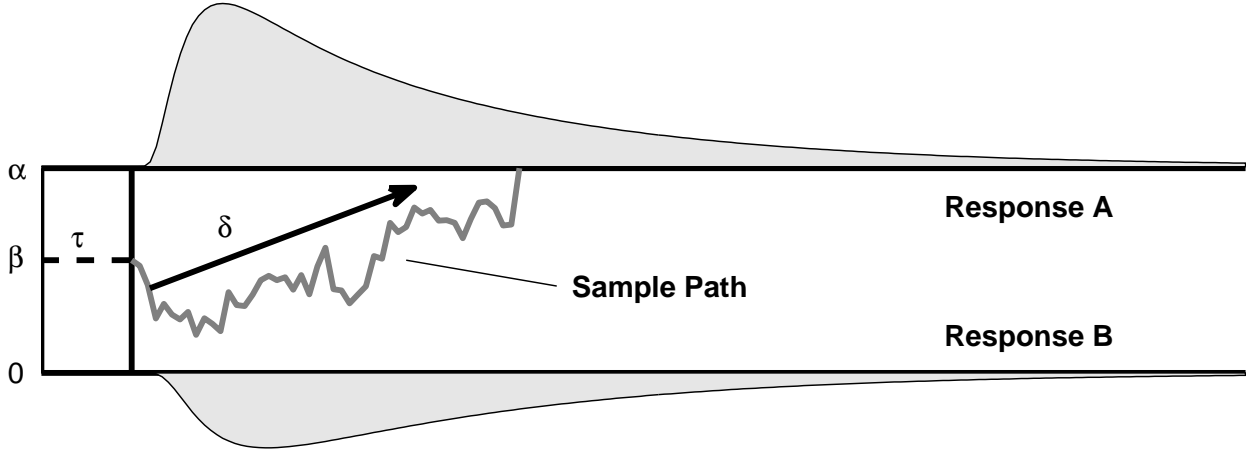


Figure 1. An illustration of the Wiener diffusion model. Evidence is accumulated over the (horizontal) time dimension, at an average rate of  $\delta$ . The decision process terminates if the evidence value reaches 0 or  $\alpha$ , and the amount of evidence at the onset of the trial is given by  $\alpha\beta$ . The nondesideration time  $\tau$  reflects independent additive processes such as stimulus encoding and response execution. Equation 1 describes the reaction time distributions that follow from these model assumptions. Figure adapted with permission from Vandekerckhove (2009).

determines the leading edge of the latency distribution, and is typically interpreted as the sum duration of all non-decision processes (and it is additionally assumed that these processes are independent of and serial to the decision process).

The PDF of the Wiener diffusion model is bivariate (with one dimension for the latency and one for the binary choice); its analytical form also contains an infinite sum and the latency distribution can therefore at best be approximated:

$$\begin{cases} p(t, x = 0 | \alpha, \beta, \tau, \delta) &= \frac{\pi}{\alpha^2} e^{-\frac{1}{2}(2\alpha\beta\delta + \delta^2(t-\tau))} \\ &\times \sum_{k=1}^{+\infty} \left[ k \sin(\pi k \beta) e^{-\frac{1}{2} \frac{k^2 \pi^2}{\alpha^2} (t-\tau)} \right] \\ p(t, x = 1 | \alpha, \beta, \tau, \delta) &= p(t, x = 0 | \alpha, 1 - \beta, \tau, -\delta) \end{cases} \quad (1)$$

Fortunately, efficient methods for the computation of the Wiener diffusion model density and distribution functions exist (Blurton, Kesselmeier, & Gondan, 2012; Navarro & Fuss, 2009, for the CDF and PDF, respectively), making it a highly tractable model. Equation 1 lacks a *diffusion coefficient* parameter, which is sometimes used to scale the evidence dimension (and typically denoted  $s$ ); the coefficient does not appear because it will be set to 1 in all applications, and it cancels out everywhere.

Figure 2 shows a graphical model representation of an un-

biased Wiener diffusion model for a data set where  $P$  participants do a task with  $T$  conditions and  $I$  trials in each condition. For conciseness,  $y$  denotes a choice RT pair  $(t, x)$ . The equations to the right of the diagram list the distributional assumptions of the model, including some example priors.

It is important to note that this data model can serve a dual purpose for researchers in psychology. On the one hand, researchers can decide to buy in to the assumptions of the model—taking the process as given and drawing conclusions that may hinge on the accuracy of these assumptions. For this particular cognitive model, the literature contains reports of experimental manipulations that selectively affect model parameters, lending some credibility to the process assumptions (e.g., Voss, Rothermund, & Voss, 2004). However, the model would remain useful even if one is unwilling to buy in to the exact process—by taking the model as a convenient data level that captures the shape of the data and serves strictly as a parsimonious description.

### Latent predictors—the third building block

De Boeck and Wilson (2004), in providing their anatomy of explanatory models, identify the three building blocks that can be used in the construction of models whose aim is to explain observed variance.

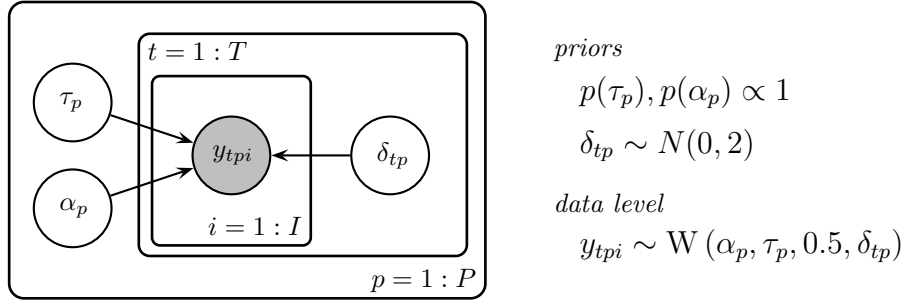


Figure 2. A graphical model representation of a Wiener diffusion model without a-priori response bias (i.e.,  $\beta = 0.5$ ). In this representation, to-be-estimated variables are shown as plain circles and data have shading. Arrows indicate “is-parent-of” relationships, with parent nodes determining the distribution of child nodes. The  $p$ -plate indicates independent repetitions over  $P$  participants,  $t$  over  $T$  conditions, and  $i$  over  $I$  trials.

The first building block is *random effects*, in which a set of model parameters are assumed to be draws from a common superordinate distribution. Random effects can be made hierarchical, so that the parameters of the superordinate distribution themselves are draws from a higher-level distribution, or they can be crossed, so that some parameters are combinations of outcomes of draws from multiple distributions. The random-effects assumption has many advantages, including the possibility of estimating population-level parameters (e.g., a person-specific parameter might be a draw from a group-level distribution, whose parameters will be descriptive of the group). Additionally, random sampling from a larger population is often a more truthful description of how participants (and, sometimes, items or stimuli) are selected. Random effects have been applied in item response models for decades, but have only relatively recently found their way into cognitive modeling (see, e.g., Rouder, Sun, Speckman, Lu, & Zhou, 2003).

The second building block is *manifest predictors*, in which external covariates are used to reduce unexplained variance in parameters. Several straightforward methods for the inclusion of manifest predictors exist; One can imagine a linear structure, where some person-specific parameter  $\theta_{(p)}$  is no longer estimated, but replaced by the linear function  $\beta_0 + \beta_1 x_{(p)}$ , where  $x_{(p)}$  is person  $p$ 's score on some external measure  $X$ . If  $X$  is continuous, this amounts to a linear regression; if it is categorical it is an ANOVA-style structure. Some caution is in order in the construction of such linear structures in order to respect the natural domain of the to-be-explained parameter. For example, if  $\theta$  is a proportion, care should be taken to constrain the explanatory structure to predicting only values in the  $[0 - 1]$  range. A standard method of enforcing such constraints is through the appli-

cation of a nonlinear link function. To constrain a parameter to the  $[0 - 1]$  range, a logistic function is one of several possible link functions, so that the regression structure becomes  $\theta_{(p)} = \{1 + \exp[-(\beta_0 + \beta_1 x_{(p)})]\}^{-1}$ . Manifest predictors for process model parameters were used by, among others, Oravecz, Tuerlinckx, and Vandekerckhove (2009) and Vandekerckhove et al. (2010).

The third building block is *latent predictors*, in which the explanatory covariates are not observed, but are inferred from the correlational structure between (for example) performance on tasks, conditions, or items (across participants) or participants (across tasks, conditions, or items). More precisely, latent variables are *at least partially unobserved variables that jointly explain the covariance between a set of observed variables* (this is called the “local independence” definition in Bollen, 2002). While latent predictors are exceedingly common in, for example, personality psychology and aptitude research, they have almost never been applied to cognitive model parameters in the manner proposed in the next section. One application of a continuous unobserved predictor to cognitive model parameters is seen in Pe, Vandekerckhove, and Kuppens (2013), whose diffusion model included a person-specific drift-gain parameter that was fully unobserved but tied together cognitive parameters with external covariates.<sup>1</sup>

Latent variables (LVs) can be used to construct explanatory structures for cognitive model parameters. With this,

<sup>1</sup>A slightly different latent variable construct, *latent class assignment*—in which the latent variable is binary as opposed to continuous—has been used in some recent publications (e.g., Bartlema, Lee, Wetzels, & Vanpaemel, in press; Lee & Wetzels, 2010; Lee, 2008; Vandekerckhove, Tuerlinckx, & Lee, 2008).

the present paper completes the set of explanatory structures available for cognitive models. This section contains a brief description of LV models as they are used classically: to explain correlations between data points, rather than model parameters. The subsequent section will then transition into the exact model that will be applied to the example data set.

### Confirmatory factor models

One of the main goals of confirmatory factor analysis (CFA), and the way it will be used here, is to determine *construct validity*. By examining interrelations between manifest variables (MVs) and explaining the interrelations in terms of a smaller number of unobserved underlying LVs, CFA enables researchers to determine *convergent validity* (i.e., confirm that MVs that should measure the same construct do so) and *discriminant validity* (i.e., confirm that MVs that should not share an LV don't).

