RT-MPTs: Process Models for Response-Time Distributions Based on Multinomial Processing Trees with Applications to Recognition Memory

Karl Christoph Klauer[a]

[a]*Corresponding Author: Department of Social Psychology and Methodology, Albert-Ludwigs-Universität Freiburg, D-79085 Freiburg, Germany. Email: klauer@psychologie.uni-freiburg.de*

David Kellen[b]

[b]*Department of Psychology, Syracuse University, 430 Huntington Hall, Syracuse, New York 13244, USA. Email: davekellen@gmail.com*

Abstract

Multinomial processing tree models have been widely used for characterizing categorical responses in terms of a finite set of discrete latent states, and a number of processes arranged serially in a processing tree. We extend the scope of this model class by proposing a method for incorporating response times. This extension enables the estimation of the completion times of each process and the testing of alternative process orderings. In line with previous developments, the proposed method is hierarchical and implemented using Bayesian methods. We apply our method to the two-high-threshold model of recognition memory, using previously published data. The results provide interesting insights into the ordering of memory-retrieval and guessing processes and show that the model performs at least as well as established benchmarks such as the diffusion model.
Keywords: Multinomial Models, Response Times, Hierarchical Models

---

Response time is perhaps the most important measure used to investigate hypotheses about mental processes in experimental psychology, going back to the pioneering work of Franciscus Donders in the nineteenth century (for reviews, see Luce, 1986; Jensen, 2006; Townsend & Ashby, 1983; Van Zandt, 2002). From a data-analytic perspective, this predominance has raised important challenges: For instance, response time distributions are positively skewed, with longer tails on the right sight of the probability density function than on the left side. Moreover, reaction-time means and variances are often found to be linearly related (Wagenmakers & Brown, 2007). These two features alone are enough to see that response times do not mesh well with the general class of statistical linear models traditionally used to analyze data. As a response to this challenge, several approaches have been developed, which can be roughly divided into two research strands: One has focused on fitting response-time data to a suitable parametric distribution (e.g., the ex-Gaussian distribution; Matzke & Wagenmakers, 2009) in order to provide economical summaries of the data in terms of a few parameters (e.g., Schmiedek, Oberauer, Wilhelm, Süß, & Wittman, 2007; for an overview, see Balota & Yap, 2011). The second research strand, which is quite active to this date, instead focuses on the development of mathematical models as psychological accounts for the data in terms of specific mental processes that unfold across time (e.g., Brown & Heathcote, 2008; Ratcliff & Rouder, 1998; Townsend & Nozawa, 1995; for reviews, see Luce, 1986; Townsend & Ashby, 1983; Schweickert, Fisher, & Sung, 2012; Van Zandt, 2002).

Beyond response times, another widely applicable tool for the study of mental processes is given by the class of *multinomial-processing tree* models (MPT models; Riefer & Batchelder, 1988). MPT models characterize categorical (frequency) data in a given paradigm by postulating a finite set of latent states. For each item type, the observed responses are the outcome of a mixture of the different latent states and associated state-to-response mappings. The probability of each state being reached is generated from a *processing tree*, the edges of which represent the outcomes of different processes. MPT models are usually tailored to a particular experimental paradigm, with trees specifying the most important processes believed to be involved in the generation of responses. The family of process-dissociation models (Jacoby, 1991) used in many lines of psychological research (Klauer, Dittrich, Scholtes, & Voss, 2015) is one prominent member of the MPT model class among many others.

Figure 1 illustrates another simple and well-known MPT model, the *two-high-threshold model* (2HT) for recognition memory (Snodgrass & Corwin, 1988). In recognition-memory research, participants study a list of items and later see these items intermixed with new items. Their task is to decide for each item whether it was previously studied or not. The 2HT model assumes three latent states:

- $\mathcal{S}_1$: Item is detected as having been previously studied
- $\mathcal{S}_2$: Item is detected as being new.
- $\mathcal{S}_3$: The status of the item could not be determined

$\mathcal{S}_1$ and $\mathcal{S}_2$ are memory-certainty states that can only be reached by studied and non-studied items, respectively, whereas the uncertainty state $\mathcal{S}_3$ can be reached by both item types. Each item type is associated with a processing subtree. Each process in the subtree can complete with one of two possible outcomes, represented by two edges, one for each outcome. The likelihood of each of the two process outcomes is governed by a parameter assigned to the respective edge.[1]

For a studied item, participants first attempt to recognize the item, which succeeds with probability D_O and fails with probability $1-D_O$. In the former case, participants enter state $\mathcal{S}_1$ and respond "old". In the latter case, they enter the uncertainty state $\mathcal{S}_3$, which in turn triggers a guessing process. With probability g, the item is guessed as having been previously studied, resulting in an "old" response. With probability $1-g$, the items is instead guessed as being absent from the study list, leading to a "new" response. For a new item, participants again attempt to recognize the item, but according to the model, they cannot succeed in recognizing it. Instead, with probability D_N, participants can sometimes infer that the item is new, based on, for example, its overall dissimilarity from the studied items (e.g., Mewhort & Johns, 2000) or memorability expectations that were not met (e.g., Strack & Bless, 1994). A successful inference of this kind, which corresponds to state $\mathcal{S}_2$, leads to the response "new". If participants cannot infer a test item's true status, they enter the uncertainty state $\mathcal{S}_3$ and the same guessing process as described for old items is assumed to operate. Based on participants' recognition judgments, we can estimate the 2HT model's parameters and test some of its properties (e.g., Dube & Rotello, 2012; Kellen & Klauer, 2014, 2015; Kellen, Singmann, Vogt, & Klauer, 2015; Province & Rouder, 2012).

Although not as ubiquitous as response-time analyses, MPT models such as the 2HT model have been found useful in an enormous range of psychological inquiries (for reviews, see Batchelder & Riefer, 1999; Erdfelder et

[1]The two possible outcomes associated with each process are often referred to as "success" and "failure", respectively. This nomenclature is intuitive in some cases; e.g., succeeding/failing to retrieve an item from memory, but it is completely arbitrary in others; e.g., when referring to guessing processes.

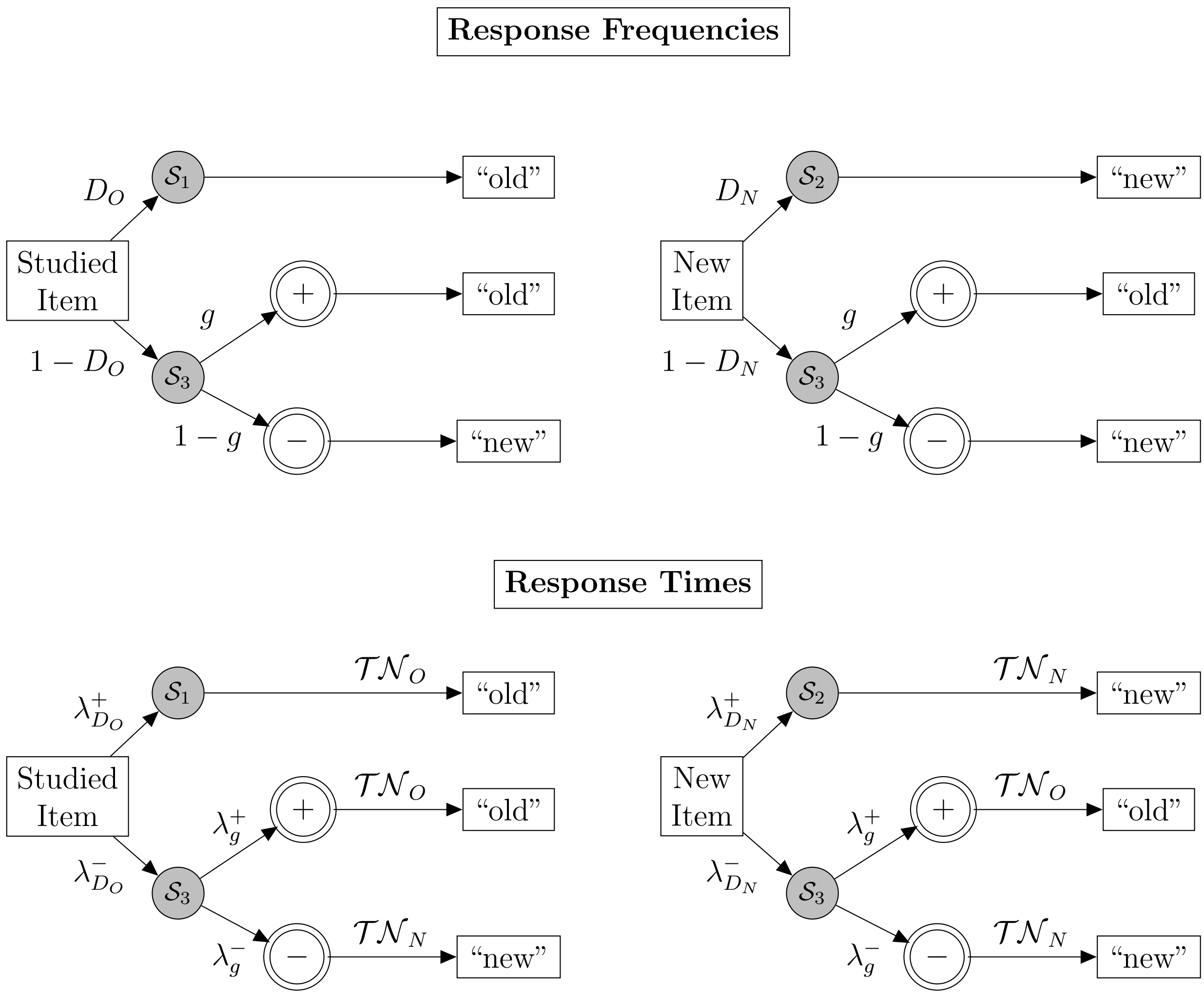


Figure 1: 2HT Model ("*Detect-Guess*" Variant). The rectangles indicate manifest states (i.e., item type and responses), and the gray circles indicate the latent states $\mathcal{S}$. The double circles provide a reference for the outcome of the guessing process, terminating with either an "old" or a "new" guess (denoted by + and −, respectively). The parameters D_o, D_n, and g refer to the probabilities of, in order, detecting an old item as old, a new item as new, and guessing "old". The λ parameters are exponential-rate parameters governing the time for these processes to complete for each outcome. $\mathcal{TN}_O$ and $\mathcal{TN}_N$ refer to truncated normal distributions governing encoding and response-execution times for response "new" and " old", respectively. Note that only nodes with process parameters attached to outgoing edges count as nodes in the technical sense detailed in Section 2.2. (i.e., the root node and the node for $\mathcal{S}_3$).

al., 2009; Hütter & Klauer, 2016). In addition to the ongoing stream of proposals for new substantive models (e.g., Gawronski, Conway, Armstrong, Friesdorf, & Hütter, 2016; Meissner & Rothermund, 2013), the current fruitfulness of this model class is attested by its growing methodological toolbox: From hierarchical and mixed-model extensions (Klauer, 2010; Matzke, Dolan, Batchelder, & Wagenmakers, 2015), to sophisticated model-selection indices (Klauer & Kellen, 2015; Wu, Myung, & Batchelder, 2010), and inequality-constraint applications (Klauer, Singmann, & Kellen, 2015).

But as useful as this model class may be, there are limits to what one can achieve on the basis of response frequencies alone. For instance, the characterization of the observed responses as a function of a mixture of latent states is ultimately silent about the duration of each of the processes that govern the access to these states, as well as about their exact order. Let us go back to the 2HT example: The tree structure of the model illustrated in Figure 1 suggests that the guessing responses occur after a failed attempt to recognize the test item. We can refer to this as the "*detect-guess*" variant of the 2HT. Alternatively, we can conceive 2HT models in which guessing occurs *prior* to any attempt to recognize the test item. For example, according to a "*default-interventionist*" variant of the 2HT shown in Figure 2, the participant first guesses whether the test item is old or new.[2] Only after this process is completed does the participant engage in a memory-retrieval process that — if successful — takes precedence over the previously established guess. Although the original detect-guess and the default-interventionist variants of the 2HT are very distinct in terms of mental-processing assumptions, they are formally equivalent in the sense that they yield the exact same range of predictions with the same parameters and parameter values.

[2] In this respect, the default-interventionist variant resembles the diffusion model of recognition memory (Dube, Starns, Rotello, & Ratcliff, 2012) in which a starting point for the diffusion process, governing response biases, is set prior to any attempt to recognize the test item.

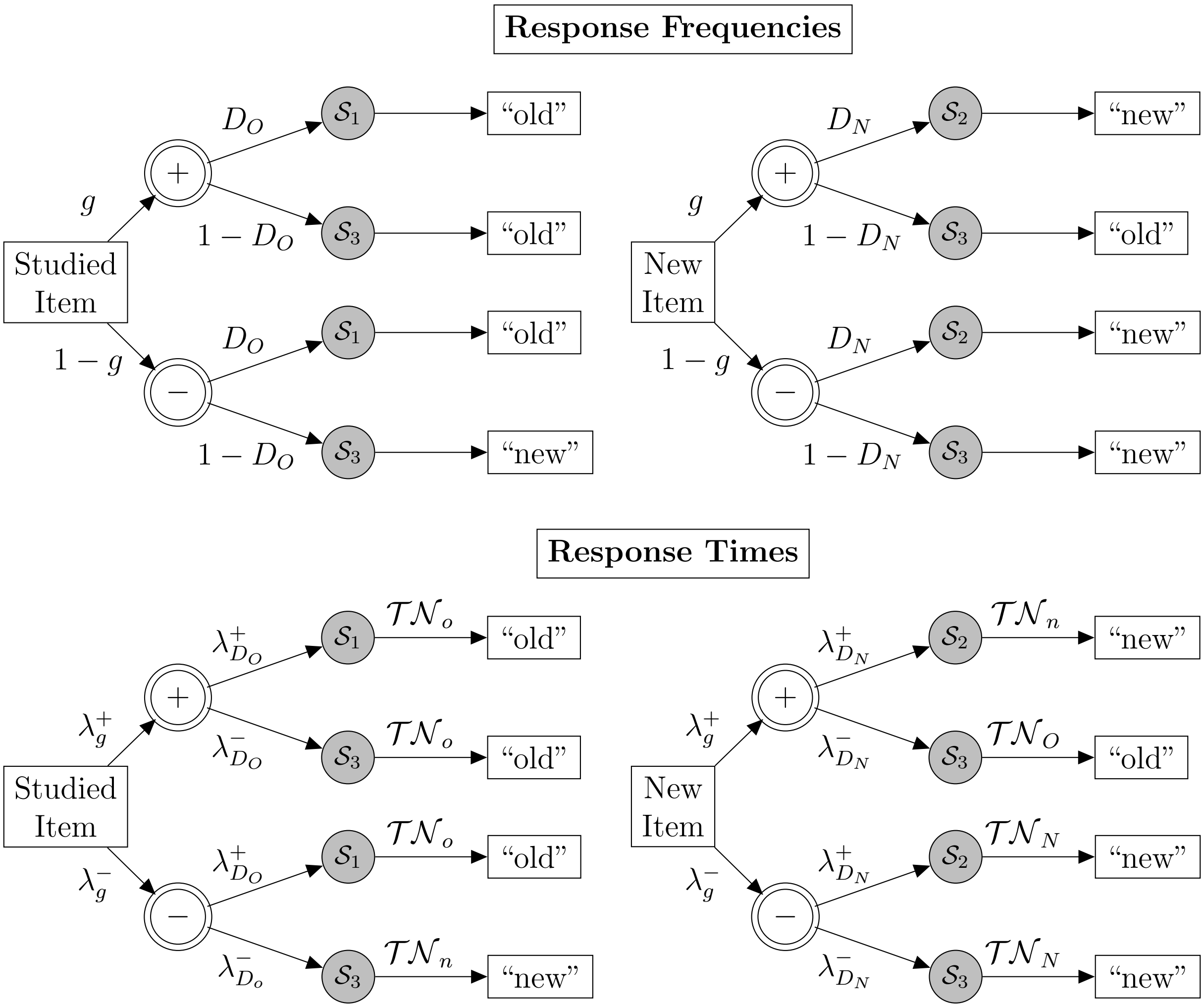


Figure 2: 2HT Model (“*Default-Interventionist*” Variant). The rectangles indicate overt states (i.e., item type and responses), and the gray circles indicate the latent states $\mathcal{S}$. The double circles provide a reference for the outcome of the guessing process, terminating with either an “old” or a “new” guess (denoted by + and −, respectively). The parameters are described in the caption of Figure 1. Note that only nodes with process parameters attached to outgoing edges count as nodes in the technical sense detailed in Section 2.2. (i.e., the root node and the double circles).

In order to test alternative tree structures and estimate the duration of the different processes they postulate, we need to extend the MPT model so that it can provide a joint characterization of response frequencies and their respective times. We present a framework within which different distributional assumptions can be explored and fully elaborate one model in this framework. The method can be applied to data from any identified MPT model if response times have been recorded along with the responses. It even extends the class of models that can be investigated because models that are not mathematically identified based on only the response frequencies can become so by the inclusion of the response times as illustrated below.

In line with modern developments, the new method is hierarchical, which is advantageous in two different ways: First, it enables models to accommodate the substantial heterogeneity that usually stems from individual differences. These differences can be expressed in many ways such as accuracy, response bias, response speed, and speed-accuracy trade-offs, among others. Second, each participant's parameter estimates are thereby informed by the other participants' estimates, which is particularly valuable when the data per individual are sparse, and group-level estimates generalize across them (Katahira, 2016; Klauer, 2010). In terms of substantive psychological applications, the proposed method can provide us with deeper insights into the architecture of processes in the modeled tasks, enabling the testing of theoretical predictions regarding the temporal characteristics of the involved processes, and the development of more diagnostic measurement models. In light of these contributions, the proposed method can be cast as a principled alternative to the currently dominant paradigm of simple diffusion-model accounts.

1. Precursors and State of the Art

Relatively direct precursors of the present approach are mixture models postulating probabilistic mixtures of response-time distributions (e.g., Falmagne, 1965; Ollman, 1966; Yantis, Meyer, & Smith, 1991). Mixture models are naturally related to MPT models if each processing-tree path terminating in an observable response category is considered as giving rise to a distinct path-specific response-time distribution. Another important precursor is the study of complex cognitive architectures that followed the seminal work of Sternberg (1969). For instance, Schweickert (1978) showed how the selective influencing of processes could be used to gain insights into their arrangement (e.g., are two processes serial or concurrent?) along an acyclical network (see also Dzhafarov & Schweickert, 1995; Goldstein & Fisher, 1991; Townsend & Nozawa, 1995; for a review, see Schweickert et al., 2012).

