Extending multinomial processing tree models to measure the relative speed of cognitive processes

Daniel W. Heck¹ · Edgar Erdfelder¹

Published online: 16 June 2016
© Psychonomic Society, Inc. 2016

Abstract Multinomial processing tree (MPT) models account for observed categorical responses by assuming a finite number of underlying cognitive processes. We propose a general method that allows for the inclusion of response times (RTs) into any kind of MPT model to measure the relative speed of the hypothesized processes. The approach relies on the fundamental assumption that observed RT distributions emerge as mixtures of latent RT distributions that correspond to different underlying processing paths. To avoid auxiliary assumptions about the shape of these latent RT distributions, we account for RTs in a distribution-free way by splitting each observed category into several bins from fast to slow responses, separately for each individual. Given these data, latent RT distributions are parameterized by probability parameters for these RT bins, and an extended MPT model is obtained. Hence, all of the statistical results and software available for MPT models can easily be used to fit, test, and compare RT-extended MPT models. We demonstrate the proposed method by applying it to the two-high-threshold model of recognition memory.

Keywords Cognitive modeling · Response times · Mixture models · Processing speed

Electronic supplementary material The online version of this article (doi:10.1007/s13423-016-1025-6) contains supplementary material, which is available to authorized users.

✉ Daniel W. Heck
dheck@mail.uni-mannheim.de

¹ Department of Psychology, School of Social Sciences,
University of Mannheim, Schloss EO 254, 68131 Mannheim,
Germany

Many substantive psychological theories assume that observed behavior results from one or more latent cognitive processes. Because these hypothesized processes can often not be observed directly, measurement models are important tools to test the assumed cognitive structure and to obtain parameters quantifying the probabilities that certain underlying processing stages take place or not. Multinomial processing tree models (MPT models; Batchelder & Riefer, 1990) provide such a means by modeling observed, categorical responses as originating from a finite number of discrete, latent processing paths. MPT models have been successfully used to explain behavior in many areas such as memory (Batchelder & Riefer, 1986, 1990), decision making (Erdfelder, Castela, Michalkiewicz, & Heck, 2015; Hilbig, Erdfelder, & Pohl, 2010), reasoning (Klauer, Voss, Schmitz, & Teige-Mocigemba, 2007), perception (Ashby, Prinzmetal, Ivry, & Maddox, 1996), implicit attitude measurement (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005; Nadarevic & Erdfelder, 2011), and processing fluency (Fazio, Brashier, Payne, & Marsh, 2015; Unkelbach & Stahl, 2009). Batchelder & Riefer (1999) and Erdfelder et al. (2009) reviewed the literature and showed the usefulness and broad applicability of the MPT model class. In the present paper, we introduce a simple but general approach to include information about response times (RTs) into any kind of MPT model.

As a running example, we will use one of the most simple MPT models, the two-high-threshold model of recognition memory (2HTM; Bröder & Schütz, 2009; Snodgrass & Corwin, 1988). The 2HTM accounts for responses in a binary recognition paradigm. In such an experiment, participants first learn a list of items and later are prompted to categorize old and new items as such. Hence, one obtains frequencies of hits (correct *old*), misses (incorrect *new*), false alarms (incorrect *old*), and correct rejections (correct

Fig. 1 The two-high threshold model of recognition memory

new responses). The 2HTM, shown in Fig. 1, assumes that hits emerge from two distinct processes: Either a memory signal is sufficiently strong to exceed a high threshold and the item is recognized as *old*, or the signal is too weak, an uncertainty state is entered, and respondents only guess *old*. The two processing stages of target detection and guessing conditional on the absence of detection are parameterized by the probabilities of their occurrence d_o and g , respectively. Given that the two possible processing paths are disjoint, the overall probability of an *old* response to an old item is given by the sum $d_o + (1 - d_o)g$. Similarly, correct rejections can emerge either from lure detection with probability d_n or from guessing *new* conditional on nondetection with probability $1 - g$. In contrast, incorrect *old* and *new* responses always result from incorrect guessing.

The validity of the 2HTM has often been tested in experiments by manipulating the base rate of learned items, which should only affect response bias and thus the guessing parameter g (Bröder & Schütz, 2009; Dube, Starns, Rotello, & Ratcliff, 2012). If the memory strength remains constant, the model predicts a linear relation between the probabilities of hits and false alarms (i.e., a linear receiver-operating characteristic, or ROC, curve; Bröder & Schütz, 2009; Kellen, Klauer, & Bröder, 2013). The 2HTM is at the core of many other MPT models that account for more complex memory paradigms such as source memory (Bayen, Mur-nane, & Erdfelder, 1996; Klauer & Wegener, 1998; Meiser & Böder, 2002) or process dissociation (Buchner, Erdfelder, Steffens, & Martensen, 1997; Jacoby, 1991; Steffens, Buchner, Martensen, & Erdfelder, 2000). These more complex models have a structure similar to the 2HTM because they assume that correct responses either result from some memory processes of theoretical interest or from some kind of guessing.