In CFA, researchers posit one or a handful of possible low-dimensional underlying structures that jointly explain the pattern of covariances between a larger number of MVs. For example, if a test of scholastic aptitude consists of six subtests (these are the MVs), three of which are tests of mathematical ability and three of language ability, a reasonable underlying model might involve only two LVs. If  $P$  students take  $T$  subtests and their scores are collected in the  $T$ -by- $P$  matrix  $Y$ , then a CFA model with  $D$  underlying factors requires  $\Lambda$ , a  $T$ -by- $D$  matrix of loadings, and  $\Phi$ , a  $D$ -by- $P$  matrix of person-specific factor scores. A typical representation of the factor model is then:  $Y = \Lambda \times \Phi + E$ , where  $E$  is a  $T$ -by- $P$  matrix of independent, zero-centered, normally distributed errors.

As it is written here, the factor model is unidentified—multiplying any row of  $\Phi$  with any real number and dividing the corresponding column of  $\Lambda$  by that number would yield identical model predictions. Hence,  $\Phi$  and/or  $\Lambda$  need to be constrained. As a result, different factor models are distinguished not only by their dimensionality  $D$ , but also by the pattern of constraints placed upon the elements of  $\Lambda$  and  $\Phi$ . For ease of implementation, the present application will constrain only elements of  $\Lambda$ , (a “ $\lambda$ -only” constraint) but it will be demonstrated how other identification constraints can be obtained by post-hoc transformations of the parameter estimates (for interpretation purposes, the “unit factor variance” constraint described below will turn out to be useful).

Constraining the loadings matrix  $\Lambda$  is tantamount to deciding which MV is allowed to be related to which LV. One possible approach is to limit each MV to load on exactly one LV—a common CFA assumption known as *simple structure*

or a *congeneric factor model*. Since simple structure by itself does not guarantee identification, a further possible constraint is to fix one loading per LV to a particular value (most commonly 1, but in principle any nonzero real value). A special case of simple structure is the one-factor model, in which all scores across tasks are scaled versions of one another, with  $\Lambda = (1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6)^T$ .

Because of the confirmatory nature of CFA, it is recommended that researchers have a strong theory underlying their factorial assumptions before analyzing data (McArdle, 2011; Williams, 1995).

In order to change the identification constraints, simple transformations of the parameter estimates can be performed. For example, to obtain the more conventional constraint of unit variance of the factor scores belonging to  $P$  participants:

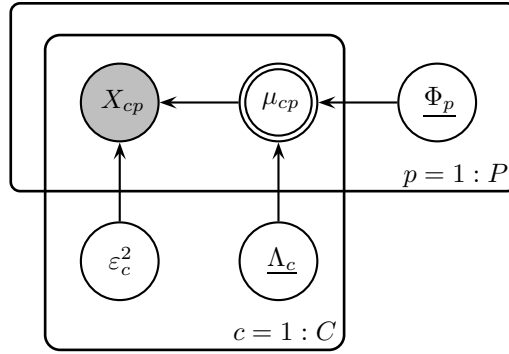
$$\forall f : \sigma_{(f)}^2 = \frac{1}{P-1} \sum_{p=1}^P (\phi_{(f,p)} - \bar{\phi}_{(f,\cdot)})^2 := 1,$$

it suffices to transform as follows:  $\phi_{(f,p)} = \hat{\phi}_{(f,p)}/\sigma_{(f)}$  and  $\lambda_{(t,f)} = \hat{\lambda}_{(t,f)}/\sigma_{(f)}$ , where the hatted parameters are the estimates under the initial (pragmatic)  $\lambda$ -only constraints and  $t$ ,  $f$ , and  $p$  index tasks, factors, and participants, respectively.  $\bar{\phi}_{(f,\cdot)}$  is the across-participant mean score on factor  $f$ . Regarding these transformations between identification schemes, it should be noted that (a) prior distributions, especially informative ones, for the affected parameters must be carefully defined, so that they do not convey spurious information after the transformation, and (b) throughout this paper, the  $\lambda$ -only constraint will be used to describe models (as in Fig 3), but the unit factor variance constraint will be used to interpret results.

Figure 3 shows a graphical model representation of a LV for  $C$  independent measures. Vector-valued nodes have as many elements as there are factors in the LV solution. Constraints are not indicated.

### Exploratory factor models

Exploratory factor models (EFAs) are identical to CFAs in their mathematical formulation, but allow for more freedom in the loadings matrix and so require much less theoretical commitment from the researcher. Typically, an EFA will have as many free parameters as possible while maintaining an identified model. The minimal requirements for identification of an EFA are nontrivial (see, e.g., Loken, 2005); one example of a minimally identified structure with  $D(D+1)/2$  values fixed is:



priors

$$\underline{\Phi}_p \sim \text{MVN}(\mathbf{0}, \mathbf{I}_p)$$

$$\underline{\Lambda}_c \sim \text{MVN}(\mathbf{0}, \mathbf{I}_f)$$

$$(1/\varepsilon_c^2) \sim \Gamma(0.1, 0.1)$$

measurement level

$$\mu_{cp} = \underline{\Lambda}_c \times \underline{\Phi}_p$$

data level

$$X_{cp} \sim \text{N}(\mu_{cp}, \varepsilon_c^2)$$

Figure 3. A graphical model representation of a latent variable model. In addition to the conventions of the previous model, this graph contains deterministic nodes (double edges) and vector-valued nodes (underlined). The  $c$ -plate indicates different independent measures and  $\times$  indicates the inner product.

$$\Lambda = \begin{pmatrix} 1 & 0 & 0 \\ \lambda_{(2,1)} & 1 & 0 \\ \lambda_{(3,1)} & \lambda_{(3,2)} & 1 \\ \lambda_{(4,1)} & \lambda_{(4,2)} & \lambda_{(4,3)} \\ \lambda_{(5,1)} & \lambda_{(5,2)} & \lambda_{(5,3)} \\ \lambda_{(6,1)} & \lambda_{(6,2)} & \lambda_{(6,3)} \end{pmatrix},$$

which considers the same hypothetical data set as before with six MVs. This EFA loadings matrix allow almost every MV to load on all LVs; the constraint is satisfied if the first MV is supported by exactly one LV, the second MV is supported by exactly two LVs, and so on until the remaining MVs are supported by all LVs. Jöreskog (1969) and Loken (2005) discuss and review further sufficient requirements.

Beyond the issue of factor identification, there is an issue of rotation invariance: Factor models are only identified up to a rotation of the factors. To obtain a unique rotation, a common strategy is the one described by Geweke and Zhou (1996), in which the loadings matrix has an upper triangle of zeros, the diagonal elements are constrained to be positive, and the factor scores are constrained to have unit variance. These constraints match exactly the ones described in the previous paragraph as the “unit variance constraint” and the “ $\lambda$ -only” constraint is equivalent.

Finally, no single latent variable model is fully exploratory. For instance, the example requires the researcher to commit to a three-dimensional latent structure, while a truly exploratory analysis would consider all seven possible dimensionalities. EFA therefore naturally takes on a model selection component.

## Discussion

This section contained a very brief overview of the most basic principles of a typical case of latent variable modeling, factor analysis, in which a number of manifest variables are considered as linear combinations of underlying, unobserved, latent variables. The weights of the linear combinations, called *loadings*, are at the center of the method of achieving model identification used in this paper, and the choices that are made to ensure identification also determine the degree to which a model is confirmatory versus exploratory.

The person-specific values of the LVs, called *factor scores* are in turn critical to the final interpretation of the model results. These scores express the degree to which a participant possesses the unobserved quality, and a participant’s score directly affects their performance in all tasks bound to that LV.

In the next section, it will be argued that a latent variable model applied not to raw data, but to *parameters of cognitive models*, is a feasible approach with practical appeal.

## The cognitive latent variable model

### Rationale

Recognizing that there exist two independent traditions with a wealth of interesting model constructs, one can combine elements from cognitive modeling and latent variable models into a new type of quantitative model. This CLVM has two distinct components. Firstly, the *data level* of the model is defined as the predicted distribution of the data,

given all the relevant parameters for a particular data point (i.e., it is the fully marginalized likelihood of the model). In the present application, the data level is a diffusion model for some data points and a normal distribution for others. The *measurement level* of the model is a set of linear equations that relate parameters at different conditions, participants, items, and possibly other experimental units to one another. Here, a confirmatory factor model will be used.

The primary property that sets this CLVM apart from classical latent variable models is the nature of the data level. While classical latent variable models have data levels that are to an extent mere restatements of the data (the mean of a group, the average accuracy in a condition, etc.), the diffusion model used here is based in cognitive science and has process parameters with distinct psychological interpretations. Consequently, this model will allow conclusions of the type “*there exists a latent ability that affects the speed of information processing in some conditions of this experiment, but not others, and that causes dependence between the behavior in these conditions across participants.*”

**Why to avoid two-stage procedures.** To address a question like the one above, it is tempting to consider a two-stage analysis. In such a procedure, one might (a) estimate the cognitive model parameters in each of  $C$  conditions and for each of  $P$  persons and collect the estimates in a  $P$ -by- $C$  matrix  $X$ , and then (b) perform latent variable analysis on  $X$ . There are two (related) reasons to object to this procedure. First, it is unclear how statistical inference should proceed in this scenario. The parameter estimates obtained in (a) have joint uncertainty (standard error of estimation or posterior variance, depending on one's statistical philosophy) associated with them and this uncertainty is lost in (b), a problem sometimes referred to as *generated regressor bias* (for an overview, see Pagan, 1984). As a result, the uncertainty on measurement-level parameters obtained in the second stage cannot be ascertained with off-the-shelf tools, prohibiting statistical inference (but see Vandekerckhove, Panis, & Wagemans, 2007, for an application of a computationally intensive bootstrap solution). Second, this procedure requires that each person-by-condition combination have enough data points so that parameters can be estimated at all. To use the diffusion model as an example, each cell would have to have at least some responses in each response category (e.g., at least some error and some correct responses).