Hu (2001) applied some of these earlier ideas in the context of the MPT model class, noting that any tree structure belonging to this class is a special type of acyclical network. Specifically, Hu investigated under what conditions it is possible to decompose mean response latency for each response category into mixtures of different path-specific means, which are further decomposed into additive components for each edge along the path, each such component describing the mean completion time of the process (and process outcome) associated with it. The inclusion of response times under this framework enabled Hu to compare two variants of the famous high-threshold source-memory model (Batchelder & Riefer, 1990) that postulate distinct relationships between the retrieval of item and source memory. Among the limitations of Hu's approach is the fact that it only operates at the level of mean response times, ignoring higher-order moments of the response-time distributions. In addition, the method does not develop tools for statistical inference. Finally, it assumes that process-completion times and the probability of a process succeeding are independent, a highly implausible assumption.

An approach based on mixture models was recently developed by Heck and Erdfelder (2016). It is based on setting up a number of response-latency bins: For example, when using two bins the responses might be split into fast and slow responses based on the overall median. A new branching is then added at the end of each path of the MPT model, distributing the responses generated along that path into the different bins. The probabilities associated with each new binary branching are governed by a path-specific parameter L. For example, in the case of the 2HT, the "old" responses emerging from state $\mathcal{S}_1$ are mapped into 'slow old' and 'fast old' bins with probabilities L_{D_O} and $1 - L_{D_O}$, respectively. The L parameters for the different paths order them in terms of their relative speed, using the overall median as benchmark.

One key advantage of Heck and Erdfelder's (2016) approach is that the binning of response times sidesteps the need to impose any parametric assumptions on the shape of the response-time distributions. Additionally, the models resulting from their approach are still members of the MPT model class as formalized by Hu and Batchelder (1994), which means that the entire methodological toolbox developed so far for MPTs can be applied without any modification. One limitation, however, is that the models resulting from this approach will often not be identified, with pathwise L parameters that cannot be uniquely estimated. Additional simplifying assumptions are needed to achieve identifiability: For example, for the 2HT model, Heck and Erdfelder assume that "old" responses to old items generated by a recognition failure and an "old" guess (with probability $[1 - D_O]g$) have the same response-time distribution as "old" responses to new items generated by a failure to infer that they are new and followed by an "old" guess (with probability $[1 - D_N]g$). An analogous assumption is made for recognition and inference failures that are followed by a "new" guess. These simplifying assumptions allow for the L parameters of the corresponding paths to be set equal, yielding an identifiable model.

Further simplifying assumptions are usually required to link the path-specific L parameters to individual processes along the different paths. For example, for the 2HT model, the L parameter attached to the processing path $(1 - D_O)g$ is interpreted as capturing the relative speed of guessing "old" although in terms of processes, the failure to recognize (with probability $1-D_O$) is also involved. Finally, it is difficult in this kind of model to integrate temporal differences across stimuli, responses, or experimental conditions due to encoding and/or response-execution processes. Such components are routinely accommodated by most response-time models usually in the form of an additive response-time component t_0 (Luce, 1986). Nevertheless, where applicable the approach by Heck and Erdfelder (2016) is relatively easy to use and has the potential to provide valuable insights into the relative speed of cognitive processes involved in the generation of the observed responses. These advantages have been spelt out in a large application by Heck and Erdfelder (2017).

The present approach pursues an alternative but complementary route by imposing specific parametric assumptions on the processes along with a serial interpretation of the processing paths as describing a succession or at least a cascade of processing steps (McClelland, 1979). In this latter respect, the present approach builds on the one by Hu (2001) and follows the intuition of many modelers in formulating MPT models. The downside of imposing parametric assumptions and a serial or cascade interpretation of processing paths is that the resulting method cannot be expected to be useful in situations where there are substantive deviations from such assumptions. More generally speaking, parametric response-time models such as the diffusion model (Ratcliff & Rouder, 1998) are most interesting where they provide a good description of observed data. Such an outcome implies that the model and its ensemble of assumptions provide one viable theoretical account of the data. A failure to describe given data is less interesting, inasmuch as it is often difficult to diagnose whether the failure goes back to a violation of central structural assumptions (e.g., the idea of a diffusion process or a particular MPT model architecture) or to a violation of ancillary parametric assumptions (e.g., the assumption of normally distributed residual encoding and response-execution times t_0). For this reason, it is desirable to develop these models for different sets of parametric assumptions, but in the present manuscript we focus on only one such set.

As will be shown below, the approach proposed here provides complete descriptions of the observed joint distribution of responses and response times, including accounts of the differences between individuals, and the correlations between the parameters governing responses and response times related to these individual differences. The methods presented here can be applied to any identifiable MPT model as well as to many MPT models that are not identified based on only the response frequencies without imposing further restrictions.

2. Model Assumptions

2.1. Overview

In its simplest form, the present method builds on binary multinomial processing trees (only two branches go out from each node) in which each processing path represents a succession of processing stages for which process-completion times add up (Sternberg & Backus, 2015). An additional additive component summarizes encoding and response-execution times. The process-completion times of different processes are assumed to be independently distributed (i.e., conditional independence is assumed), and process-completion time distributions leading to the same outcome of the same process are assumed to be identical wherever that process occurs in the processing tree.

For example, in the bottom part of Figure 1, the process of guessing "old" associated with parameter g in the tree for old items has two completion-time distributions depending upon whether it completes successfully or not. The same process occurs in the tree for new items, where it is assumed to have the same completion-time distributions.

A complete model of response-time distributions needs specific parametric assumptions (Van Zandt, 2002), where limited data are available per participant. Consider three sets of assumptions in increasing order of psychological plausibility and decreasing order of tractability:

1. Each completion-time distribution is exponentially distributed with a separate rate parameter λ for each process outcome (see the bottom part of Figures 1 and 2). The distribution of encoding and response-execution times follows a truncated Gaussian distribution $\mathcal{TN}$ (truncated so that only positive values can occur). There are separate parameters for the probabilities with which the processes complete with either of the two outcomes.
2. Like Option 1, but each completion-time distribution follows a Wald distribution (first-passage time distribution of a Brownian motion with positive drift) with separate threshold and drift-rate parameters. The encoding and response-execution times follow an exponential distribution.
3. Each process and its joint distribution of outcomes and completion times is modeled by a diffusion model with a minimal set of parameters. There is a separate distributional assumption for the distribution of

encoding and response-execution times.

Further options can of course be considered such as that processes complete as the outcome of a race between separate counters. The options listed here were motivated by models figuring importantly in the literature on response-time distributions. Option 1 is motivated by the ex-Gaussian distribution often used for modeling response times (Matzke & Wagenmakers, 2009), whereas Option 2 is motivated by the ex-Wald distribution proposed by Schwarz (2001). There are critical discussions of these distributions that focus on their empirical adequacy as well as on their psychological plausibility (e.g., Burbeck & Luce, 1982; Matzke & Wagenmakers, 2009; Sternberg & Backus, 2015). Most of these criticisms are dealt with through Option 3, motivated by the diffusion-model literature, which is however in all likelihood the most difficult to implement mathematically.[3] Note that Option 1 models, permitting different exponential-rate parameters for each process outcome, accommodate reaction-time distributions that differ as a function of response category. This means that none of the models is bound to assume that responses and their respective times are independent, a (problematic) property known as *separability* (Marley & Colonius, 1992; see also Brown, Marley, & Heathcote, 2012).

In the present manuscript, we develop a Bayesian hierarchical version of the Option 1 model with participants as random effects. Thus, each participant is associated with different parameters for the process probabilities and completion-time distributions. These parameters are constrained by a

[3] Although in its simplest form, the proposed approach assumes sequential additive stages along each processing path, it can easily accommodate technical parameters in any given path (e.g., Klauer, Singmann, et al., 2015) to which no completion-time component is attached. Furthermore, stages with overlapping or parallel processes can often be modeled by concatenating two sequential stages and assigning only one common completion-time distribution to them.

prior distribution with population means and variance-covariance structure, parameters that are also estimated from the data. The resulting models provide process-oriented accounts of the joint distribution of response categories and response times.

2.2. The Person-Level Model for Responses and Response Times

MPT models usually consist of several subtrees. For example, in the most simple 2HT model, there are two subtrees, one for trials involving studied items, and one for trials involving new items. Each tree has internal nodes, referred to simply as nodes in the following, and leaves. The leaves correspond to observable response categories such as "old" or "new". Categories for different subtrees are, however, considered different categories. For example, the category "old" is also indexed by the subtree from which it stems and thus, to be precise, there are four response categories in the most simple 2HT model. The categories are mapped on actual responses such as left or right keypresses so that responses are a function of categories. For example, the two "old" response categories may be mapped on the left key, and the two "new" response categories on the right key. We consider only binary MPT models in which each node has two children. But note that non-binary MPT models can be transformed into binary MPT models (Hu & Batchelder, 1994; see Appendix for more details).

To each node n in the tree, a process $p = p(n)$ is attached with two outcomes. For example, p might be a guessing process with two outcomes 'guess old' versus 'guess new'. The two outcomes correspond to two edges going out from the node, and we will refer to an outcome o generically as the '+' outcome, $o = +$, or the '−' outcome, $o = -$. Furthermore, let $\text{plus}(o)$ be a function of the outcome with $\text{plus}(+)=1$ and $\text{plus}(-)=0$.

Under Option 1, each process is characterized by three parameters: θ_p, λ_p^+, and λ_p^-. Parameter θ_p models the probability that the process p completes with the '+' outcome with completion time governed by the exponential rate parameter λ_p^+. The probability of the '−' outcome is given by $1-\theta_p$ with completion time described by parameter λ_p^-.

We consider paths B from root to one of the leaves and represent them in terms of the internal nodes n traversed from root to leave along with the outcomes, $+$ or $-$, attached to the edges along the path and thus, as a set of edges (n,o). The probability of path B is a product of the MPT model parameters, such as D_o and g, and their complements, as encountered along the path. Thus,

$$P(B) = \prod_{(n,o)\in B} \theta_{p(n)}^{\text{plus}(o)} (1-\theta_{p(n)})^{1-\text{plus}(o)}. \tag{1}$$

The response latency of a response generated along path B is the sum of an encoding and response-execution time δ, that follows a truncated normal distribution, and exponentially distributed process-completion times of the processes along that path. It is reasonable to assume that the encoding and response-execution time δ with mean γ and variance σ^2 depends on the particular motor response r, $r = 1,\ldots,R$, such as a left or right keypress[4] and thus, we permit its mean $\gamma = \gamma_r$ to differ between different responses.

Each path B ends in a category $c = c(B)$, which is mapped on a response $r = r(c)$. For a path B with only one node n, attached process $p = p(n)$, and outcome o that leads to response $r = r(c(B))$, the distribution of response latency thus follows the familiar ex-Gaussian distribution with truncation at zero carried over from the truncated normal component:

[4] For example, responses made with the right hand are often executed faster than left-handed responses by persons with dominant right hand; frequent responses are often executed faster than infrequent responses, etc. (see Voss, Voss & Klauer, 2010).

$$\begin{aligned} f(t|B) &= \text{ex-Gauss}_{\geq 0}(t|\lambda_p^o, \gamma_r, \sigma^2) \\ &= \lambda_p^o \exp\left[\lambda_p^o(\gamma_r + \frac{1}{2}\lambda_p^o\sigma^2 - t)\right]\left\{\Phi\left(\frac{t-\gamma_r-\lambda_p^o\sigma^2}{\sigma}\right) - \Phi\left(\frac{-\gamma_r-\lambda_p^o\sigma^2}{\sigma}\right)\right\} \\ &\times \left(1-\Phi\left(\frac{-\gamma_r}{\sigma}\right)\right)^{-1}. \end{aligned} \tag{2}$$

For longer paths, the response latencies are the sum of several exponentials and a truncated normal. The sum of exponentials is distributed as what is known as a hypoexponential or generalized Erlang distribution (Johnson, Kotz, & Balakrishnan, 1994), which after convolution with the truncated normal yields the following density:[5]

$$f(t|B) = \sum_{(n,o)\in B}\left(\text{ex-Gauss}_{\geq 0}(t|\lambda_{p(n)}^o, \gamma_r, \sigma^2) \prod_{(m,q)\in B,(m,q)\neq(n,o)} \frac{\lambda_{p(m)}^q}{\lambda_{p(m)}^q - \lambda_{p(n)}^o}\right). \tag{3}$$

Hence, the joint distribution of categories c for a given subtree (such as the category "new" for old items) and response latencies t is characterized by

$$f(c,t) = \sum_{B:B \text{ ends in } c} f(t|B)P(B). \tag{4}$$

[5]This formula assumes that the same process and outcome is not attached to two different nodes in the path. It needs to be modified if the same process is repeated along a path.

2.3. Priors and Hyperpriors

2.3.1. Priors for process-related parameters.

The parameters of the above person-level model, θ_p, λ_p^o, γ_r, and σ^2, can assume different values for each individual s and thus, they carry the additional index s, which has so far been suppressed for ease of exposition. Like in Klauer (2010), the person-level MPT parameters $\theta_{p,s} \in (0,1)$ are transformed via an inverse-probit link to the real line, yielding new parameters $\alpha_{p,s} = \Phi^{-1}(\theta_{p,s})$, where Φ is the cumulative distribution function of a standard normal distribution. Analogously, the exponential rate parameters $\lambda_{p,s}^o \geq 0$ are transformed via a log link to the entire real line, yielding transformed parameter $\beta_{o,p,s} = \log(\lambda_{p,s}^o)$.

Furthermore, we decompose the person-level parameters into the sum of a population mean μ and (zero-centered) person-level deviations from that mean:

$$
\begin{aligned}
\alpha_{p,s} &= \mu_p^{(\alpha)} + \alpha'_{p,s}, \\
\beta_{o,p,s} &= \mu_{o,p}^{(\beta)} + \beta'_{o,p,s}, \\
\gamma_{r,s} &= \mu_r^{(\gamma)} + \gamma'_{r,s}.
\end{aligned} \tag{5}
$$

The parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ characterize the accuracy and speed, respectively, of the processes specified in the MPT model. It is reasonable to assume that they can correlate across persons and thus, we assume that they are jointly distributed as a multivariate normal with variance-covariance structure $\boldsymbol{\Sigma}$ that is to be estimated from the data. Stacking the person-level deviations into vectors $\boldsymbol{\alpha}'_s = (\alpha'_{p,s})_{(p=1,\ldots,P)}$ and $\boldsymbol{\beta}'_s = (\beta'_{o,p,s})_{(o=+,-\,;\,p=1,\ldots,P)}$, their prior distribution is given by

$$\begin{pmatrix} \boldsymbol{\alpha}'_s \\ \boldsymbol{\beta}'_s \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}_{3P}, \boldsymbol{\Sigma}\right), \tag{6}$$

where $\mathbf{0}_n$ is a vector of zeros of length n and P is the number of different processes in the MPT model. This structure constrains the person-level deviations to follow a normal distribution with variances and covariances estimated from the data.

Person-level statistical information is thereby aggregated in the population means $\boldsymbol{\mu}^{(\boldsymbol{\alpha})} = (\mu_p^{(\alpha)})_{(p=1,\ldots,P)}$ and $\boldsymbol{\mu}^{(\boldsymbol{\beta})} = (\mu_{o,p}^{(\beta)})_{(o=-,+\,;\,p=1,\ldots,P)}$ for each parameter. Variances across persons, and correlations between parameters for different processes, are estimated by $\boldsymbol{\Sigma}$.

2.3.2. Hyperpriors for process-related parameters.

Hyperpriors are needed for the prior parameters $\boldsymbol{\mu}^{(\boldsymbol{\alpha})}$, $\boldsymbol{\mu}^{(\boldsymbol{\beta})}$, and $\boldsymbol{\Sigma}$. Like in Klauer (2010), the hyperprior for $\boldsymbol{\mu}^{(\boldsymbol{\alpha})}$ is a normal distribution

$$\boldsymbol{\mu}^{(\boldsymbol{\alpha})} \sim \mathcal{N}(\mathbf{0}_P, \epsilon^{-1}\boldsymbol{I}_P)$$

where $\boldsymbol{I}_P$ is the $P \times P$ identity matrix and the precision ϵ is set to one here and below for the analyses presented in this paper, implying a uniform prior distribution for the $\mu_p^{(\alpha)}$ on the probability scale. The hyperpriors for parameters $\mu_{o,p}^{(\beta)}$ are specified on the original (not log-transformed) scale and thus in terms of parameters $\exp(\mu_{o,p}^{(\beta)})$ as independent Gamma distributions with shape and rate parameters set to 1.0 and 0.1 for the analyses presented below, implying a mean and variance of 10 and 100, respectively. A mean of 10 was chosen because a mean exponential rate of 10 implies a mean process-completion time of 0.1 s or 100 ms, which seemed a reasonable prior setting for process-completion times for the applications considered in this paper. A variance of 100 ensures that the hyperprior is nevertheless reasonably uninformative.

Following Klauer (2010), the hyperprior for $\mathbf{\Sigma}$ is a scaled Inverse-Wishart distribution with $3P + 1$ degrees of freedoms, scale matrix $\boldsymbol{I}_{3P}$, and scale factors $\boldsymbol{\xi}^{(\alpha)}$ and $\boldsymbol{\xi}^{(\beta)}$ for parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively. As discussed by Gelman and Hill (2007, Chap. 13), setting the degrees of freedom to $3P+1$, that is, one plus the number of parameters in the multivariate normal distribution of Equation 6, has the effect of imposing a uniform prior distribution on the individual correlation coefficients implied by $\mathbf{\Sigma}$, a reasonably uninformative prior setting for the correlations. This distribution does, however, impose stronger constraints on the variances. To relax these constraints as well as to speed up convergence in Markov Chain Monte Carlo (MCMC) estimation, Gelman and Hill (2007) propose to use a *scaled* Inverse-Wishart distribution in which unidentified scale parameters $\boldsymbol{\xi}$ are introduced. Details are described in the Appendix.

2.3.3. Priors and hyperpriors for encoding and response-execution times.

The person-level model describes the encoding and response-execution times δ in terms of person-level means $\gamma_{r,s}$ and variances σ_s^2. It is reasonable to assume that the $\gamma_{r,s}$ pertaining to different responses are correlated across persons. Thus, we assume a multivariate normal distribution with population means $\mu_r^{(\gamma)}$ and variance-covariance matrix $\mathbf{\Gamma}$ as prior. The hyperpriors for $\mu_r^{(\gamma)}$ are again independent normal distributions with zero mean. For the application reported below, we chose a variance of 10 for these hyperpriors, which for response latencies in the range of at most a few seconds seemed to be a reasonably uninformative choice for the variance of person-level mean values, especially when considering that the variances of means are necessarily smaller than variances for individual latencies. The hyperprior for $\mathbf{\Gamma}$ is the above-discussed scaled Inverse-Wishart distribution with $R + 1$ degrees of freedom and scale matrix $\boldsymbol{I}_R$, R being the number of different responses that can occur.