Because the model proposed here is a one-stage procedure, uncertainty propagates from the data to the final parameter estimates at all levels of the model; because it is applied to an entire data set at once (allowing cross-talk be-

tween data from different conditions and participants), it is not necessary that all cells of the experimental design have many data points. These advantages are not unique to the model presented here, they are inherent to hierarchical models (Gelman & Hill, 2007; Lee, 2011).

**Why to avoid latent variable analysis on basic summary statistics.** Another conventional alternative to the approach used here would be to apply a latent variable model to the mean RTs across person-by-condition combinations. Formally, this procedure is almost identical to the two-stage analysis laid out in the previous paragraph; the only difference being that instead of cognitive model parameters being estimated in (a), parameters of a conveniently easy-to-use distribution (a Gaussian or some other member of the exponential family) are estimated. The first criticism of the two-stage approach holds exactly: uncertainty about the mean RTs is not propagated and a cell with only a handful of observations would (by default) be given equal weight to a cell with many observations. The second criticism applies only weakly: this analysis requires at least one data point per cell, which seems like a more agreeable constraint. However, this approach invites a third, more severe criticism: it does not permit the process-based conclusions that cognitive scientists often desire. While it may be possible to infer latent factors that affect mean RT, this method can shed no light on *why* the RT changes. In terms of a diffusion model, participants might differ in their ability (drift rate), in their caution (boundary separation), or in their motor response time. A cognitive process model is required to account for these differences across participants, and to decompose the interacting effects of these more elementary individual propensities.<sup>2</sup>

To summarize the rationale for the integrated model, it provides (a) *a cognitive model* because without one, one cannot draw conclusions about the process that generated the data; (b) *a latent variable model* because without one, one cannot combine data across conditions and participants to infer underlying abilities; (c) *a one-stage integrated model* because two-stage models do not propagate statistical uncertainty and statistical inference is hampered.

<sup>2</sup>It should be noted that focusing solely on descriptive statistics may be very useful from a machine-learning point of view, if the focus of the analysis is strictly to predict future mean RTs. However, from the vantage point of the cognitive scientist, a purely data-driven approach is not elucidating.

### Parameter estimation and Bayesian methods

To fit CLVMs to data, parameter estimation and inference were conducted in a Bayesian statistical framework (see, e.g., Gelman, Carlin, Stern, & Rubin, 2004). This choice was made not only because of the desirable philosophical properties of the Bayesian framework, but also because the implementation and execution of these models turns out to be comparatively easy with general-purpose Bayesian estimation software.

In the Bayesian paradigm, one's knowledge about parameters is encoded entirely as statistical distributions. Bayes' theorem (Eq. 2) is used to update knowledge about a set of parameters prior to observing the data (the *prior* distribution) with the likelihood of the data under each parameter set, in order to obtain a distribution that reflects knowledge posterior to observing the data (the *posterior* distribution):

$$p(\theta|y, \mathcal{M}) = \frac{p(y|\theta, \mathcal{M})p(\theta|\mathcal{M})}{p(y|\mathcal{M})} \quad (2)$$

Because these computations typically require high-dimensional integration with no analytical solution, numerical integration methods such as *Markov chain Monte Carlo* methods (MCMC; Robert & Casella, 1999) are a staple of applied Bayesian statistics. Without going into detail about MCMC methods in general or any sampling algorithm in particular, it bears mentioning that the procedures require some amount of quality control whenever they are applied. A common measure of quality control is to repeat the procedure multiple times with varying initial conditions, then confirming that the repeated sample chains yield similar distributions. A statistic that quantifies this convergence is Gelman and Rubin's (1992) *estimated potential scale reduction* parameter  $\hat{R}$ , which takes large values if the chains did not converge to the same distribution, and values close to 1 if they did.<sup>3</sup> Typically,  $\hat{R} < 1.1$  is considered to indicate good convergence.

Several general-purpose MCMC engines exist that are built exactly for the purpose of facilitating Bayesian analyses. These general-purpose engines include WinBUGS ("Bayesian inference Using Gibbs Sampling"; Lunn, Thomas, Best, & Spiegelhalter, 2000), JAGS ("Just Another Gibbs Sampler"; Plummer, 2003), and, more recently, Stan (Stan Development Team, 2013). JAGS and Stan are open-source, cross-platform, and easy to use. Critically, they can be extended with custom functions, distributions, and samplers, and custom cognitive models have successfully been implemented in JAGS (Wabersich & Vandekerckhove, in press).

### Model selection

A major goal in latent variable modeling (cognitive or otherwise) is dimensionality selection: the determination of the number of LVs required to account for the covariance pattern between MVs. Because the current approach involves defining a set of candidate models of different dimensionalities, model selection is a key tool. In the present application, focus will be on the Deviance Information Criterion<sup>4</sup> (DIC; Spiegelhalter, Best, Carlin, & Van Der Linde, 2002):

$$\text{DIC}_{\mathcal{M}} = D(\hat{\theta}) + p_D.$$

DIC is constructed like a classical information criterion, with a badness-of-fit component  $D(\hat{\theta})$  added to a complexity component  $p_D$ . Analogously to other information criteria like Akaike's and the Bayesian information criterion, DIC values can be transformed into model weights (Wagenmakers & Farrell, 2004):

$$w_{\mathcal{M}} = \frac{e^{-\text{DIC}_{\mathcal{M}}}}{\sum_m e^{-\text{DIC}_m}}.$$

In the application, weights  $w_{\mathcal{M}}$  will be used to select models.

### Application: Dimension reduction over hybrid data

A most typical example of latent variable modeling in cognitive science is the literature on executive functions (e.g., Miyake et al., 2000). In studies in this area, participants (typically many) are presented with batteries of related tasks, each of which taps one or more executive functions—latent constructs that are interpreted as basic functions of cognition. In the example data set (due to Pe, Raes, et al., 2013),  $P = 99$  participants performed an affective proactive interference (PI) task in which they were asked to rapidly study a set of four words, and then determine whether a probe word (presented immediately following the study set) was in the set. A *proactive interference effect* then occurs when the probe was a member of the study set in the trial directly *preceding* the current, but not in the current study set, and this sequential effect causes a decrease in performance. A typical PI task thus has four conditions, (a) a nonrecent-yes condition in which the probe was present in the current study set

<sup>3</sup> $\hat{R}$  is similar to the  $F_{df}$  statistic in analysis of variance; it is a ratio of between-chain variance and within-chain variance.

<sup>4</sup>Another model selection method that could be applicable to the models described here is automatic feature selection (AFS; Gershman & Blei, 2012).



but not in the previous, (b) a nonrecent-no condition where the probe was present in neither the current nor the previous study set, (c) a recent-yes condition in which the probe was present in both the current and previous study sets, and (d) a recent-no condition in which the probe was present in the previous set but not in the current. The PI effect shows in differential performance between conditions (b) and (d).

### A large data set with various indicators

Expanding on the popular PI paradigm, Pe, Raes, et al. (2013) also manipulated the emotional valence of the probe words over three levels: positive, negative, and neutral. This extension resulted in a total of 12 subtasks.<sup>5</sup>

Furthermore, Pe, Raes, et al. (2013) collected several clinical and personality measures in order to explore the relationship between performance in their affective PI task and emotional coping strategies. A correlation between RT and clinical measures such as dysphoria and tendency to ruminate is well documented (e.g., Bonin-Guillaume, Blin, & Hasbroucq, 2004). A novel type of question that can be addressed by a CLVM is the following: *Which* (if any) components of the task performance are related to the clinical measures of interest? Or: Can we identify interpretable components of performance in an emotional PI task that relate to depression? These questions allow a qualitatively different type of conclusion from classical analyses.

### A hybrid-data cognitive latent variable model

In order to explore this question, a series of CLVMs with *hybrid data levels* was constructed. The latent factors involved in the PI task were made to jointly predict drift rates for the PI task as well as scores on the *Center for Epidemiologic Studies Depression Scale* (CES-D; Radloff, 1977) and on the *Ruminative Response Scale* (RRS; Treynor, Gonzalez, & Nolen-Hoeksema, 2003).

**Data level.** In each model, the data level (or *marginal likelihood level*) for the behavioral data was the first-passage time distribution of an unbiased (i.e.,  $\beta = 0.5$ ) Wiener diffusion model (Eq. 1), where crossings of the lower decision boundary are interpreted as errors.<sup>6</sup> Finally, person-specific boundary separation parameters  $\alpha_{(p)}$  were allowed for, as well as person-by-task effects on the drift rates  $\delta_{(t,p)}$  and non-decision times  $\tau_{(t,p)}$ . The marginal likelihood for the choice RTs is therefore:

$$y_{(t,p,i)} \sim W\left(\alpha_{(p)}, \tau_{(t,p)}, 0.5, \delta_{(t,p)}\right), \quad (3)$$

where the distribution  $W$  is the Wiener diffusion model density as defined in Equation 1.

A separate data level needed to be defined for the covariates. A conventional choice is the normal distribution, so that if the CES-D and RRS scores of person  $p$  are  $X_{(1,p)}$  and  $X_{(2,p)}$ , respectively, then for  $c = 1, 2$ :

$$x_{(c,p)} \sim N\left(\mu_{(c,p)}, \epsilon_{(c)}^2\right).$$

**Measurement level.** In models  $\mathcal{M}_1$  through  $\mathcal{M}_7$ , the measurement level related only to the drift rates—the parameter in the diffusion model that best captures a participant’s ability at a task—and the covariates. If  $\Delta$  is the person-by-task matrix of drift rates  $\delta_{(t,p)}$ , the measurement level was the linear system  $\Delta = \Lambda \times \Phi$ , where the constraints on the loadings matrix  $\Lambda$  define the factor model. The core of these CLVMs can therefore be restated as:

$$y_{(t,p,i)} \sim W\left(\alpha_{(p)}, 0.5, \tau_{(t,p)}, \sum_{f=1}^F \lambda_{(t,f)} \phi_{(f,p)}\right).$$

Simultaneously, for the covariates, if  $M$  is the person-by-covariate matrix of traits  $\mu_{(c,p)}$ , the measurement level also included  $M = B \times \Phi$ :

$$x_{(c,p)} \sim N\left(\sum_{f=1}^F \beta_{(c,f)} \phi_{(f,p)}, \epsilon_{(c)}^2\right).$$

In order to implement the joint latent structure for the two data levels, a loadings matrix with two submatrices was constructed. For this data set, the loadings matrix had 12 rows for the PI tasks and 2 additional rows for CES-D and RRS, respectively. Now, the matrix  $\Upsilon$  contains both all the drift rate parameters  $\delta_{(t,p)}$  in a  $12$ -by- $P$  submatrix  $\Delta$  and the predicted (i.e., free of measurement error) covariate values  $\mu_{(c,p)}$  in a submatrix  $M$ . The measurement equation then takes the hybrid form

$$\begin{pmatrix} \Delta \\ M \end{pmatrix} = \begin{pmatrix} \Lambda \\ B \end{pmatrix} \times \Phi,$$

<sup>5</sup>Out of  $99 \times 152 = 15048$  trials, 55 were deleted because no response was recorded and 45 trials because their RTs were too slow to credibly represent normal task performance (more than 2s). No trials were removed because of conspicuously fast RTs (less than 0.2s). A total of 0.66% of trials were removed.

<sup>6</sup>The choice to map the boundaries to accuracy, rather than response type, was made to preserve the interpretation of the drift rate parameter as an ability parameter for which high values indicate high ability. Given this choice of mapping, the unbiased model is preferred because an a-priori bias towards whichever response option is correct (or wrong) on a given trial has no psychological meaning.

$$K = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & \lambda_{(5,2)} & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & \lambda_{(1,3)} & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & \lambda_{(9,2)} & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & \lambda_{(11,3)} & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ \hline \beta_{(1,1)} & \beta_{(1,2)} & \beta_{(1,3)} & \beta_{(1,4)} & \beta_{(1,5)} & \beta_{(1,6)} \\ \beta_{(2,1)} & \beta_{(2,2)} & \beta_{(1,3)} & \beta_{(2,4)} & \beta_{(2,5)} & \beta_{(2,6)} \end{pmatrix} \begin{array}{l} \text{nonrecent-yes, negative} \\ \text{nonrecent-no, negative} \\ \text{recent-yes, negative} \\ \text{recent-no, negative} \\ \text{nonrecent-yes, positive} \\ \text{nonrecent-no, positive} \\ \text{recent-yes, positive} \\ \text{recent-no, positive} \\ \text{nonrecent-yes, neutral} \\ \text{nonrecent-no, neutral} \\ \text{recent-yes, neutral} \\ \text{recent-no, neutral} \\ \text{CES-D} \\ \text{RRS} \end{array}$$

Equation 4: The measurement level of the hybrid-data CLVM  $\mathcal{M}_6$  in the example application.

or more concisely:  $\Upsilon = K \times \Phi$ , with  $K$  defined as in Equation 4.

In the hybrid loadings matrix  $K$ , the first column captures general task ability. This “baseline ability” determines performance in the nonrecent-no conditions. The second column captures the ability to detect a novel probe in a set (*DA—nonrecent*) and the third the detection ability (*DA*) for a recently seen probe in a set (*DA—recent*; this factor could also be thought of as ‘primed detection’). The final three columns are PI effects for the negative (*PI(−)*), positive (*PI(+)*), and neutral stimuli (*PI(∅)*), respectively. As indicated, the first four rows pertain to the drift rates of the nonrecent-yes, nonrecent-no, recent-yes, and recent-no conditions with negative target stimuli, the next eight rows are repetitions for the positive and neutral stimuli, and the bottom two rows are the loading weights for the CES-D and RRS covariates, respectively.

The model can be identified through a unit factor variance constraint<sup>7</sup>, so that all  $\lambda$ -s in the loadings matrix are free to be estimated.

In addition to the CLVM now fully defined, nine more models were constructed. By defining a large set of competing models and deciding to select between them, this procedure is in line with the advice of McArdle (2011), who recommended that model selection begin with a confirmatory phase and conclude with a more exploratory phase in which competing models are considered.

Candidate models were inspired either by ad-hoc theories of the PI task, by the desire to compare to a saturated or base-

line model, or were suggested by reviewers. The model just described is  $\mathcal{M}_6$ . Models  $\mathcal{M}_1$  through  $\mathcal{M}_7$  are variations on this model, differing only in the loadings matrix  $\Lambda$  (details of all models are given in the Appendix).

Three CLVMs ( $\mathcal{M}_8$  through  $\mathcal{M}_{10}$ ) were constructed at the suggestion of reviewers to focus on diffusion model parameters other than the drift rate. Two models will be considered with factorial structures that pertain to the nondecision time. The first such model ( $\mathcal{M}_8$ ) had the same loadings matrix as  $K$ , above. The second nondecision time model ( $\mathcal{M}_9$ ) had an identity matrix for  $\Lambda_9$ . Finally, one model ( $\mathcal{M}_{10}$ ) had latent variables simultaneously for drift rate and boundary separa-

<sup>7</sup>Under a unit factor variance constraint, the between-person variance in factor scores is constrained to be 1. The practical implementation of this constraint in a Bayesian context involves a two-stage procedure in which the model is first estimated with unconstrained factors but a constrained  $\lambda$ -structure, as above. In a second stage, the estimated factor scores ( $\hat{\phi}$ ) and loadings ( $\hat{\lambda}$  and  $\hat{\beta}$ ) are rescaled:  $\phi_{(f,p)}^{(i)} = \hat{\phi}_{(f,p)}^{(i)} / \sigma_{(f)}^{(i)}$ , where the superscript  $i$  refers to the  $i^{\text{th}}$  sample in the MCMC chain, and  $\sigma_{(f)}^{(i)}$  is the between-person standard deviation in factor  $f$  at iteration  $i$ . Additionally,  $\lambda_{(t,f)}^{(i)} = \hat{\lambda}_{(t,f)}^{(i)} \times \sigma_{(f)}^{(i)}$  and  $\beta_{(c,f)}^{(i)} = \hat{\beta}_{(c,f)}^{(i)} \times \sigma_{(f)}^{(i)}$ , for all covariates  $c = 1, \dots, C$ . In the loadings matrix, this will result in the constraint that all  $\lambda$ -s that were originally set to 1 will now be freely estimated, under the remaining constraint that  $\lambda$ -s that were originally set equal to 1 within the same column of  $\Lambda$  are equal to one another. Note that this change of constraint implies slightly different priors on the affected parameters.

tion, so that its measurement equation was:

$$\begin{pmatrix} \Delta \\ A \\ M \end{pmatrix} = \begin{pmatrix} \Lambda_{10} \\ B \end{pmatrix} \times \Phi,$$

with  $A$  referring to the vector of person-specific boundary separation parameters. The weights matrix  $B$  now contains two extra entries that connect the latent variable for boundary separation to the personality covariates. The full loadings matrix is given as Equation 5.

**Priors.** As in all Bayesian analyses, a CLVM in a Bayesian framework requires that one define a number of prior distributions. The priors used for the present analysis are given in Figure 4). These priors are generally weakly informative, specifying a plausible range for the parameters but low weight to possible but unlikely values.

The normal distribution is conjugate for the drift rate of the diffusion model (i.e., a normal prior leads to a normal full conditional distribution) and is for that reason preferred for the parameters relating to the drift rate. For the other parameters of the diffusion model, no conjugate priors are available, and priors were chosen that reflect knowledge of the scale of the variables and that have positive density across a range that is certain to contain the domain of the posterior.<sup>8</sup>

These priors, together with the data level in Equation 3 and the measurement equations  $\Upsilon = K \times \Phi$ , fully define the model. Figure 4 shows a graphical representation of the model. Note how that graph includes as components the graphical models for a diffusion model across conditions (Fig. 2) and an LV model with various measures (Fig. 3), which clearly visualizes the cross-breeding between modeling traditions.

### Results—technical

For each of the models, JAGS was used to run eight MCMC chains with 5,000 iterations each. From each chain, 3,000 samples were discarded as burn-in, leaving 16,000 posterior samples. Negligible chain autocorrelations indicated good mixing and no need for chain thinning. With few exceptions, potential scale reduction parameters ( $\hat{R}$ ; Gelman et al., 2004) were less than 1.1 (and all were below 1.2), indicating good chain convergence in all dimensions.

### Results—model evaluation

For each of the models, we computed the DIC fit measure, as well as model selection weights  $w_M$  based on DIC. The two best fitting models were  $M_{10}$  ( $DIC_{10} = -2718$ ) and  $M_6$

( $DIC_6 = -2281$ ; fit information for all models is given in the Appendix). The weights  $w$  were negligible for all models except  $M_{10}$  (and so  $w_{10} \approx 1$ ). For interpretation purposes, it is additionally worth noting a pattern across model solutions: the CESD and RRS scales consistently turn out to relate to the PI effect, and in particular to the PI effect in the negative-target condition, across all models that include a PI effect.

Though the psychometric modeling approach does not require the model to capture minor aspects of the data exactly, it is important for parameter interpretation that there is at least a coarse correspondence between model and data. Figure 5 provides diagnostic graphical contact between model  $M_{10}$  and the data. To construct that figure, *posterior predictive statistics* were generated: data sets generated from each of 5,000 samples that had been drawn from the full posterior distribution of the model parameters. On each data set so generated, as well as on the raw data, a set of summary statistics was computed. In the figure, the distribution of the statistics generated by the model (shades of grey; darker means higher model-predicted density) is overlaid with the raw data (the white markers). The posterior predictive check does not indicate any systematic misfit.