For the variances σ_s^2, a scaled inverse chi-squared distribution with scale factor ω^2 and $df = 2$ was chosen as prior, which again imposes few constraints on the variances. For ω^2, an improper uninformative prior was chosen, $p(\omega^2) \propto \frac{1}{\omega^2}$. The posterior estimate of population-level parameter ω^2 provides an overall estimate of the residual variance in response latencies that is not accounted for by the model.

3. Algorithm

The resulting model does not fall into the scope of standard software for MCMC estimation such as JAGS (Plummer, 2003), primarily because the kernel density in Equation 3 is not implemented in such software (but see Annis, Miller, & Palmeri, 2017). We devised a Gibbs sampler, more precisely a Metropolis-within-Gibbs sampler, for fitting the model that uses three steps of data augmentation.[6] The first two steps of data augmentation constitute an adaptation of the approach by Albert and Chib (1993) to replace observed categorical responses by an underlying Gaussian structure tailored to the special structure of multinomial processing trees (Klauer, 2010).

The first step is to augment each observed category c by the path B along which it was generated. Let $\mathcal{B}_c$ be the set of paths B that end in category c. The probability that an observed category c and response latency t was generated by a specific member B of $\mathcal{B}_c$ is then given by

$$p(B \mid c, t) = \frac{p(B) f(t \mid B)}{\sum_{B' \in \mathcal{B}_c} p(B') f(t \mid B')}, \tag{7}$$

[6]Program (C++) scripts of this implementation calling the NAG library can be obtained from the first author.

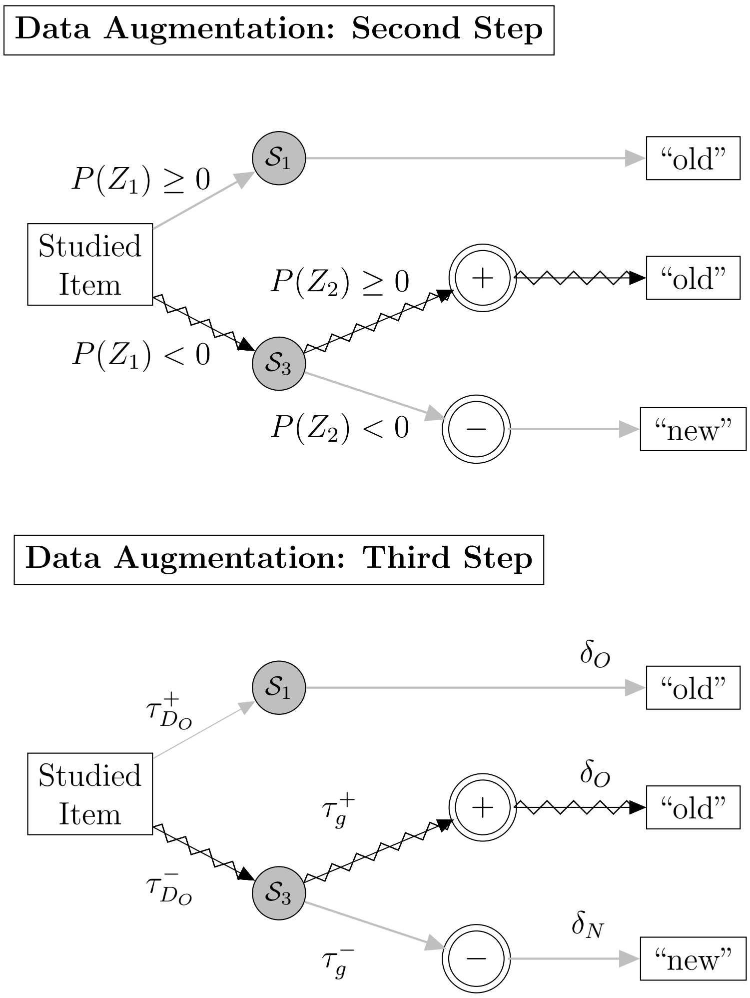
where $p(B)$ and $f(t \mid B)$ are given by Equations 1 and 3, respectively. Thus, the observed data of each trial can be augmented given the person-level parameters by sampling a path B from $\mathcal{B}_c$ based on a multinomial distribution with the above $p(B \mid c, t)$ as parameters.

The second step of data augmentation is illustrated in Figure 3. For each trial from each subtree of the model administered to person s, values $z_{n,s}$ of latent normal variables $Z_{n,s}$ are sampled for each node n of the subtree, given the person-level parameters $\boldsymbol{\alpha}_s$ and the path B sampled in the first step. The $z_{n,s}$ are sampled from normal distributions with means $\alpha_{p(n),s} = \Phi^{-1}(\theta_{p(n),s})$ and variance one. For (n, o) in B, the $Z_{n,s}$ variable is truncated from below at zero with $Z \geq 0$ for the $+$ outcome, $o = +$, and truncated at zero from above with $Z < 0$ for the $-$ outcome, $o = -$ (Albert & Chib, 1993). For edges in the subtree but not on the path B, the Z variate is not truncated.[7] The helpful aspect of this double data augmentation, by paths B and Z variates, is that it allows us to estimate the person-level parameters $\boldsymbol{\alpha}$ as in a standard hierarchical linear model in a Bayesian framework as elaborated in the Appendix.

The third step of data augmentation is also illustrated in Figure 3. For each trial from each subtree of the model administered to person s, latent process-completion times $\tau^o_{n,s}$ are generated for each edge (n, o) of the subtree along with residual encoding and motor-execution component $\delta_{r(c(B)),s}$, given the person-level parameters $\boldsymbol{\alpha_s}$, $\boldsymbol{\beta_s}$, $\boldsymbol{\gamma_s}$, and σ_s as well as the path B sampled for the trial on the basis of the category c and the response latency

[7]As discussed by Klauer (2010), these non-constrained Z variates seem redundant. Leaving them out, the number of such variates is, however, itself a random variable that takes on different values in each cycle of the resulting sampler, which in consequence would no longer be a Gibbs sampler. In consequence, the strong convergence results known for the MCMC sampler (Gill, 2008, Chap. 9) do no longer apply, and a new theoretical foundation would be required to justify the algorithm without the unconstrained Z.

For a given response "old" to a studied item with latency t:



With $t = \tau^-_{D_O} + \tau^+_g + \delta_O$

Figure 3: Illustration of the second and third steps of data augmentation. The zigzag arrows indicate the tree path that was sampled. In the second step of data augmentation (upper panel), the multinomial probability parameters are encoded in terms of (truncated) normal variates. In the third step, process completion times τ and encoding and response-execution times δ are added. For further details. see the body of text.

t that was observed on that trial. For edges (n, o) in the subtree, but not on the path B, the process-completion time is sampled from an exponential distribution with rate parameter $\lambda^o_{p(n),s} = \exp(\beta_{o,p,s})$. For links on the path B, sampling is constrained by the fact that the sums of τ-values and residual component δ along the path B have to add to the observed response latency (see Appendix for details). The helpful aspect of this data augmentation is that it allows us to estimate the exponential-rate parameters and the parameters governing the encoding and response-execution times as though we had directly observed the process-completion times τ and residual components δ.

With the observed and augmented data, most of the conditional distributions of the Gibbs sampler turn out to stem from relatively standard families of distributions for which it is easy to generate random values. One exception is the just-mentioned constrained sampling of process-completion times τ and residual component δ, which required a rejection-sampling step. A second exception is the sampling of the exponential-rate parameters $\lambda^o_{p,s}$, which required an adaptive rejection-sampling step (Gilks & Wild, 1992). A third exception concerns the sampling of parameters related to the encoding and response-execution times, which required Metropolis-Hastings steps. The joint model likelihood and details on the Gibbs sampler are provided in the Appendix.

4. Identifiability and Model Checks

4.1. Identifiability

The resulting model is identifiable whenever the underlying MPT model is identified. From Equation 4, it is easy to see that the distribution of response latencies given category c is a mixture of hypoexponential distributions convoluted with a truncated normal distribution with mixture weights given by $\frac{P(B)}{\sum_{B':B' \text{ ends in } c} P(B')}$, where B is a path that ends in c. If the multinomial model is identified, these mixture weights are identified (see Equation

1). Even without this extraneous identification stemming from the categorical responses, mixtures of members of parameterized families of continuous distributions are usually identified as shown by Titterington, Smith, and Makov (1985, Chap. 3). One way to see this is to note that expressing the $n-$th moments of the predicted response time distribution in terms of model parameters will usually yield as many non-redundant equations relating the model parameters to the moments as there are independent model parameters. Conversely, this implies that the model parameters are identified only from the higher-order moments of the RT distributions although they are further constrained by structural constraints as discussed below.

The extraneous identification of the mixture weights even deals with a remaining problem known as the *labeling problem*: When members from the same family of distributions (such as normal distributions) are mixed, the different components are exchangeable. That is, the order in which the different components are mixed makes no difference for the resulting mixture distribution. Hence, we can permute mixture weights and parameter values of the associated mixture components without changing the probability distributions predicted by the model. This possibility is however preempted when mixture weights are already identified from the categorical data alone.

These labeling problems again need to be considered where the underlying model is not identified. For example, in the basic 2HT model applied to data from a basic recognition-memory experiment, a model with parameters D_N, D_O and g will not be identified when only categorical "old"/"new" responses are available. Mixtures of pathwise response-latency distributions, including mixture weights, will nevertheless usually be identified. Even the labeling problem will usually not pose a problem, because the paths give rise to distinct families of distributions as a function of the number of edges upon them. For example, in the 2HTM, the "new" category for new items is reached via two paths, one with one edge labeled by D_N, the other one with two edges, labeled $1-D_N$ and $1-g$, respectively. The first path generates an ex-Gaussian distribution, the second a hypoexponential distribution with two exponential components convoluted with a truncated normal distribution. The labeling problem arises only if (at least some of) the different mixture components stem from families of distributions with overlap (i.e., sharing some distributions) and if such components fall into this area of overlap. There is, however, no overlap between the ex-Gaussian family of distributions and the distribution of the sum of two exponential variates and a truncated Gaussian variate. It follows that the mixture is completely identified with the relabeling possibility ruled out and hence that the mixture weights D_N and $(1 - D_N)(1 - g)$ are also identified in this case (and, mutatis mutandis, D_O and $[1 - D_O]g$). Analogous arguments show that most RT-MPT models will be identified even if the underlying MPT model is not. It is, however, possible to construct special cases in which the resulting model still suffers from the labeling problem.

How the observed reaction times are carved up into process-completion times and response-execution components depends in part on the distributional assumptions. In this respect, RT-MPTs are no different from other process-oriented models of reaction times such as the diffusion model (Jones & Dzahafarov, 2014; Heathcote, Wagenmakers, & Brown, 2014). The issue is somewhat mitigated by the fact that the same process-completion components and response-execution components appear in different combinations in different processing paths of the model, imposing structural constraints on these components and associated parameters. Nevertheless, due to such issues, estimation of the model should be accompanied by assessments of model fit as well as by selective-influence studies. In selective-influence studies manipulations are implemented that are believed to affect only one process on theoretical grounds. The model is thereby validated not only in terms of fit, but also in terms of whether or not it maps the effects of the manipulation on only this process (Heathcote, Brown, & Wagenmakers, 2015; Klauer, Stahl, & Voss, 2012). Success in these validation steps will increase one's confidence in the assumption that the parameter estimates are not unduly biased by inappropriate distributional assumptions. We illustrate both validation steps in the application below.

4.2. Model Checks

One way to assess the adequacy of a model for describing a given dataset is to conduct posterior predictive model checks (for an overview, see Gelman & Shalizi, 2013). Given the data $\mathbf{x}$, new data $\mathbf{x}^{\text{pred}}$ can be generated from the predictive posterior distribution. Referring to the collection of person-level parameters as $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_S)$, such model checks are based on a goodness-of-fit quantity $T(\mathbf{x}, \boldsymbol{\eta})$ defined as a function of the (new or old) data and of the parameters such that it quantifies specific deviations of the data from the model predictions. The probability p of $T(\mathbf{x}^{\text{pred}}, \boldsymbol{\eta}) > T(\mathbf{x}, \boldsymbol{\eta})$ is then estimated under the joint distribution of $(\mathbf{x}^{\text{pred}}, \boldsymbol{\eta})$ given the observed data. The model is considered adequate with regard to the deviations quantified by T if p is not small. Specifically, for each of the retained parameter sets generated via the MCMC sampler, a new dataset of the same size, $\mathbf{x}^{\text{pred}}$, is generated from the distribution specified in Equation 4, and the estimate of p is simply the proportion of cases with $T(\mathbf{x}^{\text{pred}}, \boldsymbol{\eta}) > T(\mathbf{x}, \boldsymbol{\theta})$.

One way to think of this procedure is by analogy to parametric Bootstrap assessments of model fit (Efron, 1982). The observed value of a test statistic is evaluated with reference to the distribution of the statistic generated from the model given the current parameter estimates. For a critique of posterior predictive model checks, see, however, O'Hagan and Forster (2004, chap. 8), Bayarri and Berger (1998), among others.

For the applications below, we routinely compute posterior predictive model checks based on three statistics, X_1, X_2, and X_3. Statistics X_1 and X_2 assess the fit of the data to the mean frequencies per category and the mean latencies per category, respectively, using statistics of the Pearson's chi-squared type, $\sum \frac{(O-E)^2}{E}$. X_3 is a summary measure of the fit of the joint distribution of categories and latencies based on the SSE statistic proposed by Van Zandt (2002). Let $F(c, t \mid \boldsymbol{\eta_s}) = \int_0^t f(c, x \,|\boldsymbol{\eta_s})\, dx$ be the defective cumulative distribution function of response latencies from category c given person s along with that person's person-level parameters collected in $\boldsymbol{\eta_s}$, where the joint density f is given by Equation 4. Let N be the total number of trials across all participants, and $N(\text{subtree}(c), s)$ the number of trials administered to the $s-$th participant from the subtree to which category c belongs. Furthermore, let the observed response latencies from category c be ordered from smallest to largest such that $t_{c,1} < t_{c,2} < \ldots < t_{c,N(c)}$, where $N(c)$ is the observed frequency of category c across trials and participants. Then,

$$X_3 = \sum_{c}^{C} \sum_{j=1}^{N(c)} \left(\frac{j}{N} - \sum_{s=1}^{S} \frac{N\big(\text{subtree}(c), s\big)}{N} F(c, t_j \mid \boldsymbol{\eta_s}) \right)^2 .$$

5. Application

We apply the method to the analysis of several datasets from recognition memory using the 2HT model.[8] The datasets stem from three experiments

[8]For all analyses reported below, we exclude extreme outliers in an individual's distribution of response latencies using Tukey's outlier criterion (Clark-Carter, 2004, Chap. 9);

by Arnold, Bröder, and Bayen (2015) and from two experiments by Dube et al. (2012) to which diffusion models have been fitted. These data contribute to an ongoing debate (Bröder & Schütz, 2009; Chen, Starns, & Rotello, 2015; Dube & Rotello, 2012; Dube et al., 2012; Kellen & Klauer, 2014, 2015; Kellen, Klauer, & Bröder, 2013; Province & Rouder, 2012) about whether or not memory-based judgments in recognition memory are better described in terms of discrete memory states as in the 2HT model or in terms of a continuous familiarity signal as in signal detection models (Kellen & Klauer, in press). In particular, Dube et al. (2012) argued that diffusion models are dynamic extensions of standard signal-detection models, and they were found to fit the frequencies of old/new decisions and associated response latencies reasonably well in two experiments. Here, we evaluate whether the 2HT model is also capable of providing a reasonable fit of these data with meaningful parameter values.

Arnold et al. (2015) manipulated baserate of old and new items (believed to selectively influence guessing g; Experiment 1), the emphasis on accuracy versus speed (Experiment 2), and studied-item strength by presenting strong items several times in the study list and weak items only once (believed to selectively influence detection parameters D_O). Baserate was also manipulated in five steps by Dube et al. (2012) along with studied-item strength. An overview of these datasets is provided in Table 1.

We sampled posterior parameter distributions using the algorithm described above to generate four MCMC chains in parallel. Note that the 2HT model is not identified based on only the categorical responses for the

that is, trials with latencies 3.0 times the interquartile range below the first quartile or above the third quartile in the individual's distribution of latencies were excluded prior to fitting the model datasets without baserate manipulation within participants. Convergence was monitored following the recommendations by Gelman, Carlin, Stern, and Rubin (2004, Chap. 11). Specifically, the $\hat{R}$ statistic was computed for all model parameters (all population-level parameters and all scaled person-level parameters) with the exception of the unidentified scale parameters $\boldsymbol{\xi}$. Sampling was continued until all of the $\hat{R}$ statistics were smaller than 1.05. Subsequently, every 11th sample from each MCMC chain was retained for analyses until a total N of 20,000 retained samples was reached.

5.1. Model Selection and Model Checks

Table 1 also shows the deviance information criterion (DIC) and the Bayesian p values for the summary model-check statistics X_1, X_2, and X_3 for the "detect-guess" (DG) and "default-interventionist" (DI) variants of the 2HT model shown in Figures 1 and 2. Note that, for both models, we allowed response-execution times to differ for the "old" and the "new" responses. Both models are equivalent in terms of their account of the response frequencies, but they imply different predicted response-latency distributions.

As can be seen in Table A1, the DI variant of the 2HT model is associated with the smaller DIC value in four of five cases. Moreover, its model checks are generally satisfactory, especially when considering that X_3 checks the entire joint distribution of responses and latencies. There is, however, some room for improvement for the Dube et al. (2012) data. Due to its superior performance, we will focus on the DI variant model in the discussion below.[9]

To place the success of the DI model in perspective, Figure 4 compares its fit of Dube et al.'s (2012, Experiment 1) data with the fit obtained with a

[9]We also fitted the same models with response-execution-times set equal across "old" and "new" responses. As detailed in the Appendix (Table A1), this resulted in much larger DIC values uniformly across datasets and for both the DG and the DI model.