### Results—substantive

Part of the factor solution of the hybrid-data CLVM ( $M_{10}$ ) is shown in Figure 6. The round markers indicating choice RT tasks are all placed on an axis, indicating that the tasks load exclusively on one dimension<sup>9</sup>, some more strongly than others. More interesting for the present data is the location of the questionnaire measures (CES-D and RRS; triangular markers). The partial factor space shows both measures in approximately the same location, loading strongly negatively on the  $PI(-)$  dimension and the *boundary separation* dimension, somewhat negatively on the  $DA-nonrecent$  factor, but very close to 0 on the  $PI(+)$  dimension. The loadings were also close to 0 on the intercept,  $DA-recent$ , and  $PI(\emptyset)$  dimensions (not shown).

Table 1 shows the loadings ( $\beta$ -s) for both measures on

<sup>8</sup>Limited robustness checks were performed by introducing changes in these priors. For example, a normal was changed to a uniform over a wide range or vice versa, and no meaningful differences in the results were observed.

<sup>9</sup>This is by design, since PI can only occur in those trials where detection is not at hand, and only one (valence-specific) type of PI can occur at a time. This structure is also seen in the loadings matrix where, disregarding the intercept (the first column), each row contains at most one nonzero element. Note that this is a property of the experimental paradigm and not a condition of the CLVM.

1	1	0	0	0	0	0	0	nonrecent-yes, negative
1	0	0	0	0	0	0	0	nonrecent-no, negative
1	0	1	0	0	0	0	0	recent-yes, negative
1	0	0	1	0	0	0	0	recent-no, negative
1	$\lambda_{(5,2)}$	0	0	0	0	0	0	nonrecent-yes, positive
1	0	0	0	0	0	0	0	nonrecent-no, positive
1	0	$\lambda_{(7,3)}$	0	0	0	0	0	recent-yes, positive
1	0	0	0	1	0	0	0	recent-no, positive
1	$\lambda_{(9,2)}$	0	0	0	0	0	0	nonrecent-yes, neutral
1	0	0	0	0	0	0	0	nonrecent-no, neutral
1	0	$\lambda_{(11,3)}$	0	0	0	0	0	recent-yes, neutral
1	0	0	0	0	1	0	0	recent-no, neutral
0	0	0	0	0	0	0	1	boundary separation
$\beta_{(1,1)}$	$\beta_{(1,2)}$	$\beta_{(1,3)}$	$\beta_{(1,4)}$	$\beta_{(1,5)}$	$\beta_{(1,6)}$	$\beta_{(1,7)}$		CES-D
$\beta_{(2,1)}$	$\beta_{(2,2)}$	$\beta_{(1,3)}$	$\beta_{(2,4)}$	$\beta_{(2,5)}$	$\beta_{(2,6)}$	$\beta_{(2,7)}$		RRS

Equation 5: The measurement level of the hybrid-data CLVM  $\mathcal{M}_{10}$  in the example application. In comparison to the measurement level of  $\mathcal{M}_6$ , this loadings matrix has an extra row and column to accommodate the person-specific boundary separation parameters, which now follow the distribution  $\alpha_{(p)} \sim N(\phi_{(7,p)}, \sigma_\alpha^2)$ .

each dimension. Out of twelve loadings, four show almost no posterior mass around 0. This is also displayed in Figure 6, where none of the credible interval ellipses intersect the horizontal axis. Additionally, the covariate scores are well recovered by the model: the correlation between CES-D and its latent proxy  $\hat{\mu}_1$  is .93 ( $\hat{\epsilon}_1^2 = 0.33$ ) and that between RRS and  $\hat{\mu}_2$  is .91 ( $\hat{\epsilon}_2^2 = 0.36$ ). The conclusion of these covariate results—that participants with higher scores on depression and rumination scales show poorer ability at inhibiting interference from negative stimuli—is in line with cognitive theories of depression (Gotlib, Roberts, & Gilboa, 1996).

### Discussion of example application

The example analysis involved fitting a series of CLVMs to a data set containing two types of data: 12 conditions of a RT experiment and 2 personality trait measures. A single underlying factor structure was defined that jointly predicted behavior in the RT task and scores on the personality traits. The use of a single unified model to tie together cognitive model parameters with personality traits allowed for a specific, novel conclusion: participants with higher dysphoria scores show more degraded information processing when a cognitive task requires the suppression of lingering negative thoughts. In the literature on cognitive theories of depression, this inertia of negatively valenced stimuli was predicted to be connected to dysphoria (e.g., Gotlib et al., 1996), and

Table 1

The loadings of the covariate measures in the hybrid-data CLVM factor solution (using a unit factor variance constraint).

Latent variable	Covariate: Depression		Covariate: Rumination	
	Mean	(SD)	Mean	(SD)
DA—recent <sup>a</sup>	0.05	(0.11)	0.04	(0.11)
DA—nonrecent <sup>a</sup>	-0.22	(0.14)	-0.20	(0.14)
PI(−) <sup>b</sup>	-0.43*	(0.14)	-0.39*	(0.14)
PI(+) <sup>b</sup>	-0.04	(0.14)	0.05	(0.14)
PI(∅) <sup>b</sup>	-0.15	(0.13)	0.01	(0.14)
Boundary separation	-0.46*	(0.16)	-0.51*	(0.15)

\* Posterior  $p(\beta > 0) < 0.01$ .

<sup>a</sup>: DA = Detection ability. <sup>b</sup>: PI = Proactive interference.

these results are in line with that prediction. Additionally, participants with high dysphoria scores turn out to have lower boundary separation parameters.

The CLVM model appears to fit the data well, and the uncovered relation between cognitive performance and dysphoria is robust across variations of the model (i.e., the relation hold in all models considered that contained a similar  $PI(-)$  factor). The data appear to support the cognitive theory of depression that involves the lingering of negative thoughts.

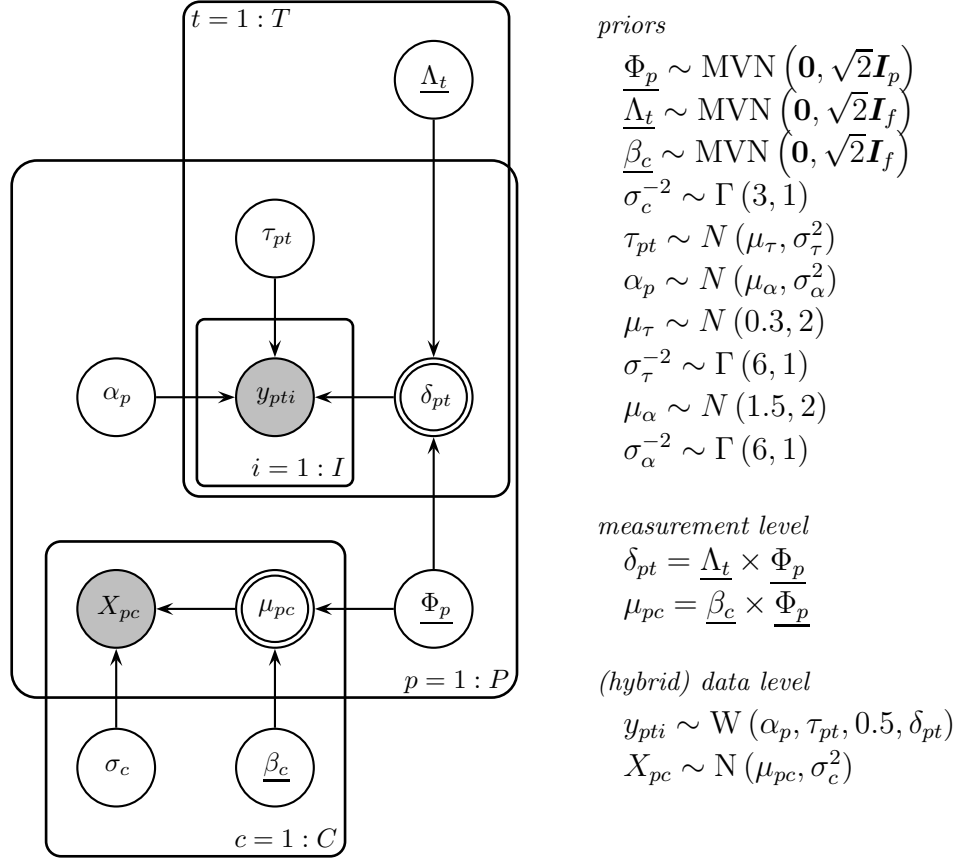


Figure 4. A graphical representation of the model used in the example application. In this graph, the  $p$ -plate indicates independent repetitions over  $P$  participants,  $t$  over  $T$  tasks,  $i$  over  $I$  trials, and  $c$  over  $C$  external covariates. The defining aspect of a *hybrid-data* CLVM is shown in the two arrows leaving the latent factor node  $\Phi$ : The single set of underlying latent variables unifies the correlational structure among and between ability parameters  $\delta_{(t,p)}$  and covariates  $X_{(c,p)}$ . The graphical models for a diffusion model (Fig. 2) and an LV model (Fig. 3) are clearly subsumed in the graphical representation of the CLVM. Note that identification constraints are not represented in this display.

### The utility of using a CLVM

Though a previous section made an a-priori rationale for using a CLVM over more conventional approaches—namely that it is a better approximation of the actual sampling scheme of the data—one might still wonder how a traditional analysis would fare with the present data. At the suggestion of a reviewer, a two-stage analysis was performed in which (a) a subject-wise measure of PI was defined as the mean RT in the recent-no condition minus that in the non-recent-no condition, and then (b) the subject-wise measure was regressed on the dysphoria scores on the tests. The regression weight between the PI score and CES-D was small but significant ( $\beta = -0.02$ ,  $t_{97} = -2.48$ ,  $p < .05$ ,  $r^2 = .06$ ).

The regression weight with RRS did not reach significance despite the large data set ( $\beta = -0.01$ ,  $t_{97} = -1.60$ ,  $ns$ ).

It is not clear how this numerical result should be interpreted. Is CES-D related to a difference in mean RT because lower dysphoria is associated with greater caution or indecisiveness, which in turn causes heteroskedasticity<sup>10</sup> and selective exaggeration of RT differences? Or do negatively valenced stimuli linger, causing processing interference if they

<sup>10</sup>Mean and standard deviation of RTs tend to be correlated (i.e., RTs show *scalar variability*). Greater caution therefore not only causes an increase in mean RT, but also greater variability. A dependent measure with greater variability will show exaggerated effects for identical manipulations, causing an otherwise spurious correlation between effect sizes and variability across participants.

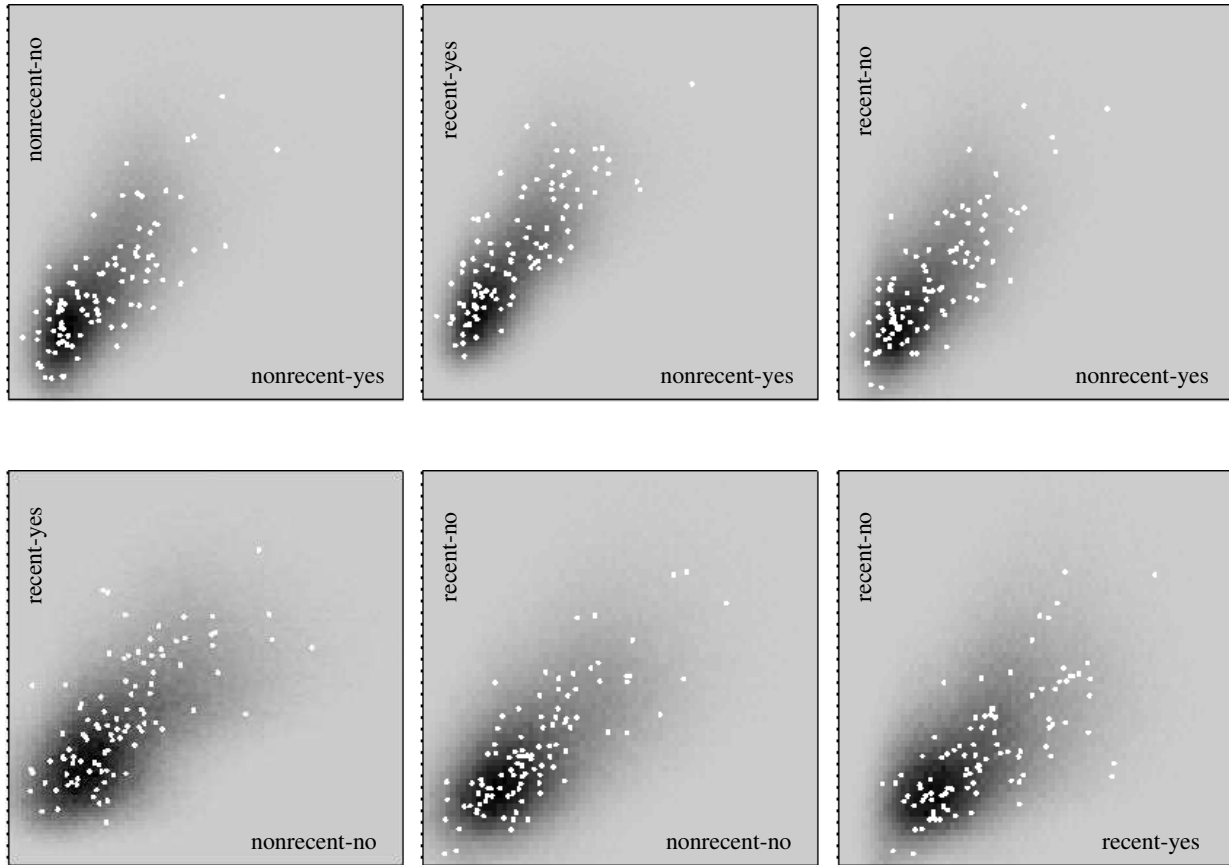


Figure 5. Posterior predictive statistics from the PI model. Each white dot represents a participant's mean RT in the task indicated by the axis. Only results from the negative-target conditions are shown, but they were similar for the other two valences. The shaded area represents the model-expected distribution of mean RTs. Several observations can be made. First, the RT means across tasks correlate in the raw data, and the model clearly shows a corresponding covariance structure. Second, all the salient aspects of the data (location, variance, covariance) are well captured by the model. Finally, there is no evidence in this posterior predictive check of systematic model misfit. Response accuracy (not displayed) is captured similarly well.

become targets in a subsequent trial? The CLVM model allowed the latter conclusion. Does RRS not carry a significant correlation because it is unrelated to PI? Perhaps the comparison to the CLVM is unfair because the CLVM took both covariates into account simultaneously and could exploit their collinearity? A multiple regression attempt using CES-D and RRS to jointly predict the negative-PI effect brought no solace for the traditional analysis (CES-D:  $\beta = -0.02$ ,  $t_{96} = -1.86$ , *ns*, RRS:  $\beta = -0.00$ ,  $t_{96} = -0.01$ , *ns*). More likely, the effect of RRS is occluded by the loss of information going from one stage to the next.

In contrast to the traditional approach, the CLVM provides parameter estimates with no such statistical pitfalls, and that

may be readily interpreted in process model terms.

### General discussion

The present paper introduces and demonstrates a *cognitive latent variable model*, a model that is a blend of cognitive modeling and psychometric latent variable modeling. This model permitted conclusions about structural relations between cognitive constructs in a way that was not possible through either component alone.

The current approach leaves room for a number of extensions. For example, the structural level could be extended to include higher level structures, so that not only the manifest variables have a correlational structure, but the latent

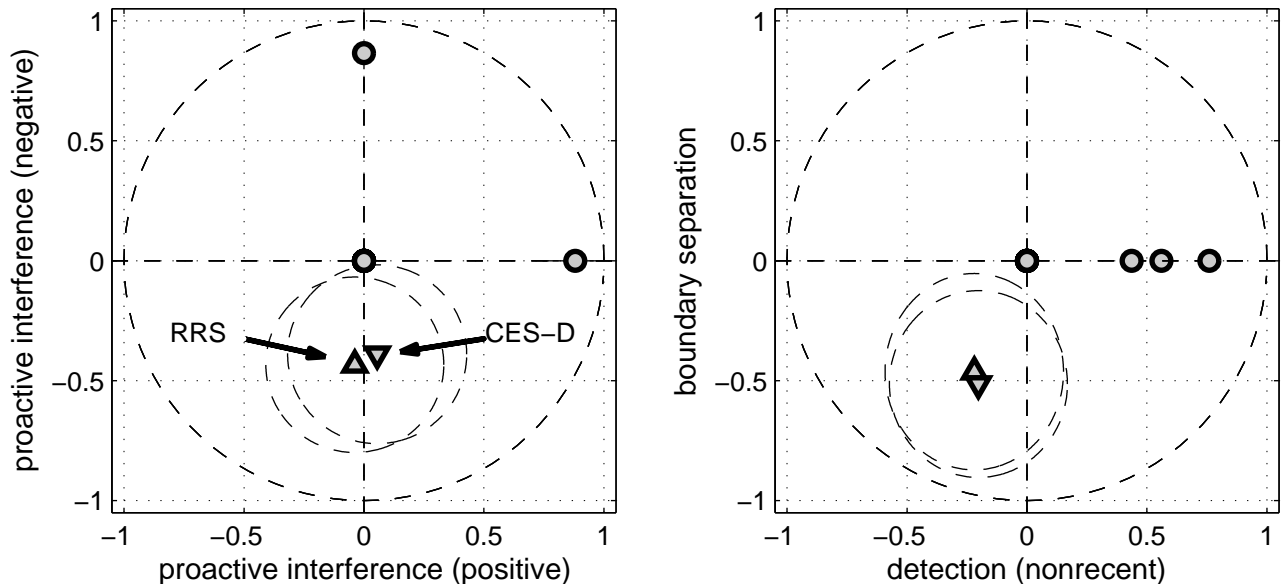


Figure 6. Four dimensions of the factor solution obtained from the hybrid-data CLVM  $\mathcal{M}_{10}$ . **Left panel:** The latent factors  $PI(+)$  and  $PI(-)$  (corresponding to the fourth and fifth columns of the loadings matrix). **Right panel:** The latent factors *boundary separation* and *DA—nonrecent* (corresponding to the seventh and third columns of the loadings matrix). In each panel, an axis represents a latent factor, a round marker a condition in the experiment, and a triangular marker a covariate (upward pointing for RSS, downward pointing for CES-D). The location of a marker indicates the loadings on each latent factor, so that markers close to the origin are unrelated to the latent factors and markers closer to the unit circle (dashed circle, drawn for reference only) are strongly related. The markers for covariates are surrounded by a dashed ellipse, indicating the 99% Bayesian credibility interval of the location. Note that these loadings were obtained under a unit factor variance constraint.

variables as well. Such models could be called *cognitive structural equation models*. In such a model, the factor matrix  $\Phi$  would be subject to further constraints similar to the ones implied by the measurement models used here, so that  $\Phi = \Theta \times \Gamma$ . A higher-order loadings matrix  $\Theta$  together with a low-dimensional set of more abstract abilities  $\Gamma$  would then generate the basic factors  $\Phi$ —repeating the analogy used earlier,  $\Phi$  could contain *mathematical ability* (which underlies positive correlations among mathematics tests) as well as *language ability* (causing covariance among language tests), but these two abilities might themselves be correlated with one another due to the higher-order, more abstract ability *intelligence*.

Going one step further, one could consider *nonlinear* structural equations, in which MVs are a function of the interaction between multiple latent variables. Using the same example, one might imagine that a test taker’s mathematical ability  $\phi_M$  is expressed in a test through their language ability  $\phi_L$ —that is, their mathematics score is partly modulated by their language ability. In this case,  $\mu_{(c,p)} = \lambda_{(c,M \times L)} \phi_{(M,p)} \phi_{(L,p)} + \dots$  (where the ellipsis is used to omit

other potential additive factors).