Table 1
Reanalyzed Data Sets, DIC values, and Model Checks (Posterior p Values) for the DG and the DI Models

			DG Variant				DI Variant			
Data	S	Manipulation	ΔDIC	X_1	X_2	X_3	ΔDIC	X_1	X_2	X_3
			— Arnold et al. (2015) —							
Exp. 1	60	Baserate (b.p.)	22.12	.50	.47	.63	0.00	.50	.40	.66
Exp. 2	60	Speed-accuracy tradeoff (b.p.)	0.00	.50	.33	.49	41.19	.48	.48	.40
Exp. 3	30	Target strength (w.p.)	23.55	.48	.65	.51	0.00	.50	.66	.45
			— Dube et al. (2012) —							
Exp. 1	21	Base-rate (w.p.) × target strength (w.p.)	62.58	.0001	.01	.04	0.00	.29	.008	.11
Exp. 2	26	Base-rate (w.p.) × target strength (w.p.)	81.40	<.0001	.04	.31	0.00	.0001	.01	.44

Note. DG = "Detect-Guess"; DI = "Default-Interventionist"; b.p. = Between participants; w.p. = Within participants; S = Number of participants in the dataset. ΔDIC = DIC difference from the lowest DIC.

diffusion model under the maximum-likelihood parameter estimates reported by Dube et al. Specifically, Figure 4 shows the model fits to the response frequency data (panel “ROC”), and mean correct and false response latencies for the three kinds of items (strong studied-items, weak studied-items, and new items). As can be seen, both models capture the major trends in the data. The mean frequencies and latencies are generally well captured by the DI model's posterior predictions, whereas the diffusion model encounters difficulties in accounting for false-response latencies to new items (see points for "errors" in the lower right panel). The analogous figure for Dube et al.'s Experiment 2 shows the same patterns and is therefore omitted.

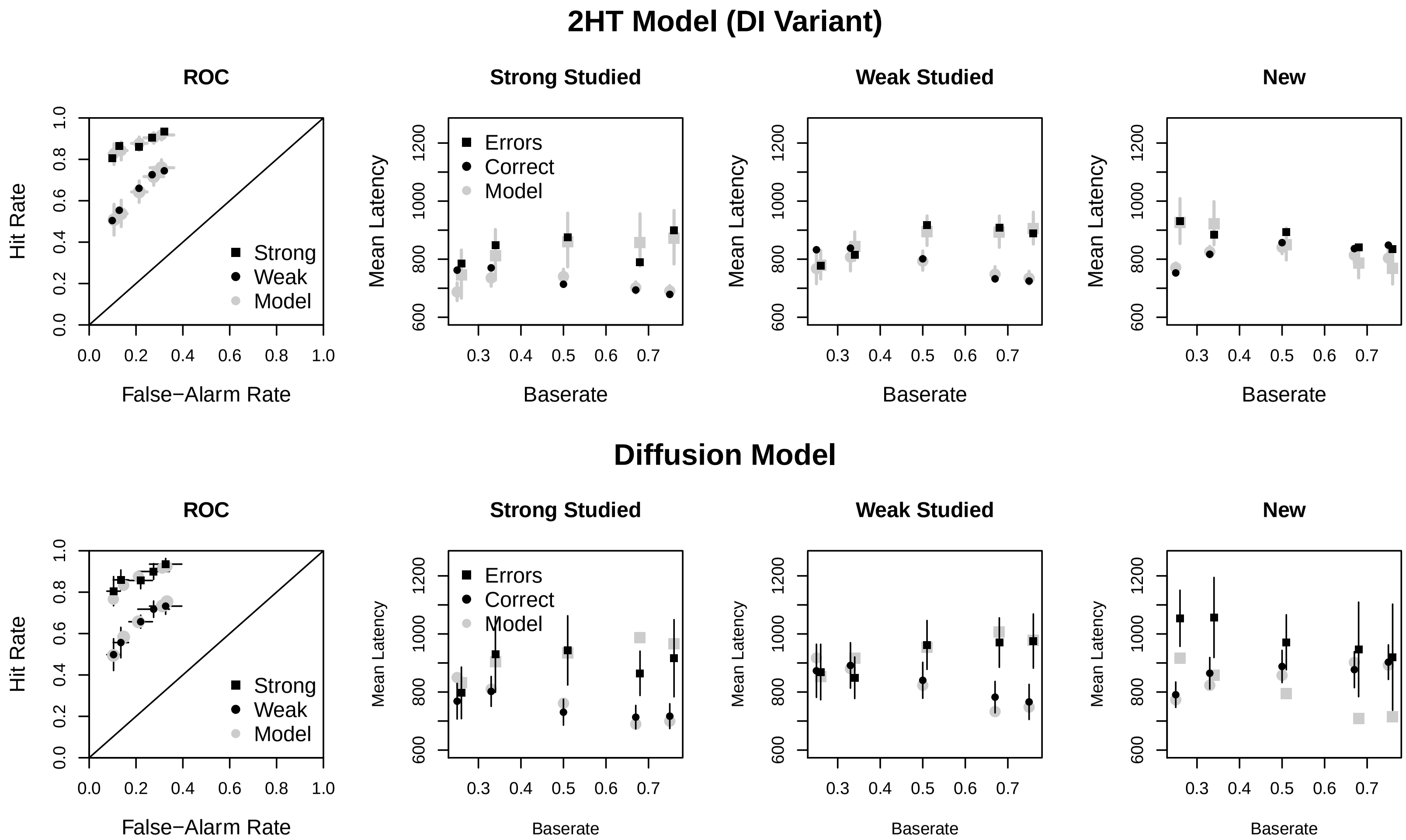


Figure 4: Fits of the 2HT model to false-alarm and hit rates (upper left panel) and to correct and false response latencies for strong studied-items, weak studied-items, and new items (upper right panels) per baserate condition. Lower panels: Fits of the diffusion model. In the upper panels, the error bars give the 95% HDIs of the model predictions, whereas in the lower panels the error bars correspond to the 95% confidence intervals associated with each observed data point. Note that the data points differ slightly between upper and lower panels due to differences in outlier treatment (see Footnote 7).

The superior fit of the DI variant of the 2HT model only serves as an example of the feasibility of the RT-MPT approach, not as a direct comparison against the diffusion model. One confounding factor is the difference in the number of parameters between the two models. The DI model requires more parameters per person than the diffusion model: the former needs to capture the outcomes and completion times of three separate two-outcomes processes (2 detection processes, one guessing process), whereas the latter model needs to characterize the outcomes and completion times of only one such process (the diffusion process). For a basic recognition-memory paradigm, the 2HT models used here require, per process, three parameters (one parameter governing accuracy, two parameters governing completion times), and three parameters to model response-execution times (two for the means of old and new responses and one for the variance), for a total of twelve parameters per person. The diffusion model uses only seven parameters in this situation (although the model could be extended to include additional components such as *drift criterion*; see Starns, Ratcliff, & White, 2012). A formal comparison of the fits of the 2HT model and the diffusion model should therefore take differences in flexibility into account via an adequate quantification of model parsimony (e.g., Vandekerckhove, Matzke, & Wagenmakers, 2015).

5.2. Parameter Estimates and Effects of Experimental Manipulations

Tables 2 and 3 report the median posterior population-level parameters and their respective 95% highest-density intervals (HDI, reported in square brackets) for the data from Arnold et al. (2015) and Dube et al. (2012), respectively. The parameters governing process-completion times were re-transformed to the original millisecond scale for ease of interpretation.

Consider the detection parameters. Detection failures (see rows/columns for D parameter under $\boldsymbol{\mu}_{-}^{(\beta)}$) generally required more time than detection successes (see $\boldsymbol{\mu}_{+}^{(\beta)}$), perhaps reflecting multiple retrieval attempts before an uncertainty state is entered. This result could be interpreted in light of the notion that successful detection emerges from a self-terminating process, whereas the failure to detect corresponds to an exhaustive process (for a general discussion, see Cox & Criss, 2017).

The effects of studied-item strength. Moreover, detection parameters for studied items (D_O, D_S, and D_W) were sensitive to studied-item strength, as captured by their respective population means $\boldsymbol{\mu}^{(\alpha)}$.[10] There was, however, little evidence for an effect of study strength on the process-completion times of the detection process, whatever its outcome (see $\boldsymbol{\mu}_{-}^{(\beta)}$ and $\boldsymbol{\mu}_{+}^{(\beta)}$). This result is in line with previous studies showing that response speed-ups observed in study-strength manipulations can be largely attributed to differences in the mixture of detection ($\mathcal{S}_1$ and $\mathcal{S}_2$) and uncertainty ($\mathcal{S}_3$) states (see Kellen et al., 2015; Province & Rouder, 2012), as reflected in the differences in $\boldsymbol{\mu}^{(\alpha)}$ parameters for the detection parameters.

The effects of baserate. The baserate manipulation had the expected effect of increasing $\boldsymbol{\mu}^{(\alpha)}$ for the guessing parameters as the proportion of old items increased. This pattern was present in in both Arnold et al.'s (2015) Experiment 1 and Dube et al.'s (2012) datasets. The 95% HDI of the contrast between low and high base rate in the former experiment contains zero, however (see the respective column Δ in Table 2). Furthermore, in Dube et al.'s (2012) datasets, the process-completion time for the guessing process increases for guessing "new" ($\boldsymbol{\mu}_{-}^{(\beta)}$) and decreases for guessing "old" ($\boldsymbol{\mu}_{+}^{(\beta)}$)

[10]Because Dube et al.'s (2012) studies always involved a within-participant manipulation of study strength, in Table 3 we denote the detection probabilities for weak and strong items by D_W and D_S, respectively. In the case of Arnold et al. (2015), only Experiment 3 included such a manipulation, but between subjects. We therefore denote all detection probabilities as D_O in Table 2.

Table 2

Parameter Estimates and 95% HDIs for the Arnold et al. (2015) Data

	Exp. 1			Exp. 2			Exp. 3
Par.	Low	High	Δ	Acc.	Speed	Δ	
				— $\boldsymbol{\mu}^{(\alpha)}$ —			
D_N	.59 [.28, .76]	.53 [.07, .67]	[−.20, .55]	.43 [.06, .57]	.45 [.07, .59]	[−.32, .18]	.79 [.73, .85]
D_O	.34 [.14, .52]	.15 [.00, .58]	[−.26, .39]	.28 [.13, .55]	.10 [.00, .40]	[−.04, .43]	.01 [.00, .09]; .54 [.43, .68][a]
g	.38 [.20, .52]	.65 [.30, .74]	[−.44, .07]	.55 [.30, .65]	.56 [.29, .64]	[−.21, .13]	.58 [.49, .66]
				— $\boldsymbol{\mu}_{-}^{(\beta)}$ (ms) —			
D_N	237 [187, 287]	302 [83, 363]	[−144, 143]	465 [362, 568]	83 [39, 109]	[285, 483]	187 [136, 224]
D_O	164 [49, 191]	250 [97, 291]	[−157, 80]	448 [377, 532]	168 [84, 304]	[285, 456]	155 [127, 183]; 138 [116, 159][a]
g	136 [49, 191]	97 [57, 245]	[−126, 118]	49 [16, 81]	45 [22, 90]	[-49, 41]	55 [28, 81]
				— $\boldsymbol{\mu}_{+}^{(\beta)}$ (ms) —			
D_N	70 [35, 118]	202 [39, 258]	[−196, 32]	295 [129, 385]	77 [49, 102]	[50, 307]	81 [61, 104]
D_O	87 [34, 160]	59 [16, 166]	[−75, 123]	168 [84, 305]	59 [20, 98]	[17, 251]	56 [35, 83]; 84 [17, 203][a]
g	171 [119, 216]	76 [37, 287]	[−113, 150]	100 [55, 197]	69 [40, 106]	[-27, 113]	88 [68, 107]
				— $\boldsymbol{\mu}^{(\gamma)}$ (ms) —			
New	623 [597, 650]	684 [652, 711]	[−97, −21]	773 [722, 819]	551 [510, 591]	[155, 282]	620 [593, 648]
Old	638 [613, 663]	593 [568, 618]	[10, 79]	716 [667, 765]	517 [475, 560]	[134, 263]	583 [560, 607]

Note. Par. = Parameter; Low= Low base-rate; High = High base-rate; Acc. = Accuracy; Δ=95% HDI of the group difference.

[a]Values for weak and strong studied items, respectively.

Table 3

Parameter Estimates and 95% HDIs for the Dube al. (2012) Data

Par.	D_N	D_W	D_S	g_1	g_2	g_3	g_4	g_5
				— Exp. 1 —				
$\boldsymbol{\mu}^{(\alpha)}$	.51 [.38, .63]	.35 [.27, .43]	.79 [.74, .85]	.19 [.14, .24]	.29 [.23, .36]	.47 [.38, .57]	.58 [.50, .65]	.65 [.58, .71]
Δ		Strong vs. Weak: [.36, .52]				Linear Trend: [.29, .79]		
$\boldsymbol{\mu}_{-}^{(\beta)}$	203 [158, 245]	189 [150, 223]	160 [110, 212]	20 [10, 34]	62 [32, 101]	114 [68, 162]	118 [66, 179]	135 [85, 185]
Δ		Strong vs. Weak: [−76, 21]				Linear Trend: [49, 132]		
$\boldsymbol{\mu}_{+}^{(\beta)}$	127 [98, 156]	107 [67, 157]	78 [51, 107]	169 [115, 226]	163 [101, 223]	79 [40, 127]	33 [15, 57]	20 [10, 35]
Δ		Strong vs. Weak: [−70, 10]				Linear Trend: [−179, −91]		
				— Exp. 2 —				
$\boldsymbol{\mu}^{(\alpha)}$	.48 [.34, .61]	.21 [.12, .30]	.72 [.63, .80]	.21 [.14, .29]	.28 [.21, .36]	.44 [.35, .52]	.57 [.48, .65]	.63 [.57, .70]
Δ		Strong vs. Weak: [.42, .62]				Linear Trend: [.32, .44]		
$\boldsymbol{\mu}_{-}^{(\beta)}$	240 [190, 294]	201 [163, 244]	232 [170, 301]	24 [11, 43]	57 [34, 82]	137 [84, 198]	155 [109, 204]	208 [155, 260]
Δ		Strong vs. Weak: [−19, 87]				Linear Trend: [105, 190]		
$\boldsymbol{\mu}_{+}^{(\beta)}$	156 [117, 202]	103 [58, 152]	100 [72, 134]	177 [126, 234]	210 [164, 261]	89 [52, 133]	51 [28, 79]	18 [8, 33]
Δ		Strong vs. Weak: [−50, 44]				Linear Trend: [−190,−110]		

Note. Par. = Parameter; Δ=95% HDI of planned contrasts, D_w =target detection for weak studied items, D_s =target detection for strong studied items. $\boldsymbol{\mu}^{(\gamma)}$ for new and old responses were, respectively, 518 ms [490, 547] and 554 ms [531, 578] in Exp. 1; 534 ms [505, 563] and 557 ms [518,594] in Exp. 2.

as the proportion of old items increases (from g_1 to g_5). This agrees well with the idea of the DI version of the 2HT model that guessing first suggests a default response and that a clear default is available to the extent to which the baserate departs from 50%. There is, however, no such trend in Experiment 1 by Arnold et al. (2015) in which a baserate manipulation was also implemented, perhaps because response-execution times for "old" and "new" responses and completion times for guessing "old" and "new" are strongly confounded in Arnold et al.'s design, making it difficult to estimate guessing completion-times with precision.

The effects of speed-accuracy instructions. We did not have clear expectations for the effects of the speed-accuracy instructions other than that we expected an emphasis on speed to speed up response execution and perhaps detection processes, as participants might refrain from continuing their retrieval efforts after a certain period. Furthermore, an emphasis on speed should not increase the probability of detection (e.g., Ludwig & Davies, 2011).

As can be seen in Table 2, an emphasis on accuracy (Arnold et al., 2015, Exp. 2) has relatively little effect on the accuracy parameters relative to an emphasis on speed. The only exception was a trend for better detection of studied items ($\boldsymbol{\mu}^{(\alpha)}$ for D_O) in Arnold et al.'s (2015) Experiment 2. This is in line with the absence of a significant effect of the manipulation on signal-detection sensitivity in traditional analyses (Arnold et al., 2015).

However, an emphasis on speed led to a speed-up of all process-completion times as well as response execution relative to an emphasis on accuracy. For example, detection processes under accuracy instructions might be based on repeated retrieval attempts, which do not add much to accuracy over and above the first attempt, whereas fewer retrieval attempts might be performed under speed instructions. Only the completion times for the guessing processes are not affected, in line with the idea that guessing always provides a fast first response proposal in the DI variant of the 2HT model.[11]

Summary. Taken together, the DI model provides a reasonable fit of the data, and its parameters react meaningfully to the different experimental manipulations. This suggests that a discrete-state model is able to provide a process account of extant response-time data in recognition memory at least to a similar extent as the diffusion model does. Limitations of the RT-MPT account are considered in the General Discussion.

5.3. Correlations

One advantage of the present hierarchical approach is the possibility to model and estimate correlations between the person-level process parameters across persons. To illustrate, we consider correlations between the person-level parameters $\beta'_{+,p,s}$, averaged across processes p, and $\beta'_{-,p,s}$, averaged across processes p based on the estimated variance-covariance matrix $\mathbf{\Sigma}$. For Arnold et al.'s (2015) Experiments 1 to 3, they amounted to (95% HDIs in brackets), in order, 0.71 $[0.38, 0.93]$, 0.82 $[0.61, 0.96]$, and 0.35 $[-0.24, 0.81]$. For Dube et al.'s (2012) Experiments 1 and 2, they amounted to 0.41 $[-0.24, 0.81]$, and 0.79 $[0.59, 0.92]$, respectively. Thus, there is some evidence for a general speed factor, such that persons completing processes with a '+' outcome fast also tend to arrive at the '−' outcome fast.

We also probed for speed-accuracy trade-off between persons by computing the correlations between the person-level parameters governing accuracy, $\alpha'_{p,s}$ corresponding to the detection parameters D_N and D_O, averaged across

[11] It is worth noting that our evidence for selective influence with Arnold et al.'s (2015) Experiment 2 data is not replicated when fitting the data with the diffusion model. Arnold et al. reported differences in the boundary-separation parameters (as expected) but also in the drift rates for new items.

the different detection processes on the one hand and the person-level parameters governing the speed of responses $\beta'_{+,p,s}$ and $\beta'_{-,p,s}$ averaged across these same detection processes on the other hand. None of these correlations was substantial, however, ranging from -0.15 to 0.28, and all 95% HDIs contained zero. Thus, there is little evidence for speed-accuracy trade-offs between persons in these datasets.