On the strictly technical and implementational side, there will be a need for more efficient parameter estimation routines. While the Monte Carlo methods we applied were effective, the analyses in the example took well over a day of computing time. This computational expense is partly due to the complex likelihood evaluation, but inefficient sampling increased the computing time ten- or twentyfold. The JAGS computing platform is highly customizable, modular, and extendable, so that the current sampler could be substituted for a more efficient one (e.g., one that takes into account new conjugacy relationships) without changing the model specification. Alternatively, the models could be implemented in Stan, which is at the time of writing still under development but may turn out to be more efficient due to its use of the Hamiltonian Monte Carlo sampler (Hoffman & Gelman, 2011). Finally, the issue of model selection and model identification in a CLVM context will require careful attention.

Finding latent structure in interpretable cognitive model parameters seems a highly appropriate endeavor for cognitive scientists, and an integrative CLVM has many poten-

tial applications. Many areas of psychology deal with latent structures that are tapped by batteries of tests, and that are only observed in the correlational pattern across tasks. One example is working memory research, where batteries of working memory tasks are administered in order to infer the low-dimensional structure of working memory (e.g., Oberauer et al., 2000). Similarly, the structure of executive functions is typically studied through large sets of smaller tasks, each potentially with a cognitive model underlying it (e.g., Miyake et al., 2000). Here the focus was on a diffusion model data level, but these future applications may occasion tailor-made data levels.

### References

- Bartholomew, D. J., Knott, M., & Moustaki, I. (2011). *Latent variable models and factor analysis: A unified approach*. Wiley.
- Bartlema, A., Lee, M., Wetzels, R., & Vanpaemel, W. (in press). A Bayesian hierarchical mixture approach to individual differences: Case studies in selective attention and representation in category learning. *Journal of Mathematical Psychology*.
- Batchelder, W. (1998). Multinomial processing tree models and psychological assessment. *Psychological Assessment*, 10(4), 331–344.
- Batchelder, W. (2007). Cognitive psychometrics: Combining two psychological traditions. *CSCA Lecture, Amsterdam, The Netherlands*.
- Blurton, S., Kesselmeier, M., & Gondan, M. (2012). Fast and accurate calculations for cumulative first-passage time distributions in Wiener diffusion models. *Journal of Mathematical Psychology*, 56(6), 470–475.
- Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual review of psychology*, 53(1), 605–634.
- Bonin-Guillaume, S., Blin, O., & Hasbroucq, T. (2004). An additive factor analysis of the effect of depression on the reaction time of old patients. *Acta Psychologica*, 117(1), 1–11.
- Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, 71(3), 425–440.
- Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12(11), 671–684.
- De Boeck, P., & Wilson, M. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. Springer.
- Estes, W. (1956). The problem of inference from curves based on group data. *Psychological Bulletin*, 53(2), 134.
- Estes, W. (2002). Traps in the route to models of memory and decision. *Psychonomic Bulletin & Review*, 9(1), 3–25.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis*. Chapman & Hall/CRC.
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 457–472.
- Gershman, S. J., & Blei, D. M. (2012). A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*, 56(1), 1–12.
- Geweke, J., & Zhou, G. (1996). Measuring the pricing error of the arbitrage pricing theory. *Review of Financial Studies*, 9(2), 557–587.
- Gotlib, I. H., Roberts, J. E., & Gilboa, E. (1996). Cognitive interference in depression. In I. G. Sarason, G. R. Pierce, & B. R. Sarason (Eds.), *Cognitive interference: Theories, methods, and findings* (pp. 347–377). Lawrence Erlbaum Associates.
- Hamaker, E. L. (2012). Why researchers should think “within-person”: A paradigmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life*. New York, NY: Guilford Publications.
- Heathcote, A., Brown, S., & Mewhort, D. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review*, 7(2), 185–207.
- Hoffman, M. D., & Gelman, A. (2011).
- Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34(2), 183–202.
- Kamphaus, R. W., Petoskey, M. D., & Morgan, A. W. (1997). A history of intelligence test interpretation. *Contemporary intellectual assessment: Theories, tests, and issues*, 3–16.
- Lee, M. D. (2008). Three case studies in the Bayesian analysis of cognitive models. *Psychonomic Bulletin & Review*, 15(1), 1–15.
- Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology*, 55(1), 1–7.
- Lee, M. D., & Wetzels, R. (2010). Individual differences in attention during category learning. In *Proceedings*



- of the 32nd annual conference of the cognitive science society (pp. 387–392).
- Loken, E. (2005). Identification constraints and inference in factor models. *Structural Equation Modeling*, 12(2), 232–244.
- Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS—A Bayesian modeling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10(4), 325–337.
- McArdle, J. J. (2011). Some ethical issues in factor analysis. *Handbook of Ethics in Quantitative Methodology*, 313.
- 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.
- Navarro, D., & Fuss, I. (2009). Fast and accurate calculations for first-passage times in Wiener diffusion models. *Journal of Mathematical Psychology*, 53(4), 222–230.
- Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity—facets of a cognitive ability construct. *Personality and Individual Differences*, 29(6), 1017–1045.
- Oravecz, Z., Tuerlinckx, F., & Vandekerckhove, J. (2009). A hierarchical Ornstein–Uhlenbeck model for continuous repeated measurement data. *Psychometrika*, 74(3), 395–418.
- Pagan, A. (1984). Econometric issues in the analysis of regressions with generated regressors. *International Economic Review*, 25(1), 221–247.
- Pe, M. L., Raes, F., Koval, P., Brans, K., Verduyn, P., & Kuppens, P. (2013). Interference resolution moderates the impact of rumination and reappraisal on affective experiences in daily life. *Cognition & Emotion*(27), 492–501.
- Pe, M. L., Vandekerckhove, J., & Kuppens, P. (2013). A diffusion model account of the relationship between the emotional flanker task and rumination and depression. *Emotion*, 13, 739–747.
- 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 (DSC 2003)* (pp. 20–22).
- Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1(3), 385–401.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2), 59.
- Ratcliff, R., & Rouder, J. (1998). Modeling response times for two-choice decisions. *Psychological Science*, 9(5), 347–356.
- Robert, C. P., & Casella, G. (1999). *Monte carlo statistical methods* (Vol. 58). Springer New York.
- Rouder, J., Sun, D., Speckman, P., Lu, J., & Zhou, D. (2003). A hierarchical Bayesian statistical framework for response time distributions. *Psychometrika*, 68(4), 589–606.
- Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. CRC Press.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4), 583–639.
- Stan Development Team. (2013). *Stan: A C++ Library for Probability and Sampling, Version 1.3*.
- Stone, M. (1960). Models for choice-reaction time. *Psychometrika*, 25(3), 251–260.
- Treynor, W., Gonzalez, R., & Nolen-Hoeksema, S. (2003). Rumination reconsidered: A psychometric analysis. *Cognitive Therapy and Research*, 27(3), 247–259.
- Vandekerckhove, J. (2009). *Extensions and applications of the diffusion model for two-choice response times*. Unpublished doctoral dissertation, University of Leuven.
- Vandekerckhove, J., Panis, S., & Wagemans, J. (2007). The concavity effect is a compound of local and global effects. *Attention, Perception, & Psychophysics*, 69(7), 1253–1260.
- Vandekerckhove, J., Tuerlinckx, F., & Lee, M. (2008). A Bayesian approach to diffusion process models of decision-making. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1429–1434).
- Vandekerckhove, J., Verheyen, S., & Tuerlinckx, F. (2010). A crossed random effects diffusion model for speeded semantic categorization decisions. *Acta Psychologica*, 133(3), 269–282.
- Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition*, 32(7), 1206–1220.
- Wabersich, D., & Vandekerckhove, J. (in press). Extending JAGS: Adding custom distributions to JAGS, with a diffusion model example. *Behavior Research Methods*.

- Wagenmakers, E.-J. (2009). Methodological and empirical developments for the Ratcliff Diffusion Model of response times and accuracy. *European Journal of Cognitive Psychology, 21*(5), 641–671.
- Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic bulletin & review, 11*(1), 192–196.
- White, C., Ratcliff, R., Vasey, M., & McKoon, G. (2009). Dysphoria and memory for emotional material: A diffusion-model analysis. *Cognition and Emotion, 23*(1), 181–205.
- Williams, L. J. (1995). Covariance structure modeling in organizational research: Problems with the method versus applications of the method. *Journal of Organizational Behavior, 16*(4), 225–234.

## Appendix

Parameter estimates for the various models in the example application

This Appendix lists parameter estimates for each of the models considered in the example application.

**Models under consideration**

There were ten different models, all with the same hybrid data level but differing in their measurement levels. The estimated loadings matrices  $\Lambda$  are shown in this Appendix. In all matrices, loadings that were not estimated are displayed in italics. Estimates are posterior means. Loadings are displayed with an asterisk if less than 1% of their posterior mass is on the side of zero opposite the posterior mean (i.e., the posterior probability of the displayed sign of the loading is at least .99). Unless otherwise noted, the first factor in all solutions is an intercept; the others are explained below. The bottom two rows (or rightmost columns in the transposition) in a loadings matrix always refer to CESD and RRS, respectively.