Based on the posterior distribution of the variance-covariance matrix $\mathbf{\Gamma}$ for response-execution parameters, the speed of executing "old" and "new" responses correlated positively across participants with correlations ranging from 0.35 to 0.91 across studies. None of the associated 95% HDIs contained the value zero.

5.4. Precision of Estimates

Another advantage of the present approach is that the precision of the estimation of the traditional MPT parameters governing the categorical data can be expected to increase as a side effect of including the response-time data. To see this, consider the population mean parameters $\mu_p^{(\alpha)}$ for process p. The MCMC algorithm samples from the distribution of this parameter given the categorical frequency data C and the response-time data T, that is from $P(\mu_p^{(\alpha)} \mid C, T)$, and the estimate of $\mu_p^{(\alpha)}$ is a measure of the central tendency of that distribution. The measurement precision can thus be quantified in terms of the variability of $\mu_p^{(\alpha)}$.

The traditional Bayesian approach without response-time data (e.g., Klauer, 2010) samples from $P(\mu_p^{(\alpha)} \mid C)$. It is well known that the variances of these distributions, $P(\mu_p^{(\alpha)} \mid C, T)$ and $P(\mu_p^{(\alpha)} \mid C)$, are related via $\mathrm{var}(\mu_p^{(\alpha)} \mid C) = \mathrm{E}_T[\mathrm{var}(\mu_p^{(\alpha)} \mid C, T)] + \mathrm{var}_T(\mathrm{E}[\mu_p^{(\alpha)} \mid C, T])$, where E_T and var_T refer to taking the expectation and variance with respect to T (Gelman et al., 2004, p. 24). This implies that the variance of $\mu_p^{(\alpha)}$ given C and T,

$\text{var}(\mu_p^{(\alpha)} \,|\, C, T)$, can be expected to be smaller than the variance given only C by the variance of the expected value of $\mu_p^{(\alpha)}$ given C and T, across repeated response-time measurements, T. In other words, including the response-time data can be expected to decrease the variability of the posterior distribution of $\mu_p^{(\alpha)}$.

To illustrate, we computed the lengths of the HDI's for the 16 $\mu_p^{(\alpha)}$ parameters in the two experiments by Dube et al. (2012) under the DI variant (see Tables 2 and 3). In these experiments, the 2HT model is also identified on the basis of only the categorical data, and we applied the hierarchical Bayesian approach on the basis of only the categorical data using exactly the same priors and hyperpriors as in the present approach to estimate these same parameters. HDIs based on categorical data and response times were shorter than those based on only the categorical data for 13 of the 16 parameters. The mean saving in the length of the HDIs across the 16 parameters was 39%.

6. Recovery Study

Strong theoretical results on MCMC estimation guarantee that the posterior estimates will approach the true underlying values as the sample size increases. However, such results do not excuse researchers from investigating the recoverability of parameters under any new modeling development (Heathcote, Brown, & Wagenmakers, 2015). Ideally, these recovery studies should be based on realistic datasets. We therefore conducted a recovery study based on the parameters of Dube et al.'s (2012) Experiment 1. We chose this dataset because it is the smallest one in terms of numbers of participants ($S = 21$) with a large number of process-related parameters ($P = 8$) and few trials per cell of the baserate $\times$ item-type design (between 12 and

48 trials per person). Recovery results in this framework are thereby likely to provide a lower baseline of the estimation accuracy to be expected.

Based on the population-level parameters estimated for Dube et al.'s (2012) Experiment 1, we generated 2000 artificial datasets of the same size as the original data in terms of participants and numbers of trials per participant. Each dataset was then submitted to the present algorithm.

Table 4 presents recovery results for the population-level process parameters. It can be seen that the estimates tend to track the underlying values quite closely, but there is a tendency to overestimate small process-completion times. The 50% and 95% HDIs for each parameter appear to be reasonably well calibrated as quantifying estimation uncertainty inasmuch as the underlying value tends to fall into the respective interval with approximately the nominal percentages. There are, however, again larger deviations for small process-completion times, and in general there is a tendency for the actual intervals to contain the true value somewhat less often than suggested by their nominal percentages even for large estimated process-completion times. The standard deviation of parameter values in the posterior sample provides an adequate estimate of the standard deviation of the posterior median across simulated datasets.

Table 5 presents the same information for the population-level standard deviations of the process parameters as estimated in $\mathbf{\Sigma}$. Despite the small size of the underlying datasets, the model performs quite well in estimating the underlying standard deviations in the person-level process parameters and succeeds well in quantifying estimation uncertainty in terms of the HDIs. Again, the standard deviation of parameter values in the posterior sample provides a good estimate of the standard error of the posterior median across the board.

Table 4

Parameter Recovery Study: Population-level Means of Process Parameters

Par.	True	Est.	50%[a]	95%[a]	SE^b	SD^c
			— $\boldsymbol{\mu}^{(\alpha)}$ —			
D_N	.51	.51	46.35	92.85	0.06	0.06
D_W	.35	.36	48.50	94.40	0.04	0.04
D_S	.79	.79	49.30	95.35	0.03	0.03
g_1	.21	.21	47.05	93.65	0.04	0.04
g_2	.28	.28	51.70	94.45	0.03	0.04
g_3	.44	.44	46.70	93.25	0.04	0.04
g_4	.57	.57	49.25	94.90	0.04	0.04
g_5	.63	.63	49.45	94.20	0.03	0.03
			— $\boldsymbol{\mu}_{-}^{(\beta)}$ —			
D_N	202.70	194.23	43.25	90.15	21.53	21.96
D_W	189.20	184.63	47.05	92.45	15.95	16.44
D_S	159.90	151.09	43.15	89.70	21.04	21.12
g_1	19.94	34.82	10.70	60.50	7.83	8.97
g_2	62.26	77.18	44.95	91.40	17.84	19.89
g_3	114.30	118.46	51.50	95.20	20.66	22.42
g_4	117.80	126.89	49.80	95.05	26.46	28.17
g_5	134.70	139.46	49.40	94.40	23.09	23.79
			— $\boldsymbol{\mu}_{+}^{(\beta)}$ —			
D_N	126.90	120.53	43.85	90.05	12.73	12.86
D_w	107.20	99.51	40.90	88.75	18.75	18.27
D_s	78.40	73.51	42.85	89.40	12.50	12.58
g_1	168.60	174.91	46.95	93.35	29.57	29.40
g_2	162.60	167.59	49.25	94.50	27.54	28.42
g_3	79.10	93.72	46.95	92.70	21.11	22.61
g_4	33.11	51.89	23.95	77.40	12.46	14.32
g_5	20.46	36.24	10.50	64.30	8.32	9.96

Note. Par. = Parameter; Est. = posterior median (mean across simulated datasets).

[a]Percent of simulated datasets with true value in the HDI of this percentage.

[b]Standard error of posterior medians across simulated datasets. [c]Posterior standard deviation (mean across simulated datasets).

Table 5

Parameter Recovery Study: Standard Deviations of Person-Level Process Parameters

Par.	True	Est.	50%[a]	95%[a]	SE^b	SD^c
			— $\boldsymbol{\alpha'}$ —			
D_N	0.57	0.61	52.45	96.05	0.12	0.13
D_w	0.27	0.30	48.95	92.55	0.10	0.10
D_s	0.33	0.38	47.70	92.65	0.09	0.09
g_1	0.48	0.51	50.85	94.10	0.12	0.13
g_2	0.34	0.34	46.80	91.90	0.11	0.11
g_3	0.28	0.26	51.10	91.85	0.10	0.10
g_4	0.29	0.29	48.40	92.35	0.11	0.11
g_5	0.19	0.20	45.75	95.55	0.09	0.10
			— $\boldsymbol{\beta'_-}$ —			
D_N	0.42	0.46	50.85	95.85	0.09	0.10
D_w	0.30	0.32	48.90	94.55	0.08	0.08
D_s	0.40	0.43	46.50	92.90	0.15	0.15
g_1	1.09	0.86	30.70	82.70	0.27	0.26
g_2	1.08	1.03	49.00	94.95	0.21	0.23
g_3	0.70	0.72	52.10	96.70	0.16	0.17
g_4	0.83	0.85	52.70	96.65	0.18	0.20
g_5	0.56	0.58	52.20	95.25	0.16	0.17
			— $\boldsymbol{\beta'_+}$ —			
D_N	0.24	0.26	46.70	94.80	0.11	0.12
D_w	0.52	0.52	47.70	92.50	0.19	0.19
D_s	0.60	0.65	50.75	95.00	0.15	0.16
g_1	0.46	0.46	49.80	93.65	0.17	0.18
g_2	0.53	0.54	51.55	92.40	0.17	0.18
g_3	0.96	0.90	48.80	93.65	0.20	0.22
g_4	1.21	1.01	33.75	85.75	0.27	0.27
g_5	1.19	0.94	32.10	82.70	0.29	0.28

Note. Par. = Parameter; Est. = posterior median (mean across simulated datasets).

[a]Percent of simulated datasets with true value in the HDI of this percentage.

[b]Standard error of posterior medians across simulated datasets. [c]Posterior standard deviation (mean across simulated datasets).

Table 6 presents the same information for selected correlations, specifically for the eight lowest, eight median, and eight highest correlations. Considering the small number of persons across which these correlations are estimated and the small numbers of trials per cell of the design, the correlations are estimated reasonably well. The estimates faithfully track the sign of the underlying correlations. Perhaps not surprisingly, given the small sizes of the datasets, they underestimate the absolute sizes of the true correlations, and standard deviations of posterior medians across simulated datasets are systematically smaller than the posterior standard deviation of parameter values. Nevertheless, the HDIs quantify the estimation uncertainty reasonably well.

A reviewer suggested to provide information on model discrimination via simulation. Specifically, the concern was that the DI model might outperform the DG model due to greater flexibility. To assess this possibility, we generated 100 more datasets from the DI model as well as 100 datasets from the DG model using the same parameters and procedures as just described, and fitted these datasets with the DG and the DI model, computing DIC for both models. When generating from the DI model, DIC was higher for the DG model than for the DI model for all 100 artificial datasets; when generating from the DG model, DIC was higher for the DI model than the DG model in 98 cases. The mean differences in DIC values in favor of the generating model were 379.02 and 197.79, respectively. Not surprisingly, the differences were significant in a t test across the 100 datasets in both cases: $t = 40.34$, $df = 99$, $p < .001$ and $t = 20.11$, $df = 99$, $p < .001$, respectively. These DIC differences indicate that none of the models is particularly apt in mimicking the other, providing additional support for the present results favoring the DI variant of the 2HT model.

Table 6
Parameter Recovery Study: Selected Correlations of Person-Level Process Parameters

Pars.	True	Est.	50%[a]	95%[a]	SE^b	SD^c
	— Eight Lowest Correlations —					
$\alpha'_{D_s}, \beta'_{g_5,-}$	−.53	−.23	37.95	95.45	0.17	0.28
$\alpha'_{D_s}, \beta'_{g_5,+}$	−.51	−.21	38.25	96.15	0.17	0.29
$\alpha'_{g_1}, \beta'_{D_N,-}$	−.50	−.33	50.05	94.95	0.17	0.22
$\alpha'_{g_1}, \beta'_{g_5,-}$	−.50	−.30	50.25	96.60	0.17	0.24
$\beta'_{g_1,-}, \beta'_{g_5,+}$	−.48	−.20	41.90	94.90	0.18	0.28
$\alpha'_{g_1}, \beta'_{D_s,-}$	−.47	−.24	46.95	96.90	0.17	0.26
$\alpha'_{g_3}, \beta'_{D_N,-}$	−.47	−.23	48.10	95.75	0.18	0.27
$\beta'_{g_1,-}, \beta'_{g_5,-}$	−.44	−.21	49.85	95.10	0.19	0.27
	— Eight Median Correlations —					
$\beta'_{g_5,-}, \beta'_{D_w,+}$	.03	.04	59.20	99.30	0.20	0.29
$\alpha'_{D_s}, \beta'_{g_5,+}$	.03	.00	58.30	98.85	0.19	0.27
$\alpha'_{g_5}, \beta'_{D_N,-}$	.03	.09	60.25	99.35	0.18	0.29
$\beta'_{g_2,-}, \beta'_{g_1,+}$	.03	.02	65.70	99.50	0.18	0.28
$\alpha'_{g_4}, \beta'_{g_1,+}$	.04	.02	74.85	99.95	0.15	0.30
$\beta'_{g_1,-}, \beta'_{D_s,+}$	.04	.01	55.40	97.75	0.21	0.27
$\alpha'_{g_3}, \beta'_{D_w,+}$	.04	.00	68.90	99.65	0.17	0.30
$\beta'_{g_2,-}, \beta'_{D_s,+}$	.04	.03	55.20	96.85	0.20	0.24
	— Eight Largest Correlations —					
$\beta'_{D_N,-}, \beta'_{g_3,-}$	.48	.35	55.95	97.40	0.17	0.22
$\beta'_{D_w,-}, \beta'_{g_3,-}$	.48	.34	58.50	98.15	0.16	0.23
$\alpha'_{D_w}, \beta'_{D_w,-}$	.48	.21	40.30	94.20	0.17	0.27
$\beta'_{D_w,+}, \beta'_{D_s,+}$	.51	.27	47.70	96.45	0.18	0.27
$\beta'_{D_N,-}, \beta'_{g_5,-}$	.53	.37	57.10	97.05	0.16	0.23
$\beta'_{g_3,-}, \beta'_{g_5,-}$	.54	.35	51.65	95.70	0.17	0.23
$\alpha'_{g_3}, \beta'_{g_1,-}$	.56	.21	28.60	92.90	0.17	0.29
$\beta'_{D_N,-}, \beta'_{D_w,-}$	.60	.41	50.80	95.00	0.15	0.22

Note. Pars. = Parameters; Est. = posterior median (mean across simulated datasets).
[a]Percent of simulated datasets with true value in the HDI of this percentage.
[b]Standard error of posterior medians across simulated datasets. [c]Posterior standard deviation (mean across simulated datasets).

7. General Discussion

The modeling tools currently available in researchers' toolboxes constitute a major source of constraint for the type of experimental paradigms that are ultimately adopted. In the case of response-time modeling, the need for a large number of observations has been relaxed due to recent advances in hierarchical Bayesian methods, making them available to a large number of applications (e.g., Rouder, Province, Morey, Gomez, & Heathcote, 2015; Vandekerckhove, Tuerlinckx, & Lee, 2011). However, most response-time models like the prominent diffusion model (Ratcliff & Rouder, 1998) are restricted to two-choice paradigms (one exception being the family of linear-ballistic accumulator models; Brown & Heathcote, 2008), which limits their overall usefulness. No such limitations exist in the case of MPT models.

The present work aims to enrich the current toolbox by proposing a method for combining two long-standing modeling traditions that are typically seen as somewhat disjoint — process-oriented response-time modeling and multinomial-processing-tree modeling. As traditionally assumed in MPT models, the probability of observing a certain response corresponds to a mixture of different processing paths. We propose that the latencies associated with a given response can be captured by ascribing a completion-time distribution to each process outcome included in the different paths that lead to that response, in addition to encoding and response-execution times. Because our proposed method can be applied to any existing member of the MPT model class, it imposes no a priori constraints on the type of MPT paradigm that one can consider, as long as response times can be reliably recorded. In fact, the inclusion of response time might even lead to less constrained model accounts, as parameters that were not identified can become so.

As an application example, we extended two variants of the well-known 2HT model and tested their ability to capture recognition-memory data across a range of experimental manipulations. Interestingly, we found that a "default-interventionist" variant of the 2HT, in which the guessing process precedes the attempts to retrieve the item from memory, provided the best account of the datasets considered here. This tree structure deviates from the way the 2HT is typically conceptualized, namely in terms of the "detect-guess" variant. However, it should be emphasized that the two variants are indistinguishable on the basis of the typical categorical data collected and analyzed in recognition-memory experiments. The conceptualization of the model in terms of the "detect-guess" variant has therefore been a convention based on tacit and previously untestable assumptions about processing order. Future work is required to determine whether the processing-tree structure found to be most adequate here depends on the characteristics of the experimental design, a likely possibility in any complex faculty such as human memory (e.g., Humphreys, Bain, & Pike, 1989; Meyer & Kieras, 1997). One indication in this direction may be the finding that the traditional "detect-guess" variant outperformed the "default-interventionist" variant in one of the analyzed datasets in which instructions emphasized either speed or accuracy. In any case, the present results demonstrate that by incorporating response times, one can overcome a long-standing inability to distinguish between different processing tree structures (e.g., Kellen & Singmann, in press; Kellen, Singmann, & Klauer, 2014).

The applications also illustrate two additional advantages of incorporating response times in the present framework: First, models that are not identifiable on the basis of only categorical responses will typically be identified when response times are included via assumptions about process-completion distributions and distributions of encoding and response-execution components. For example, in the above applications, models with different detection parameters D_N and D_O for detecting new and old items, respectively, could be fit even where this would not have been possible in the traditional MPT analysis. Second, the present method provides a principled alternative to the currently dominant diffusion-model analyses (e.g., Matzke & Wagenmakers, 2009; Voss, Voss, & Lerche, 2015), at least in the many cases in which theory-driven and validated MPT models have been formulated for the accuracy data in the experimental paradigm under scrutiny. In such cases, RT-MPT models establish a principled competitor for process-oriented accounting of the data. RT-MPT models build on the MPT models that were successful in describing the accuracy data, now addressing the same range of data at the same level of detail as the diffusion-model analyses of them as illustrated here for recognition-memory paradigms.

Note, however, that RT-MPT models and diffusion models differ in the breadth and depth of the process accounts they provide. RT-MPT models postulate that several processes are at work and aim at quantifying the different processes' relative contributions to observed responses and response times, conditional on the constellation of the processes' interactions as coded in the tree structure. This is especially helpful in modeling paradigms in which multiple processes are likely involved. These processes are, however, not modeled in any more detail beyond what they contribute to the frequencies of responses and the observed response times. In contrast, diffusion models postulate one (diffusion) process, which they model in more depth. For example, the threshold parameter in diffusion models allows one to model speed-accuracy trade-offs in a parsimonious and compact way, and the starting-point parameter captures negative correlations between response frequency and response time in a natural fashion. Options 2 and 3 discussed in the introduction for alternative distributional assumptions may enable one to graft some of this elegance onto response-time extensions of MPT models.