Model 1 had only one latent factor for detection, with no PI effect.

$$(\Lambda|\mathcal{M}_1)^T = \begin{pmatrix} 0.51^* & 0.51^* & 0.51^* & 0.51^* & 0.51^* & 0.51^* & 0.51^* & 0.51^* & 0.51^* & 0.51^* & 0.51^* & 0.51^* & 0.09 & 0.11 \\ 0.47^* & 0 & 0.47^* & 0 & 0.79^* & 0 & 0.22^* & 0 & 0.73^* & 0 & 0.23^* & 0 & -0.18 & -0.21 \end{pmatrix}$$

Model 2 had only one latent factor for detection and a constant PI effect across valences. Preferring this model over  $\mathcal{M}_1$  would indicate the existence of some PI effect.

$$(\Lambda|\mathcal{M}_2)^T = \begin{pmatrix} 0.51^* & 0.51^* & 0.51^* & 0.51^* & 0.51^* & 0.51^* & 0.51^* & 0.51^* & 0.51^* & 0.51^* & 0.51^* & 0.51^* & 0.21^* & 0.22^* \\ 0.47^* & 0 & 0.47^* & 0 & 0.78^* & 0 & 0.24^* & 0 & 0.73^* & 0 & 0.25^* & 0 & -0.32^* & -0.34^* \\ 0 & 0 & 0 & 0.64^* & 0 & 0 & 0 & 0.64^* & 0 & 0 & 0 & 0.64^* & -0.27 & -0.15 \end{pmatrix}$$

Model 3 had a single latent factor for detection and three valence-specific factors for PI. Preferring this model over  $\mathcal{M}_2$  would indicate that the PI effect depends on the valence of the stimulus.

$$(\Lambda|\mathcal{M}_3)^T = \begin{pmatrix} 0.57^* & 0.57^* & 0.57^* & 0.57^* & 0.57^* & 0.57^* & 0.57^* & 0.57^* & 0.57^* & 0.57^* & 0.57^* & 0.57^* & 0.16^* & 0.15^* \\ 0.61^* & 0 & 0.61^* & 0 & 0.61^* & 0 & 0.61^* & 0 & 0.61^* & 0 & 0.61^* & 0 & -0.28^* & -0.31^* \\ 0 & 0 & 0 & 0.80^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.63^* & -0.60^* \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.88^* & 0 & 0 & 0 & 0 & -0.00 & 0.08 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.92^* & -0.06 & 0.07 \end{pmatrix}$$

Model 4 had a single latent factor for detection and three valence-specific factors for PI, but the degree to which the various valence conditions depend on the detection factor is no longer constant. Preferring this model over  $\mathcal{M}_3$  would indicate that the degree to which the detection factor affects the drift rate depends on the valence of the stimulus.

$$(\Lambda|\mathcal{M}_4)^T = \begin{pmatrix} 0.51^* & 0.51^* & 0.51^* & 0.51^* & 0.51^* & 0.51^* & 0.51^* & 0.51^* & 0.51^* & 0.51^* & 0.51^* & 0.51^* & 0.19^* & 0.18^* \\ 0.47^* & 0 & 0.47^* & 0 & 0.80^* & 0 & 0.24^* & 0 & 0.73^* & 0 & 0.24^* & 0 & -0.29^* & -0.31^* \\ 0 & 0 & 0 & 0.81^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.63^* & -0.61^* \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.88^* & 0 & 0 & 0 & 0 & 0.17 & 0.24 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.92^* & 0.03 & 0.15 \end{pmatrix}$$

Model 5 had only one latent factor for detection, with detection no longer forced to be equal between the recent and nonrecent yes conditions. Preferring this model over  $\mathcal{M}_4$  would indicate that the influence of the detection factor scales differently between the recent and nonrecent conditions.

$$(\Lambda|\mathcal{M}_5)^T = \begin{pmatrix} 0.50^* & 0.50^* & 0.50^* & 0.50^* & 0.50^* & 0.50^* & 0.50^* & 0.50^* & 0.50^* & 0.50^* & 0.50^* & 0.50^* & 0.14 & 0.14 \\ 0.95^* & 0 & 0.29^* & 0 & 0.75^* & 0 & 0.23^* & 0 & 0.68^* & 0 & 0.22^* & 0 & -0.21 & -0.24^* \\ 0 & 0 & 0 & 0.82^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.63^* & -0.62^* \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.89^* & 0 & 0 & 0 & 0 & 0.36 & 0.41 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.92^* & 0.18 & 0.29 \end{pmatrix}$$

Model 6 had two latent factors for detection, and different PI effects across valences. Preferring this model over  $\mathcal{M}_5$  would indicate that the drift rates are not strongly correlated between recent and nonrecent conditions, and merit separate latent abilities.

$(\Delta|\mathcal{M}_6)^T =$

$$\begin{pmatrix} 0.55^* & 0.55^* & 0.55^* & 0.55^* & 0.55^* & 0.55^* & 0.55^* & 0.55^* & 0.55^* & 0.55^* & 0.55^* & 0.09 & 0.08 \\ 0.78^* & 0 & 0 & 0 & 0.66^* & 0 & 0 & 0 & 0.60^* & 0 & 0 & -0.03 & -0.04 \\ 0 & 0 & 0.75^* & 0 & 0 & 0 & 0.43^* & 0 & 0 & 0 & 0.53^* & 0 & -0.41^* & -0.41^* \\ 0 & 0 & 0 & 0.81^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.61^* & -0.57^* \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.87^* & 0 & 0 & 0 & 0 & -0.10 & -0.02 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.92^* & -0.10 & 0.04 \end{pmatrix}$$

Model 7 is a saturated model with respect to drift rates: it has one latent factor for each condition, and each of the latent factors has a loading for the two covariates. Note that this model has no intercept:

$(\Delta|\mathcal{M}_7)^T =$

$$\begin{pmatrix} 0.84^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.37 & -0.28 \\ 0 & 0.79^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.21 & 0.37 \\ 0 & 0 & 0.67^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.47^* & -0.30 \\ 0 & 0 & 0 & 0.87^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.44^* & -0.37 \\ 0 & 0 & 0 & 0 & 0.81^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.47^* & 0.39^* \\ 0 & 0 & 0 & 0 & 0 & 0.71^* & 0 & 0 & 0 & 0 & 0 & 0 & 0.09 & 0.04 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.69^* & 0 & 0 & 0 & 0 & 0 & 0.03 & -0.11 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.86^* & 0 & 0 & 0 & 0 & 0.07 & 0.15 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.80^* & 0 & 0 & 0 & -0.11 & -0.12 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.79^* & 0 & 0 & 0.35 & 0.09 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.74^* & 0 & 0.19 & 0.13 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.91^* & 0.04 & 0.17 \end{pmatrix}$$

Model 8 does not apply a structure to drift rates, but on nondecision time. The structure is identical to the one used in

$\mathcal{M}_6$ :

$(\Delta|\mathcal{M}_8)^T =$

$$\begin{pmatrix} 0.07^* & 0.07^* & 0.07^* & 0.07^* & 0.07^* & 0.07^* & 0.07^* & 0.07^* & 0.07^* & 0.07^* & 0.07^* & 0.07^* & 0.00 & -0.01 \\ 0.06^* & 0 & 0 & 0 & 0.05^* & 0 & 0 & 0 & 0.05^* & 0 & 0 & 0 & -0.00 & 0.04 \\ 0 & 0 & 0.07^* & 0 & 0 & 0 & 0.06^* & 0 & 0 & 0 & 0.06^* & 0 & -0.05 & -0.02 \\ 0 & 0 & 0 & 0.08^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.01 & -0.03 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.07^* & 0 & 0 & 0 & 0 & -0.02 & 0.00 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.07^* & 0.09 & 0.07 \end{pmatrix}$$

Model 9 is a saturated model with respect to nondecision time: it takes one latent factor for each condition's nondecision time. This model differs from  $\mathcal{M}_7$  in that the covariates are now tied to nondecision time instead of drift rate. Preferring this model over  $\mathcal{M}_8$  would indicate that nondecision time does not follow the design of the experiment.

$(\Delta|\mathcal{M}_9)^T =$

$$\begin{pmatrix} 0.07^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.02 & 0.02 \\ 0 & 0.09^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.06 & -0.10 \\ 0 & 0 & 0.08^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.02 & -0.02 \\ 0 & 0 & 0 & 0.10^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.02 & -0.05 \\ 0 & 0 & 0 & 0 & 0.07^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.02 & 0.02 \\ 0 & 0 & 0 & 0 & 0 & 0.09^* & 0 & 0 & 0 & 0 & 0 & 0 & 0.08 & 0.01 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.08^* & 0 & 0 & 0 & 0 & 0 & 0.05 & 0.10 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.09^* & 0 & 0 & 0 & 0 & -0.06 & -0.03 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.07^* & 0 & 0 & 0 & -0.02 & 0.00 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.08^* & 0 & 0 & 0.09 & 0.08 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.08^* & 0 & -0.11 & -0.09 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.10^* & 0.07 & 0.04 \end{pmatrix}$$

Model 10 defines a factor structure that jointly involves drift rates, boundary separation, and the two covariates. The first twelve rows of the loadings matrix apply the structure of the experimental design to the drift rate parameters, while the thirteenth row relates an additional latent factor to the boundary separation. The final two rows relate all seven latent factors to the personality covariates:

$(\Lambda|\mathcal{M}_{10})^T =$

$$\begin{pmatrix} 0.54^* & 0.54^* & 0.54^* & 0.54^* & 0.54^* & 0.54^* & 0.54^* & 0.54^* & 0.54^* & 0.54^* & 0.54^* & 0.54^* & 0 & 0.34^* & 0.36^* \\ 0.80^* & 0 & 0 & 0 & 0.68^* & 0 & 0 & 0 & 0.61^* & 0 & 0 & 0 & 0 & 0.05 & 0.04 \\ 0 & 0 & 0.76^* & 0 & 0 & 0 & 0.44^* & 0 & 0 & 0 & 0.56^* & 0 & 0 & -0.22 & -0.20 \\ 0 & 0 & 0 & 0.86^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.43^* & -0.39^* \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.88^* & 0 & 0 & 0 & 0 & 0 & -0.04 & 0.05 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.93^* & 0 & -0.15 & 0.01 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.54^* & -0.46^* & -0.51^* \end{pmatrix}$$

### Model fit indices

The DIC values for the ten models were:  $\text{DIC}_1 = -2196$ ,  $\text{DIC}_2 = -1959$ ,  $\text{DIC}_3 = -1771$ ,  $\text{DIC}_4 = -1989$ ,  $\text{DIC}_5 = -2265$ ,  $\text{DIC}_6 = -2281$ ,  $\text{DIC}_7 = 1772$ ,  $\text{DIC}_8 = -1858$ ,  $\text{DIC}_9 = -1944$ , and  $\text{DIC}_{10} = -2718$ .