The proposed approach is also likely to be valuable in areas in which the incorporation of response times into existing MPT models represents a major development. One such area of research is *implicit social cognition* — which includes prominent paradigms such as the Implicit Association Test or the Weapon Identification Task — where several MPT models capturing a diverse set of automatic and controlled processes have been proposed (e.g., Bishara & Payne, 2009; Meissner & Rothermund, 2013; for reviews, see Hütter, & Klauer, 2016; Sherman, Klauer, & Allen, 2010). Despite their merits, these models currently ignore response times, a key aspect of participants' judgments (e.g., decide as quickly as possible whether the object being held by a Black or White person is a gun) and a key aspect in traditional analyses of such data. Incorporating response times into the model analyses would allow the MPT literature on social-cognition paradigms to speak more directly to the vast social-cognition literature relying on response times as the major dependent variable. One important question that response times can help answering concerns the relative temporal sequencing of controlled detection and inhibition processes and their dominance over automatic processes. Another contribution is that the present method, by disentangling process-completion times, allows one to characterize the speed of the different modeled processes and to test theoretical predictions about their different speeds. For example, one defining characteristic of automatic processes such as the activation of response proposals based on stereotypic associations in the Implicit Association Task or the Weapon Identification Task is that they are believed to complete faster than the controlled detection and inhibition processes that are also assumed to operate in these tasks. Efforts to incorporate response times into some of these models from social cognition are ongoing.

Finally, since its inception, the MPT model class has been framed as a way to obtain theoretically motivated and empirically validated decompositions of the major processes at work, which can be useful when characterizing individuals from different populations (Riefer & Batchelder, 1988). When used as a *measurement tool*, a MPT model can provide important insights, such as the attribution of a given cognitive deficit to differences in a specific parameter (e.g., Riefer, Knapp, Batchelder, Bamber, & Manifold, 2002). By providing simultaneous descriptions of accuracy and response-time data, the extended RT-MPT models obviously make use of more information than traditional MPT models thereby likely enhancing the usefulness of such models as measurement tools. For example, the precision of the estimation of the traditional MPT parameters governing the categorical data can be expected to increase as a side effect of including the response-time data. As another simple example, the new RT-MPT method can accommodate differences between persons in the relative emphasis an individual puts on accuracy at the expense of speed and vice versa — individual differences to which traditional MPT models, relying on only the accuracy data, are vulnerable. For instance, it would now be possible to diagnose cognitive deficits that are revealed primarily in processing speed rather than processing accuracy and to pinpoint the processes most affected, addressing important issues in, for example, the study of cognitive aging (e.g., Kliegl, Mayr, & Krampe, 1994).

To conclude, consider just one more area where measurement modeling using the RT-MPT method seems especially promising — the characterization of *workload capacity* conceptualized in terms of processing speed as a function of the number of concurrent stimuli (Miller, 1982). The assessment of individual workload capacity can often be challenging due to the need for a considerable number of trials per person, as well as the possibility of contamination by guessing-based responses (for an overview, see Gondan & Minakata, 2016). Alternatively, one could follow up on Ollman's (1966) earlier work and build an RT-MPT that estimates the latencies of the stimulus-dependent and stimulus-independent (i.e., guessing) processes. Such an approach has the advantage of capturing information in the response-time distributions in a small number of parameters, with workload capacity being assessed by comparing process-latency parameters across conditions (see Eidels, Donkin, Brown, & Heathcote, 2010). Moreover, the RT-MPT method is hierarchical, which improves parameter estimation by allowing individual estimates to inform each other, an advantage of critical importance when data are sparse (Katahira, 2016; Klauer, 2010).

Reaping these benefits rests on these models fitting the to-be-analyzed data well and on successfully passing a validation program based on selective-influence studies (Heathcote et al., 2015; Klauer et al., 2012) as tentatively illustrated here for the response-time extension of the 2HT model. Selective-influence studies implement experimental manipulations believed to affect only one process in the model. The question then is whether each such manipulation will be reflected primarily in the parameters for the targeted process while leaving parameters pertaining to non-manipulated processes unaffected. As already mentioned, one limitation of the present development in this context is that like diffusion models, RT-MPTs rely on a set of specific auxiliary assumptions about the distributions of process-completion times and encoding and response-execution times. It is likely that cases exist in which these specific auxiliary assumptions do not even approximately describe the data-generating process, leading to misfit and failure of demonstrating selective influence, even though the core structural and psychological assumptions of the underlying MPT model as such may still be viable. We hope to be able to relax this limitation to some extent through future work in which we aim to develop the model for the alternative sets of distributional assumptions outlined above (see Section 2.1.).

Acknowledgements

Christoph Klauer received support from grant DFG Kl 614/39-1 from the Deutsche Forschungsgemeineschaft (DFG). David Kellen received support from the Swiss National Science Foundation (SNF) Grant 100014_165591.

References

Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, *88*, 669–679. doi:10.1080/01621459.1993.10476321

Annis, J., Miller, B. J., & Palmeri, T. J. (2017). Bayesian inference with Stan: A tutorial on adding custom distributions. *Behavior Research Methods*, *49*, 863–886. doi:10.3758/s13428-016-0746-9

Arnold, N., Bröder, A., & Bayen, U. (2015). Empirical validation of the diffusion model for recognition memory and a comparison of parameter-estimation methods. *Psychological Research*, *79*, 882–898. doi:10.1007/s00426-014-0608-y

Balota, D. A., & Yap, M. J. (2011). Moving beyond the mean in studies of mental chronometry: The power of response time distributional analyses. *Current Directions in Psychological Science*, *20*, 160–166. doi:10.1177/0963721411408885

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review*, *97*, 548–564. doi:10.1037/0033-295X.97.4.548

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial processing tree modeling. *Psychonomic Bulletin & Review*, *6*, 57–86. doi:10.3758/BF03210812

Bayarri, M. J., & Berger, J. O. (1998). Quantifying surprise in the data and model verification. In J. M. Bernardo, J. O. Bernardo, A. P. Dawid, & A. F. M.S Smith (Eds.), *Bayesian Statistics* (Vol. 6, pp. 53–82). Oxford, England: Oxford University Press.

Bishara, A. J., & Payne, B. K. (2009). Multinomial process tree models of control and automaticity in weapon misidentification. *Journal of Experimental Social Psychology*, *45*, 524–534. doi:10.1016/j.jesp.2008.11.002

Bröder, A., & Schütz, J. (2009). Recognition ROCs are curvilinear - or are they? On premature arguments against the two-high-threshold model of recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 587–606. doi:10.1037/a0015279

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

Brown, S. D., Marley, A. A. J., & Heathcote, A. (2012). *Decision times for preference models.* Unpublished manuscript, University of Newcastle, Callaghan, Australia.

Burbeck, S. L., & Luce, R. D. (1982). Evidence from auditory simple reaction times for both change and level detectors. *Perception & Psychophysics*, *32*, 117–133. doi:10.3758/BF03204271

Chen, T., Starns, J. J., & Rotello, C. M. (2015). A violation of the conditional independence assumption in the two-high-threshold model of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 1215–1222. doi:10.1037/xlm0000077

Clark-Carter, D. (2004). *Quantitative psychological research: A student's handbook*. New York: Psychology Press.

Cox, Gregory E., & Criss, Amy H. (2017). Parallel interactive retrieval of item and associative information from event memory. *Cognitive Psychology*, *97*, 31–61. doi:10.1016/j.cogpsych.2017.05.004

Dube, C., & Rotello, C. M. (2012). Binary ROCs in perception and recognition memory are curved. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *38*, 130–151. doi:10.1037/a0024957

Dube, C., Starns, J. J., Rotello, C. M., & Ratcliff, R. (2012). Beyond ROC curvature: Strength effects and response time data support continuous-evidence models of recognition memory. *Journal of Memory and Language*, *67*, 389–406. doi:10.1016/j.jml.2012.06.002

Dzhafarov, E. N., & Schweickert, R. (1995). Decompositions of Response Times: an Almost General Theory. *Journal of Mathematical Psychology*, *39*, 285–314. doi:10.1006/jmps.1995.1029

Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans.* Philadelphia, PA: Society for Industrial and Applied Mathematics.

Eidels, A., Donkin, C., Brown, S. D., & Heathcote, A. (2010). Converging measures of workload capacity. *Psychonomic Bulletin & Review*, *17*, 763–771. doi:10.3758/PBR.17.6.763

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Journal of Psychology*, *217*, 108–124. doi:10.1027/0044-3409.217.3.108

Falmagne, J. C. (1965). Stochastic models for choice reaction time with applications to experimental results. *Journal of Mathematical Psychology*, *2*, 77–124. doi:10.1016/0022-2496(65)90018-0

Gawronski, B., Conway, P, Armstrong, J., Friesdorf, R., & Hütter, M. (2016). Moral dilemma judgments: Disentangling deontological inclinations, utilitarian inclinations, and general action tendencies. In J. P.

Forgas, P. A. M. Van Lange, & L. Jussim (Eds.), *Social psychology of morality* (pp. 91–110). New York, NY: Psychology Press.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, B. D. (2004). *Bayesian data analyses* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York, USA: Cambridge University Press.

Gelman, A., & Shalizi, C. R. (2013). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical & Statistical Psychology*, *66*, 8–38. doi:10.1111/j.2044-8317.2011.02037.x

Gilks, W. R.., & Wild, P. (1992). Adaptive Rejection Sampling for Gibbs Sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *41*, 337.

Gill, J. (2008). *Bayesian methods: A social and behavioral sciences approach* (2nd ed.). Boca Raton, Florida: Chapman & Hall/CRC.

Goldstein, W. M., & Fisher, D. L. (1991). Stochastic networks as models of cognition: Derivation of response time distributions using the Order-of-Processing method. *Journal of Mathematical Psychology*, *35*, 214–241. doi:10.1016/0022-2496(91)90026-P

Gondan, M., & Minakata, K. (2016). A tutorial on testing the race model inequality. *Attention, Perception, & Psychophysics*, *78*, 723–735. doi:10.3758/s13414-015-1018-y

Heathcote, A., Brown, S. D., & Wagnemakers, E.-J. (2015). An introduction to good practices in cognitive modeling. In B. U. Forstmann

& E.-J. Wagenmakers (Eds.), *An introduction to model-based cognitive neuroscience* (pp. 25–48). New York, NY: Springer.

Heathcote, A., Wagenmakers, E.-J., & Brown, S. D. (2014). The falsifiability of actual decision-making models. *Psychological Review*, *121*, 676–678. doi:10.1037/a0037771

Heck, D. W., & Erdfelder, E. (2016). Extending multinomial processing tree models to measure the relative speed of cognitive processes. *Psychonomic Bulletin & Review*, *23*, 1440–1465. doi:10.3758/s13423-016-1025-6

Heck, D. W., & Erdfelder, E. (2017). Linking process and measurement models of recognition-based decisions. *Psychological Review*, *124*, 442–471. doi:10.1037/rev0000063

Heck, D. W., & Wagenmakers, E.-J. (2016). Adjusted priors for Bayes factors involving reparameterized order constraints. *Journal of Mathematical Psychology*, *73*, 110–116. doi:10.1016/j.jmp.2016.05.004

Hütter, M., & Klauer, K. C. (2016). Applying processing trees in social psychology. *European Review of Social Psychology*, *27*, 116–159. doi:10.1080/10463283.2016.1212966

Hu, X. (2001). Extending general processing tree models to handle/analyze reaction time experiments. *Journal of Mathematical Psychology*, *45*, 603–634. doi:10.1006/jmps.2000.1340

Hu, X., & Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika*, *59*, 21–47. doi:10.1007/BF02294263

Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review*, *96*, 208–233. doi:10.1037/0033-295X.96.2.208

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513–541. doi:10.1016/0749-596X(91)90025-F

Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences.* Amsterdam: Elsevier.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). *Continuous univariate probability distributions (Vol. 1).* New York: Wiley.

Jones, M., & Dzhafarov, E. N. (2014). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. *Psychological Review*, *121*, 1–32. doi:10.1037/a0034190

Katahira, K. (2016). How hierarchical models improve point estimates of model parameters at the individual level. *Journal of Mathematical Psychology*, *73*, 37–58. doi:10.1016/j.jmp.2016.03.007

Kellen, D., & Klauer, K. C. (2014). Discrete-state and continuous models of recognition memory: Testing core properties under minimal assumptions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 1795–1804. doi:10.1037/xlm0000016

Kellen, D., & Klauer, K. C. (2015). Signal detection and threshold modeling of confidence-rating ROCs: A critical test with minimal assumptions. *Psychological Review*, *122*, 542–557. doi:10.1037/a0039251

Kellen, D., & Klauer, K. C. (in press). Elementary signal detection and threshold theory. In J. Wixted & E.-J. Wagenmakers (Eds.), *Stevens' handbook of experimental psychology and cognitive neuroscience*. New York: Wiley.

Kellen, D., Klauer, K. C., & Bröder, A. (2013). Recognition memory models and binary-response ROCs: A comparison by minimum description length. *Psychonomic Bulletin & Review*, *20*, 693–719. doi:10.3758/s13423-013-0407-2

Kellen, D., & Singmann, H. (in press). Memory representations, tree structures, and parameter polysemy: Comment on Cooper, Greve, and Henson (2017). *Cortex*.

Kellen, D., Singmann, H., & Klauer, K. C. (2014). Modeling source-memory overdistribution. *Journal of Memory and Language*, *76*, 216–236. doi:10.1016/j.jml.2014.07.001

Kellen, D., Singmann, H., Vogt, J., & Klauer, K. C. (2015). Further Evidence for Discrete-State Mediation in Recognition Memory. *Experimental Psychology*, *62*, 40–53. doi:10.1027/1618-3169/a000272

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika*, *75*, 70–98. doi:10.1007/s11336-009-9141-0

Klauer, K. C., Dittrich, K., Scholtes, C., & Voss, A. (2015). The invariance assumption in process-dissociation models: An evaluation across three domains. *Journal of Experimental Psychology: General*, *144*, 198–221. doi:10.1037/xge0000044

Klauer, K. C., & Kellen, D. (2015). The flexibility of models of recognition memory: The case of confidence ratings. *Journal of Mathematical Psychology*, *67*, 8–25. doi:10.1016/j.jmp.2015.05.002

Klauer, K. C., Singmann, H., & Kellen, D. (2015). Parametric order constraints in multinomial processing tree models: An extension of Knapp and Batchelder (2004). *Journal of Mathematical Psychology*, *64-65*, 1–7. doi:10.1016/j.jmp.2014.11.001

Klauer, K. C., Stahl, C., & Voss, A. (2012). Multinomial models and diffusion models. In K. C. Klauer, A. Voss, & C. Stahl (Eds.), *Cognitive methods in social psychology. Abridged edition* (pp. 331–354). New York: Guilford.

Kliegl, R., Mayr, U., & Krampe, R. T. (1994). Time-accuracy functions for determining process and person differences: An application to cognitive aging. *Cognitive Psychology*, *26*, 134–164. doi:10.1006/cogp.1994.1005

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* New York: Oxford University Press.

Ludwig, C. J., & Davies, J. R. (2011). Estimating the growth of internal evidence guiding perceptual decisions. *Cognitive Psychology*, *63*, 61–92. doi:10.1016/j.cogpsych.2011.05.002

Marley, A. A. J., & Colonius, H. (1992). The "horse race" random utility model for choice probabilities and reaction times, and its competing risks interpretation. *Journal of Mathematical Psychology*, *36*, 1–20. doi:10.1016/0022-2496(92)90050-H

Matzke, D., Dolan, C. V., Batchelder, W. H., & Wagenmakers, E.-J. (2015). Bayesian Estimation of Multinomial Processing Tree Models with Heterogeneity in Participants and Items. *Psychometrika*, *80*, 1–31. doi:10.1007/s11336-013-9374-9

Matzke, D., & Wagenmakers, E.-J. (2009). Psychological interpretation of the ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review*, *16*, 798–817. doi:10.3758/PBR.16.5.798

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, *86*, 287–330. doi:10.1037/0033-295X.86.4.287

Meissner, F., & Rothermund, K. (2013). Estimating the Contributions of Associations and Recoding in the Implicit Association Test: The ReAL Model for the IAT. *Journal of Personality and Social Psychology*, *104*, 45–69. doi:10.1037/a0030734

Mewhort, D. J. K., & Johns, E. E. (2000). The extralist-feature effect: Evidence against item matching in short-term recognition memory. *Journal of Experimental Psychology: General*, *129*, 262–284. doi:10.1037/0096-3445.129.2.262

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanism. *Psychological Review*, *104*, 3–65. doi:10.1037/0033-295X.104.1.3

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*, 247–279. doi:10.1016/0010-0285(82)90010-X

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

Ollman, R. (1966). Fast guesses in choice reaction time. *Psychonomic Science*, *6*, 155–156. doi:10.3758/BF03328004

Plummer, M. et al. (2003). Jags: A program for analysis of Bayesian graphical models using Gibbs sampling. *In Proceedings of the 3rd international workshop on distributed statistical computing* (Vol. 124, p. 125). Vienna, Austria.

Province, J. M., & Rouder, J. N. (2012). Evidence for discrete-state processing in recognition memory. *Proceedings of the National Academy of Sciences*, *109*, 14357–14362. doi:10.1073/pnas.1103880109

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*, 347–356. doi:10.1111/1467-9280.00067

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, *95*, 318–339. doi:10.1037/0033-295X.95.3.318

Riefer, D. M., Knapp, B. R., Batchelder, W. H., Bamber, D., & Manifold, V. (2002). Cognitive psychometrics: Assessing storage and retrieval deficits in special populations with multinomial processing tree models. *Psychological Assessment*, *14*, 184–201. doi:10.1037/1040-3590.14.2.184

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. doi:10.1007/s11336-013-9396-3

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

Schwarz, W. (2001). The ex-Wald distribution as a descriptive model of response times. *Behavior Research Methods, Instruments, & Computers*, *33*, 457–469. doi:10.3758/BF03195403

Schweickert, R. (1978). A critical path generalization of the additive factor method: Analysis of a Stroop task. *Journal of Mathematical Psychology*, *18*, 105–139. doi:10.1016/0022-2496(78)90059-7

Schweickert, R., Fisher, D. R., & Sung, K. (2012). *Discovering cognitive architecture by electively influencing mental processes*. Singapore: World Scientific.

Sherman, J. W., Klauer, K. C., & Allen, T. (2010). Mathematical modeling of implicit social cognition: The machine in the ghost. In Gawronski B. & Payne B. K. (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 156–175). New York: New York: Guilford Press.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*, 34–50. doi:10.1037/0096-3445.117.1.34

Starns, J. J., Ratcliff, R., & White, C. N. (2012). Diffusion model drift rates can be influenced by decision processes: An analysis of the strength-based mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1137–1151. doi:10.1037/a0028151

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, *30*, 276–315. doi:10.1016/0001-6918(69)90055-9

Sternberg, S., & Backus, B. T. (2015). Sequential processes and the shapes of reaction time distributions. *Psychological Review*, *122*, 830–837. doi:10.1037/a0039658

Strack, F., & Bless, H. (1994). Memory for nonoccurences: Metacognition and presuppositional strategies. *Journal of Memory and Language*, *33*, 203–217. doi:10.1006/jmla.1994.1010

Titterington, D. M., Smith, A. F. M., & Makov, U. E. (1985). *Statistical analysis of finite mixture distributions*. New York: Wiley.

Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes*. New York: Cambridge University Press.

Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology*, *39*, 321–359. doi:10.1006/jmps.1995.1033

Van Zandt, T. (2002). Analysis of Response Time Distributions. In J. T. Wixted & H. Pashler (Eds.), *Stevens' Handbook of Experimental Psychology* (Vol. 4, 3rd ed., pp. 461–516). New York: John Wiley & Sons, Inc.

Vandekerckhove, J., Matzke, D., & Wagenmakers, E.-J. (2015). Model comparison and the principle of parsimony. In J. R. Busemeyer, Z.

Townsend, J. Wang, & A. Eidels (Eds.), *Oxford Handbook of Computational and Mathematical Psychology* (pp. 300–319). Oxford, UK: Oxford University Press.

Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2011). Hierarchical diffusion models for two-choice response times. *Psychological Methods*, *16*, 44–62. doi:10.1037/a0021765

Voss, A., Voss, J., & Klauer, K. C. (2010). Separating response-execution bias from decision bias: Arguments for an additional parameter in Ratcliff's diffusion model. *British Journal of Mathematical and Statistical Psychology*, *63*, 539–555. doi:10.1348/000711009X477581

Voss, A., Voss, J., & Lerche, V. (2015). Assessing cognitive processes with diffusion model analyses: A tutorial based on fast-dm-30. *Frontiers in Psychology*, *6*, 336. doi:10.3389/fpsyg.2015.00336

Wagenmakers, E.-J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review*, *114*, 830–841. doi:10.1037/0033-295X.114.3.830

Wu, H., Myung, J. I., & Batchelder, W. H. (2010). Minimum description length model selection of multinomial processing tree models. *Psychonomic Bulletin & Review*, *17*, 275–286. doi:10.3758/PBR.17.3.275

Yantis, S., Meyer, D. E., & Smith, J. E. K. (1991). Analyses of multinomial mixture distributions: New tests for stochastic models of cognition and action. *Psychological Bulletin*, *110*, 350–374. doi:10.1037/0033-2909.110.2.350

Appendix

A.1. Non-Binary MPT Models

Models containing nodes with more than two children can be transformed into binary MPT models. For that purpose, each node with more than two children is replaced by a sequence of linked nodes each of which has only two children. Hu and Batchelder (1994) show how to parameterize the links to achieve equivalence between the original non-binary MPT model and the resulting binary MPT model. To maintain equivalence of the response-time predictions, the technical links connecting the series of binary nodes that replace a non-binary node should not be assigned a completion-time component so that having such links in a processing path does not add to the response times. This guarantees equivalence of the person-level models. In hierarchical models, priors and hyperpriors also need to be adjusted to guarantee equivalence for the entire hierarchical model following parameter transformations (Gelman et al., 2004, Chap. 2; see also Heck & Wagenmakers, 2016).

A.2. The Likelihood of the Joint Distribution of Parameters and Data

On each trial x, administered to subject $s = s(x)$, a category $c = c(x)$ with response $r = r(c(x))$ and latency $t = t(x)$ is observed from one of the subtrees of the models, denoted subtree(x). Subtree(x) is represented as the set of edges (n, o) in the subtree; the set of nodes n in the subtree will be referred to as nodes(subtree(x)).

The data observed for each trial x are augmented as described above by the path B_x along which the category $c(x)$ was reached, by $z-$variates $\boldsymbol{z}_x = (z_{n,x})_{n \in \text{nodes}(\text{subtree}(x))}$ for each node of the relevant subtree, and process-completion times $\boldsymbol{\tau}_x = (\tau^o_{n,x})_{(n,o) \in \text{subtree}(x)}$ for each edge of that subtree along with the residual encoding and motor-execution component δ_x.

Let 1_C be an indicator function that takes on the value one if the condition C is satisfied and the value zero otherwise. Let $\boldsymbol{\theta}$ be a vector that stacks all of the model parameters. For N trials x, the probability function for the joint distribution of parameters and observed and augmented data is given by:

$$\begin{aligned}
p(\;& (B_x, \boldsymbol{z}_x, \boldsymbol{\tau}_x, \delta_x)_{x=1,\ldots,N}, (c_x, t_x)_{x=1,\ldots,N}, \boldsymbol{\theta}) \\
\propto \; & \prod_{x=1}^{N} \Bigg[\left(\prod_{n \in \text{nodes(subtree}(x))} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(z_{n,x} - \alpha_{p(n),s(x)}\right)^2} \right) \\
& \left(\prod_{(n,o) \in B_x : o=+} 1_{\{z_{n,x} \geq 0\}} \right) \left(\prod_{(n,o) \in B_x : o=-} 1_{\{z_{n,x} < 0\}} \right) 1_{\{B_x \text{ ends in } c_x\}} \\
& \left(\prod_{(n,o) \in \text{subtree(x)}} \lambda^o_{p(n),s(x)} e^{-\lambda^o_{p(n),s(x)} \tau^o_{n,x}} \right) \left(\sqrt{2\pi\sigma^2_{s(x)}} \Phi(\frac{\gamma_{r(c(x)),s(x)}}{\sigma_{s(x)}}) \right)^{-1} \\
& e^{-\frac{1}{2} \frac{\left(\delta_x - \gamma_{r(c(x)),s(x)}\right)^2}{\sigma^2_{s(x)}}} 1_{\{\delta_x \geq 0\}} 1_{\{\delta_x + \sum_{(n,o) \in B_x} \tau^o_{n,x} = t_x\}} \Bigg] g(\boldsymbol{\theta}),
\end{aligned} \tag{A1}$$

where g summarizes the prior and hyperprior distributions as unpacked below. The likelihood is the product of five factors:

1. The product of the densities of the independent normal variables $z_{n,x}$ with means $\alpha_{p,s} = \mu^{(\alpha)}_p + \alpha'_{p,s}$,
2. the product of indicator variables coding whether $\boldsymbol{z}_x$ is consistent with the path B_x and whether the path B_x is consistent with the observed category c_x,
3. exponential densities for the process-completion times $\tau^o_{n,x}$ with rate parameters $\lambda^o_{p,s} = \exp(\mu^{(\beta)}_{o,p} + \beta'_{o,p,s})$ and the truncated normal density for the residual encoding and motor-execution time δ_x with parameters

$\gamma_{r,s} = \mu_r^{(\gamma)} + \gamma'_{r,s}$ and σ_s^2,

4. an indicator function coding whether these times are consistent with the observed response latency, and
5. the function g characterizing the prior and hyperprior distributions.

In this equation, we suppressed, as is usual for exponential variates, indicator functions coding that the individual process-completion times must be non-negative, but we will need this fact below where we consider the conditional distribution of $(\boldsymbol{\tau}_x, \delta_x)_{x=1,\ldots,N}$.

A.3. The Gibbs Sampler

The Gibbs sampler is an MCMC algorithm for sampling from the posterior distribution of the model parameters given the data $(c_x, t_x)_{x=1,\ldots,N}$. It cycles through blocks of parameters. For each block, one sample is drawn from the conditional distribution of the parameters of the block given the data and the remaining parameters. In what follows, we characterize the conditional distributions involved and briefly describe how we sampled from them for non-standard distributions.

A.3.1. The Augmented Data

The conditional distribution of $(B_x, \boldsymbol{z}_x, \boldsymbol{\tau}_x, \delta_x)_{x=1,\ldots,N}$ is sampled trial by trial. For a given trial x, we first sample B_x from the conditional distribution of paths B with normal variates $\boldsymbol{z}_x$ and process-completion and encoding/response-execution times $(\boldsymbol{\tau}_x, \delta_x)$ integrated out, followed by sampling from the conditional distribution of $\boldsymbol{z}_x$ given B_x, the other parameters, and the data, with $(\boldsymbol{\tau}, \delta_x)$ integrated out, followed by sampling from the conditional distribution of $(\boldsymbol{\tau}_x, \delta_x)_{x=1,\ldots,N}$ given B_x, $\boldsymbol{z}_x$, the other parameters, and the data.

Like in Klauer (2010), it can be shown that the distribution of B_x ending in c_x is given by Equation 7 in the body of the paper. This defines a multinomial distribution from which paths were sampled. Given a path B_x, the normal variates $\boldsymbol{z}_x$ can be sampled from truncated and non-truncated normal distributions as described in the paper (Section "Algorithm").

Consider next the conditional distribution of $(\boldsymbol{\tau}, \delta_x)$ given the path B_x, $\boldsymbol{z}_x$, and the other parameters. For a given trial x and path B_x, the conditional distribution of $(\boldsymbol{\tau}_x, \delta_x)$ is a function of the exponential rates $\lambda^o_{p(n),s(x)}$ attached to the edges (n, o) of the path as well as the parameters $\gamma_{r(c(x)),s(x)} = \mu^{(\gamma)}_{r(c(x))} + \gamma'_{r(c(x)),s(x)}$ and $\sigma_{s(x)}$ governing the residual time δ_x.

Process-completion times $\tau^o_{n,x}$ for edges (n, o) in subtree(x), but not on the given path B_x can be sampled from the exponential distribution with rate $\lambda^o_{p(n),s}$ without further constraint (see also Footnote 7 in the body of the paper). The process-completion times and the residual time along the given path B_x must, however, add up to t_x. This means that one of these component times can be expressed in terms of the other times. Let (m, q) be one of the edges (n, o) in B_x with minimum rate parameter: $\lambda^q_{p(m),s(x)} = \min_{(n,o)\in B_x} \lambda^o_{p(n),s(x)}$. We sample $((\tau^o_{n,x})_{(n,o)\in B_x}, \delta_x)$ in three steps.

First, we sample process-completion times $\tau^o_{n,x}$ for edges other than (m, q) from the conditional distribution with δ_x integrated out, followed by sampling δ_x from the conditional distribution of δ_x given $\boldsymbol{\tau}_x$, all other parameters and the data. Finally, $\tau^q_{m,x}$ is set to $t_x - \delta_x - \sum_{(n,o)\in B_x:(n,o)\neq(m,q)} \tau^o_{n,x}$.

Collecting the other parameters and data in $\mathcal{P}$, the conditional distribution $p = p((\tau^o_{n,x})_{(n,o)\in B_x:(n,o)\neq(m,q)}, \delta_x) \mid \mathcal{P})$ of $((\tau^o_{n,x})_{(n,o)\in B_x:(n,o)\neq(m,q)}, \delta_x)$ given $\mathcal{P}$ is characterized by:

$$\begin{aligned}
p &\propto \lambda^q_{p(m),s(x)} e^{-\lambda^q_{p(m),s(x)}\tau^q_{m,x}} \prod_{(n,o)\in B_x:(n,o)\neq(m,q)} \lambda^o_{p(n),s(x)} e^{-\lambda^o_{p(n),s(x)}\tau^o_{n,x}} \\
&\quad 1_{\{\delta_x\geq 0\}} 1_{\{\tau^q_{m,x}\geq 0\}} e^{-\frac{1}{2}\frac{\left(\delta_x-\gamma_{r(c(x)),s(x)}\right)^2}{\sigma^2_{s(x)}}} \\
&\propto \prod_{(n,o)\in B_x:(n,o)\neq(m,q)} (\lambda^o_{p(n),s(x)}-\lambda^q_{p(m),s(x)}) e^{-(\lambda^o_{p(n),s(x)}-\lambda^q_{p(m),s(x)})\tau^o_{n,x}} \\
&\quad 1_{\{0\leq\delta_x\leq t_x-\sum_{(n,o)\in B:(n,o)\neq(m,q)}\tau^o_{n,x}\}} e^{-\frac{1}{2}\frac{\left(\delta_x-(\gamma_{r(c(x)),s(x)}+\lambda^q_{p(m),s(x)}\sigma^2_{s(x)})\right)^2}{\sigma^2_{s(x)}}},
\end{aligned}$$

where components in the above product of exponential densities with rate parameters $\lambda^o_{p(n),s(x)} - \lambda^q_{p(m),s(x)}$ and $\lambda^o_{p(n),s(x)} = \lambda^q_{p(m),s(x)}$ should be replaced by a constant. The second expression in the above characterization of p is obtained by replacing $\tau^q_{m,x}$ by $t_x - \delta_x - \sum_{(n,o)\in B_x:(n,o)\neq(m,q)} \tau^o_{n,x}$ and simple manipulations. Integrating out δ_x, the conditional distribution of $(\tau^o_{n,x})_{(n,o)\in B_x:(n,o)\neq(m,q)}$ given the other parameters and data is proportional to:

$$\begin{aligned}
p(&\ (\tau^o_{n,x})_{(n,o)\in B_x:(n,o)\neq(m,q)} \mid \mathcal{P}) \propto \\
&\prod_{(n,o)\in B_x:(n,o)\neq(m,q)} (\lambda^o_{p(n),s(x)}-\lambda^q_{p(m),s(x)}) e^{-(\lambda^o_{p(n),s(x)}-\lambda^q_{p(m),s(x)})\tau^o_{n,x}} \\
&\left[\Phi\left(\frac{t_x-\sum_{(n,o)\in B:(n,o)\neq(m,q)}\tau^o_{n,x}-(\gamma_{r(c(x)),s(x)}+\lambda^q_{p(m),s(x)}\sigma^2_{s(x)})}{\sigma_{s(x)}}\right)\right. \\
- &\left.\Phi\left(\frac{-(\gamma_{r(c(x)),s(x)}+\lambda^q_{p(m),s(x)}\sigma^2_{s(x)})}{\sigma_{s(x)}}\right)\right] 1_{\{t_x-\sum_{(n,o)\in B:(n,o)\neq(m,q)}\tau^o_{n,x}\geq 0\}}.
\end{aligned}$$

To sample from this distribution, we sequentially sample the τ values along the edges (n, o) of B_x other than (m, q) from the respective exponential distribution with rate parameters $\lambda^o_{p(n),s(x)} - \lambda^q_{p(m),s(x)}$ truncated from above by t_x (or from a uniform distribution on $[0, t_x]$ in the event that $\lambda^o_{p(n),s(x)} = \lambda^q_{p(m),s(x)}$). If in this process, the sum of values already sampled exceeds t_x, we start afresh, amounting to rejection sampling to satisfy the constraint encoded in the indicator function $1_{\{t_x - \sum_{(n,o)\in B:(n,o)\neq(m,q)} \tau^o_{n,x} \geq 0\}}$. A complete set of τ values for (n, o) in B_x with $(n, o) \neq (m, q)$ emerging from this sampling scheme follows a distribution with density proportional to the above (non-normalized) density without the factor given by the difference of the two cumulative normal distributions, which we refer to as $\Phi_1 - \Phi_2$. We can thus "add" this factor in a final rejection-sampling step by drawing a random value u from a uniform distribution, accepting the set of $\tau-$values if

$$u < \frac{\Phi_1 - \Phi_2}{\Phi\left(\frac{t_x - (\gamma_{r(c(x)),s(x)} + \lambda^q_{p(m),s(x)}\sigma^2_{s(x)})}{\sigma_{s(x)}}\right) - \Phi_2}$$

and starting anew otherwise.

Next, from the above expression for the conditional distribution of $((\tau^o_{n,x})_{(n,o)\in B_x:(n,o)\neq(m,q)}, \delta_x)$, it is easy to see that the conditional distribution of δ_x given the $\tau-$values other than $\tau^q_{m,x}$ and given the other parameters and data is a doubly truncated normal distribution with mean $\gamma_{r(c(x)),s(x)} + \lambda^q_{p(m),s(x)}\sigma^2_{s(x)}$, variance $\sigma^2_{s(x)}$, lower bound zero, and upper bound $t_x - \sum_{(n,o)\in B:(n,o)\neq(m,q)} \tau^o_{n,x}$. Having sampled a new δ_x from this distribution, the new $\tau^q_{m,x}$ is finally set to $t_x - \delta_x - \sum_{(n,o)\in B:(n,o)\neq(m,q)} \tau^o_{n,x}$.

A.3.2. The Person-Level Process Parameters

Sampling $(\boldsymbol{\alpha}'_s)_{s=1,\ldots,S}$. Following Gelman and Hill (2007), we implement the scaled inverse Wishart distribution for the variance-covariance matrix $\boldsymbol{\Sigma}$ of $\begin{pmatrix}\boldsymbol{\alpha}'_s \\ \boldsymbol{\beta}'_s\end{pmatrix}$ by further decomposing $\alpha'_{p,s}$ and $\beta'_{o,p,s}$ into $\alpha'_{p,s} = \xi^{(\alpha)}_p \alpha''_{p,s}$ and $\beta'_{o,p,s} =$

$\xi^{(\beta)}_{o,p}\beta''_{o,p,s}$ using the scale factors $\boldsymbol{\xi} = \begin{pmatrix}\boldsymbol{\xi}^{(\alpha)}\\ \boldsymbol{\xi}^{(\beta)}\end{pmatrix}$. A multivariate normal with zero mean and variance-covariance matrix $\boldsymbol{Q}$ is assumed as prior for the unscaled person-level parameters $\begin{pmatrix}\boldsymbol{\alpha}''_s\\ \boldsymbol{\beta}''_s\end{pmatrix}$, and normal priors with mean 1.0 and variance ϵ^{-1} are assumed for the scale factors. $\boldsymbol{\Sigma}$ is thereby decomposed into

$$\boldsymbol{\Sigma} = \mathrm{Diag}\left(\boldsymbol{\xi}\right)\boldsymbol{Q}\;\mathrm{Diag}\left(\boldsymbol{\xi}\right),$$

where Diag is a diagonal matrix of dimension $3P \times 3P$ with the elements of the vector it takes as argument as diagonal elements. If Q follows the Inverse-Wishart distribution (with the identity matrix as scale matrix and $3P+1$ degrees of freedom), then $\boldsymbol{\Sigma}$ is distributed as a scaled Inverse-Wishart distribution as desired. Reflecting this decomposition, we separately sample from the conditional distributions of $(\boldsymbol{\alpha}''_s)_{s=1,\ldots,S}$ and $\boldsymbol{\xi}^{(\alpha)}$.

Denote by $\mathcal{T}(p,s)$ the subset of pairs of nodes and trials, (n,x), with trial x administered to participant $s(x) = s$ and node n stemming from the trial's subtree, $n \in \mathrm{subtree}(x)$, such that process p is attached to the node, $p(n) = p$. It follows:

$$p((\boldsymbol{\alpha}''_s)_{s=1,\ldots,S} \mid \mathcal{P}) \;\propto\; \prod_{s=1}^{S}\left[\prod_{p=1}^{P}\prod_{(n,x)\in\mathcal{T}(p,s)}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(z_{n,x}-\xi_p^{(\alpha)}\alpha''_{p,s}-\mu_p^{(\alpha)}\right)^2}\right] e^{-\frac{1}{2}\begin{pmatrix}\boldsymbol{\alpha}''_s\\ \boldsymbol{\beta}''_s\end{pmatrix}^t \boldsymbol{Q}^{-1}\begin{pmatrix}\boldsymbol{\alpha}''_s\\ \boldsymbol{\beta}''_s\end{pmatrix}}$$

Partition $\boldsymbol{Q}^{-1}$ as follows:

$$\boldsymbol{Q}^{-1} = \begin{pmatrix} (\boldsymbol{Q}^{-1})_{11} & (\boldsymbol{Q}^{-1})_{12} \\ (\boldsymbol{Q}^{-1})_{21} & (\boldsymbol{Q}^{-1})_{22} \end{pmatrix}$$

where the dimensions of $(\boldsymbol{Q}^{-1})_{11}$, $(\boldsymbol{Q}^{-1})_{12}$, and $(\boldsymbol{Q}^{-1})_{22}$ are, in order, $P \times P$, $P \times 2P$, and $2P \times 2P$. Let $N(p,s)$ be the number of pairs (n,x) in $\mathcal{T}(p,s)$. Finally, let $\boldsymbol{R} = [(\boldsymbol{Q}^{-1})_{11} + \mathrm{Diag}\left((N(p,s)(\xi_p^{(\alpha)})^2)_{p=1,\dots,P}\right)]^{-1}$. Standard manipulations show that the conditional distribution of $\boldsymbol{\alpha}_s''$ for a given s is a multivariate normal with mean

$$\boldsymbol{R}\left[\begin{pmatrix} \dots \\ \xi_p^{(\alpha)} \sum_{(n,x)\in\mathcal{T}(p,s)} (z_{n,s,x} - \mu_p) \\ \dots \end{pmatrix} - (\boldsymbol{Q}^{-1})_{12}\boldsymbol{\beta}_s''\right]$$

and variance-covariance matrix $\boldsymbol{R}$. Standard methods exist for sampling from multivariate normal distributions.

The conditional distribution of $\boldsymbol{\xi^{(\alpha)}}$, on the other hand, is given by

$$p(\boldsymbol{\xi^{(\alpha)}} \mid \mathcal{P}) \;\propto\; \prod_{p=1}^{P}\left[\prod_{s=1}^{S}\prod_{(n,x)\in\mathcal{T}(p,s)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(z_{n,s,x} - \xi_p^{(\alpha)}\alpha_{p,s}'' - \mu_p^{(\alpha)}\right)^2}\right] e^{-\frac{1}{2}\epsilon(\xi_p^{(\alpha)}-1)^2},$$

where ϵ is the prior precision. It is easy to see that the $\xi_p^{(\alpha)}$ thereby follow independent normal distributions with posterior variance $\sigma^2_{\text{post}} = (\sum_{s=1}^{S} N(p,s)(\alpha_{p,s}'')^2 + \epsilon)^{-1}$ and mean $\sigma^2_{\text{post}}(\sum_{s=1}^{S} \alpha_{p,s}'' \sum_{(n,x)\in\mathcal{T}(p,s)} (z_{n,s,x} - \mu_p) + \epsilon)$.

Sampling $(\boldsymbol{\beta}_s')_{s=1,\dots,S}$. Again, we sample separately from the conditional distribution of $\boldsymbol{\beta}''$ and from the conditional distribution of $\boldsymbol{\xi^{(\beta)}}$. Consider $\boldsymbol{\beta}''$ first:

$$p((\boldsymbol{\beta}''_s)_{s=1\ldots,S} \mid \mathcal{P}) \propto \prod_{s=1}^{S} \left[\left(\prod_{p=1}^{P} \prod_{(n,x)\in\mathcal{T}(p,s)} \prod_{o=-,+} \exp(\xi^{(\beta)}_{o,p}\beta''_{o,p,s}) e^{-\exp(\mu^{(\beta)}_{o,p}+\xi^{(\beta)}_{o,p}\beta''_{o,p,s})\tau^{o}_{n,x}} \right) e^{-\frac{1}{2}\begin{pmatrix}\boldsymbol{\alpha}''_s\\ \boldsymbol{\beta}''_s\end{pmatrix}^{t} \boldsymbol{Q}^{-1} \begin{pmatrix}\boldsymbol{\alpha}''_s\\ \boldsymbol{\beta}''_s\end{pmatrix}} \right].$$

There is no easy way to sample from this distribution. We proceeded by sampling from the conditional distribution of each individual $\beta''_{o,p,s}$ given the other $\beta''-$parameters, and the other model parameters and data $\mathcal{P}$. It is not difficult to show that the density of this distribution is log-concave throughout and hence, amenable to adaptive rejection sampling (Gilks & Wild, 1992), which is the sampling method adopted.

The conditional distribution of $\boldsymbol{\xi}^{(\boldsymbol{\beta})}$ on the other hand is proportional to:

$$p(\boldsymbol{\xi}^{(\beta)}|\mathcal{P}) \propto \prod_{p=1}^{P} \prod_{o=-,+} e^{-\frac{1}{2}\epsilon(\xi^{(\beta)}_{o,p}-1)^2} \left(\prod_{s=1}^{S} \prod_{(n,x)\in\mathcal{T}(p,s)} \exp(\xi^{(\beta)}_{o,p}\beta''_{o,p,s}) e^{-\exp(\mu^{(\beta)}_{o,p}+\xi^{(\beta)}_{o,p}\beta''_{o,p,s})\tau^{o}_{n,x}} \right)$$

Again, there is no easy way to sample from this distribution, and we employed adaptive rejection sampling to sample from the conditional distribution of each $\xi^{(\beta)}_{o,p}$ given the other parameters using adaptive rejection sampling.

A.3.3. The Population-Level Process-Related Parameters

The conditional distribution of $\boldsymbol{\mu}^{(\boldsymbol{\alpha})}$. Using standard manipulations, it is not difficult to see that the $\mu^{(\alpha)}_p$ follow independent normal distributions with means

$$\frac{\sum_{s=1}^{S}\sum_{(n,x)\in\mathcal{T}(p,s)}(z_{n,x}-\xi_p^{(\alpha)}\alpha''_{p,s})}{\sum_{s=1}^{S}N(p,s)+\epsilon}$$

and variance $(\sum_{s=1}^{S}N(p,s)+\epsilon)^{-1}$. Sampling proceeded from these normal distributions.

Sampling $\boldsymbol{\mu^{(\beta)}}$. We sample the parameters $\boldsymbol{\mu^{(\beta)}}$ on the original (not log-transformed) scale and thus in terms of parameters $\rho_p^o = \exp(\mu_{o,p}^{(\beta)})$. The conditional distribution of ρ_p^o given the other parameters and data, collected in $\mathcal{P}$ is characterized by:

$$p(\rho_p^o \mid \mathcal{P}) \quad \propto \quad (\rho_p^o)^{[\sum_{s=1}^{S}N(p,s)+1]-1} e^{-\rho_p^o\left([\sum_{s=1}^{S}\exp(\xi_{o,p}^{(\beta)}\beta''_{o,p,s})\sum_{(n,x)\in\mathcal{T}(p,s)}\tau_{n,x}^o]+\frac{1}{10}\right)},$$

which defines a Gamma distribution with shape parameters $\sum_{s=1}^{S}N(p,s)+1$ and rate parameter $[\sum_{s=1}^{S}\exp(\xi_{o,p}^{(\beta)}\beta''_{o,p,s})\sum_{(n,x)\in\mathcal{T}(p,s)}\tau_{n,x}^o]+\frac{1}{10}$. Sampling proceeded from these Gamma distributions.

The conditional distribution of $\boldsymbol{Q}$. The conditional distribution is an Inverse-Wishart with $S+3P+1$ degrees of freedom and $I+C$ as scale matrix, where C is the sum of cross-products of the person-level deviations:

$$C = \sum_{s=1}^{S}\begin{pmatrix}\boldsymbol{\alpha}''_s\\ \boldsymbol{\beta}''_s\end{pmatrix}\begin{pmatrix}\boldsymbol{\alpha}''_s\\ \boldsymbol{\beta}''_s\end{pmatrix}^t.$$

A.3.4. The Person-Level Encoding and Response-Execution Parameters

Sampling $\boldsymbol{\gamma}'_s$. Like before, we implement the scaled Wishart distribution for the variance-covariance matrix $\boldsymbol{\Gamma}$ of the person effect parameters $\boldsymbol{\gamma}'_s$ by further decomposing $\gamma'_{r,s}$ into $\gamma'_{r,s} = \xi_r^{(\gamma)}\gamma''_{r,s}$ using the scale factors $\boldsymbol{\xi}^{(\gamma)}$. A multivariate prior with zero mean and variance-covariance matrix $\boldsymbol{S}$ is assumed for the unscaled person-level parameters $\boldsymbol{\gamma}''_s$, and as before, independent normal priors with mean 1.0 and variance ϵ^{-1} are assumed for the scale factors.

Denote by $\mathcal{U}(r,s)$ the subset of observed trials x administered to participant $s(x) = s$ with response $r = r(c(x))$, and by $N_{r,s}$ the number of such trials. It follows:

$$p((\boldsymbol{\gamma}''_s)_{s=1,\dots,S} \mid \mathcal{P}) \;\propto\; \prod_{s=1}^{S}\left[\prod_{r=1}^{R}\left(\Phi(\frac{\mu_r^{(\gamma)} + \xi_r^{(\gamma)}\gamma''_{r,s}}{\sigma_s})\right)^{-N_{r,s}} \prod_{x\in\mathcal{U}(r,s)} \frac{1}{\sqrt{2\pi}\sigma_s} e^{-\frac{1}{2}\frac{\left(\delta_x - \xi_r^{(\gamma)}\gamma''_{r,s} - \mu_r^{(\gamma)}\right)^2}{\sigma_s^2}}\right] e^{-\frac{1}{2}(\boldsymbol{\gamma}''_s)^t \boldsymbol{S}^{-1}(\boldsymbol{\gamma}''_s)}$$

If it were not for the factors $\left(\Phi(\frac{\mu_r^{(\gamma)}+\xi_r^{(\gamma)}\gamma''_{p,s}}{\sigma_s})\right)^{-N_{r,s}}$, the parameters $\boldsymbol{\gamma}''_s$ could be sampled for each participant s from a multivariate normal analogous to the sampling of $\boldsymbol{\alpha}''_s$. We sample from this multivariate distribution, using it as the proposal distribution for a Metropolis-within-Gibbs step, in which we accept the proposal $\boldsymbol{\gamma}^*_s$, if for a random value u drawn from a uniform distribution

$$u \leq \frac{\prod_{r=1}^{R}\left(\Phi(\frac{\mu_r^{(\gamma)}+\xi_r^{(\gamma)}\gamma^*_{r,s}}{\sigma_s})\right)^{-N_{r,s}}}{\prod_{r=1}^{R}\left(\Phi(\frac{\mu_r^{(\gamma)}+\xi_r^{(\gamma)}\gamma''_{r,s}}{\sigma_s})\right)^{-N_{r,s}}},$$

and keep the old values $\boldsymbol{\gamma}''_s$ otherwise. Because the person-wise mean of residual encoding and response-execution times $\mu_r^{(\gamma)} + \xi_r^{(\gamma)}\gamma''_{r,s}$ is usually large relative to the residual variance, σ_s, the factors $\left(\Phi(\frac{\mu_r^{(\gamma)}+\xi_r^{(\gamma)}\gamma''_{r,s}}{\sigma_s})\right)^{-1}$ are very close to one both in the nominator and in the denominator of the above fraction so that the fraction itself is close to one and most proposals are accepted.

The conditional distribution of $\boldsymbol{\xi}^{(\gamma)}$, on the other hand, is given by

$$p(\boldsymbol{\xi}^{(\gamma)} \mid \mathcal{P}) \propto \prod_{r=1}^{R} \Bigg[\prod_{s=1}^{S} \left(\Phi(\frac{\mu_r^{(\gamma)} + \xi_r^{(\gamma)} \gamma''_{r,s}}{\sigma_s}) \right)^{-N_{r,s}} \prod_{x \in \mathcal{U}(r,s)} \frac{1}{\sqrt{2\pi}\sigma_s} e^{-\frac{1}{2} \frac{\left(\delta_x - \xi_r^{(\gamma)} \gamma''_{r,s} - \mu_r^{(\gamma)}\right)^2}{\sigma_s^2}} \Bigg] e^{-\frac{1}{2}\epsilon(\xi_r^{(\gamma)}-1)^2},$$

where ϵ is the prior precision. If it were not for the factors involving Φ, the parameters $\boldsymbol{\xi}^{(\gamma)}$ could be sampled from independent normal distributions for each response r analogous to parameters $\boldsymbol{\xi}^{(\alpha)}$. Like before, we sample proposal values from these normal distributions and "add" the factors involving the Φ through a Metropolis-within-Gibbs step.

The conditional distribution of $(\sigma_s^2)_{s=1,\ldots,S}$. Let N_s be the number of trials x administered to participant s. It is not difficult to see that the conditional distribution of σ_s^2 is proportional to:

$$p(\sigma_s^2 \mid \mathcal{P}) \propto \left(\frac{1}{\sigma_s^2}\right)^{[\frac{N_s+2}{2}+1]} e^{-\frac{N_s+2}{2\sigma_s^2}\left(\frac{\sum_{x:s(x)=s}\left(\delta_x - \xi_r^{(\gamma)}\gamma''_{r,s} - \mu_r^{(\gamma)}\right)^2 + 2\omega^2}{N_s+2}\right)} \prod_{x:s(x)=s} \left(\Phi(\frac{\mu_r^{(\gamma)} + \xi_r^{(\gamma)} \gamma''_{r,s}}{\sigma_s})\right)^{-1},$$

where $r = r(c(x))$ is the response attached to trial x. If it were not for the last factors involving Φ, σ_s^2 could thus be sampled from a scaled inverse χ^2-distribution with degrees of freedom $N_s + 2$ and scale factor $\left(\frac{\sum_{x:s(x)=s}\left(\delta_x - \xi_r^{(\gamma)}\gamma''_{r,s} - \mu_r^{(\gamma)}\right)^2 + 2\omega^2}{N_s+2}\right)$. Again, we used this distribution to generate a proposal value and "add" the factors via a Metropolis-within-Gibbs step.

A.3.5. The Population-Level Encoding and Response-Execution Parameters

The conditional distribution of $\boldsymbol{\mu}^{(\gamma)}$. It is not difficult to see that the conditional distribution of $\mu_r^{(\gamma)}$ is proportional to a normal distribution with variance $\sigma^2_{\text{post}} = (\sum_s \frac{N_{r,s}}{\sigma_s^2} + \frac{1}{10})^{-1}$, and mean $\sigma^2_{\text{post}} \sum_s \frac{\sum_{x \in \mathcal{U}(r,s)} (\delta_x - \xi_r^{(\gamma)} \gamma''_{r,s})}{\sigma_s^2}$ up to the factor $\prod_s \Phi(\frac{\mu_r^{(\gamma)} + \xi_r^{(\gamma)} \gamma''_{r,s}}{\sigma_s})^{-N_{r,s}}$. We sample from the normal distribution to generate a proposal value and again "add" the factors via a Metropolis-within-Gibbs step.

The conditional distribution of $\mathbf{\Gamma}$. Like for the person-level parameters α' and β', the decomposition of $\gamma'_{r,s}$ into $\gamma'_{r,s} = \xi_r^{(\gamma)} \gamma''_{r,s}$ implies that $\mathbf{\Gamma}$ is decomposed into

$$\mathbf{\Gamma} = \text{Diag}\left(\boldsymbol{\xi}^{(\gamma)}\right) \boldsymbol{S} \ \text{Diag}\left(\boldsymbol{\xi}^{(\gamma)}\right).$$

The conditional distribution of $\boldsymbol{S}$ is Inverse-Wishart with $S + R + 1$ degrees of freedom and $I + C$ as scale matrix, where C is the sum of cross-products of the person-level deviations:

$$C = \sum_{s=1}^{S} (\boldsymbol{\gamma}''_s) (\boldsymbol{\gamma}''_s)^t .$$

The conditional distribution of ω^2. It is not difficult to see that the conditional distribution of ω^2 is a Gamma distribution with shape and rate parameters equal to $\frac{S \times df}{2}$ and $\frac{df}{2} \sum_s \frac{1}{\sigma_s^2}$, respectively, where $df = 2$ for the applications presented in the main body of the text.

A.4. Model Selection for the Models with Equal Response-Execution Times for Old and New Responses

Table A1 shows DIC values and the Model Checks for DG and DI Models with response-execution times for old and new responses set equal for the datasets analyzed in the body of the paper (see Table 1). As can be seen, DIC is uniformly larger than for the same models with unequal response-execution times for old and new responses. The model checks are often, but not always associated with smaller p values. The relatively smaller impact of setting equal the response-execution times on the model checks than on the DIC values suggests that having separate parameters for old and new responses was less important for fitting the mean frequencies and response times averaged across participants – as also suggested by the fact that the HDI's of the $\mu_r^{(\gamma)}$ for old and new responses in most cases overlap (see Tables 2 and 3) – than for accounting for individual differences, perhaps due to differences in handedness, between participants.

Table A1
DIC values and Model Checks (Posterior p Values) for the DG and the DI Models with Equal Response-Execution Times for the Old and the New Response

	DG Variant				DI Variant			
Data	ΔDIC	X_1	X_2	X_3	ΔDIC	X_1	X_2	X_3
— Arnold et al. (2015) —								
Exp. 1	267.70	.48	.27	.51	283.63	.44	.44	.28
Exp. 2	189.27	.49	.37	.37	203.62	.48	.29	.44
Exp. 3	243.99	.40	.57	.08	218.22	.43	.64	.05
— Dube et al. (2012) —								
Exp. 1	157.80	<.0001	.02	.15	106.25	.27	.01	.23
Exp. 2	224.66	<.0001	.05	.48	133.31	.0001	.01	.62

Note. DG = "Detect-Guess"; DI = "Default-Interventionist"; ΔDIC = DIC difference from the lowest DIC of the dataset from Table 1 in the body of the text.