Whereas MPT models are valuable tools to disentangle cognitive processes based on categorical data, they lack the ability to account for response times (RTs). Hence, MPT models cannot be used to test hypotheses about the speed of the assumed cognitive processes, for example, whether one underlying process is faster than another one. However, modeling RTs has a long tradition in experimental

psychology, for instance, in testing whether cognitive processes occur serially or in parallel (Luce, 1986; Townsend & Ashby, 1983). Given that many MPT models have been developed for cognitive experiments that are conducted with the help of computers under controlled conditions, recording RTs in addition to categorical responses comes at a small cost. Even more importantly, substantive theories implemented as MPT models might readily provide predictions about the relative speed of the hypothesized processes or about the effect of experimental manipulations on processing speeds. For instance, the 2HTM can be seen as a two-stage serial process model in which guessing occurs only after unsuccessful detection attempts (see, e.g., Dube et al., 2012; Erdfelder, Küpper-Tetzl, & Mattern, 2011). Given this assumption of serial processing stages, the 2HTM predicts that, for both target and lure items, responses based on guessing are slower than responses based on detection (Province & Rouder, 2012). This hypothesis, however, cannot directly be tested because it concerns RT distributions of unobservable processes instead of directly observable RT distributions.

To our knowledge, there are mainly three general approaches to use information about RTs in combination with MPT models. First, Hu (2001) developed a method, based on an approach of Link (1982), that decomposes the mean RTs of the observed categories based on the estimated parameters of an MPT model. This approach assumes a strictly serial sequence of processing stages. Each transition from one latent stage to another is assigned with a mean processing time that is assumed to be independent of the original core parameters of the MPT model. Within each branch, all of the traversed mean processing times sum up to a mean observed RT. Despite its simplicity, this approach has not been applied often in the literature. One reason might be that the assumption of a strictly serial sequence of processing stages is too restrictive for many MPT models. Moreover, the method does not allow for testing different structures on the latent processing times because MPT parameters and mean processing speeds are estimated separately in two steps and not jointly in a single model.

The second approach of combining RTs with MPT models involves directly testing qualitative predictions for specific response categories (e.g., Dube et al., 2012; Erdfelder et al., 2011; Hilbig & Pohl, 2009). For instance, Erdfelder et al. (2011) and Dube et al. (2012) derived the prediction for the 2HTM that the mean RT of guessing should be slower than that of detection if guesses occur serially after unsuccessful detection attempts.¹ Moreover, a stronger

¹Note that this order constraint on mean RTs of guessing and detection is implied by the stronger assumption of stochastic dominance discussed below.

response bias towards *old* items will result in a larger proportion of slow *old*-guesses. Assuming that response bias manipulations do not affect the speed of detection and the speed of guessing itself, we would thus expect an increase of the mean RT of hits with increasing guessing bias towards *old* responses (Dube et al., 2012). Whereas such indirect, qualitative tests may help clarify theoretical predictions, they often rest on assumptions of unknown validity and might require many paired comparisons for larger MPT models.

A third approach was used by Province and Rouder (2012) and Kellen et al. (2015). Similar to the two-step method by Hu (2001), the standard MPT model is fitted to the individual response frequencies to obtain parameter estimates first. Next, each response and the corresponding RT is assigned to the cognitive process from which it most likely emerged. For example, if a participant with 2HTM parameter estimates $\hat{d}_o = .80$ and $\hat{g} = .50$ produces a hit, then the recruitment probability that this hit was due to detection is estimated as $.80/ (.80 + .20 \cdot .50) = .89$ (cf. Fig. 1). Hence, this response and its RT are assigned to the detection process rather than to the guessing process for which the recruitment probability estimate is only $(.20 \cdot .50) / (.80 + .20 \cdot .50) = .11$. Note that although assignments of responses to latent processes may vary between participants with different MPT parameter estimates, they are necessarily identical within participants. Hence, all RTs of a participant corresponding to a specific response type are always assigned to the same process. Classification errors implied by this procedure are likely to be negligible if recruitment probabilities are close to 1 as in our example but may be substantial if the latter are less informative (i.e., close to .50). Despite this problem, the method provides an approximate test of how experimental manipulations affect process-specific mean RTs. For example, both Province and Rouder (2012) and Kellen et al. (2015) found that the overall mean RTs across processes decreased when the number of repetitions during the learning phase increased. However, the mean RTs for the subsets of responses that were assigned to detection and guessing processes, respectively, were not affected by this experimental manipulation. Instead, the faster overall mean RTs for higher repetition rates could solely be explained by a larger proportion of fast detection responses (i.e., an increase of d_o with repetitions), or equivalently, by a smaller proportion of slow guesses.

All of these three approaches rely on separate, two-step analyses of categorical responses and RTs and do not allow to account for response frequencies and RTs in a single statistical model. Therefore, we propose a novel method that directly includes information about RTs into any MPT model. As explained in the following sections, the method rests on the fundamental idea that MPT models imply finite mixture distributions of RTs for each response

category because they assume a finite number of underlying cognitive processes. Second, instead of modeling RTs as continuous variables, each observed response category is split into discrete bins representing fast to slow responses, similar to a histogram. Based on these more fine-grained categories, the relative speed of each processing branch is represented by probability parameters for the discrete RT bins, resulting in an RT-extended MPT model. Importantly, the underlying, unobservable RT distributions are modeled in a distribution-free way to avoid potentially misspecified distributional assumptions. Third, we introduce a strategy how to impose restrictions on the latent RT distributions in order to ensure their identifiability. Moreover, we discuss how to test hypotheses concerning the ordering of latent processes. Note that the RT-extended MPT model can be fitted and tested using the existing statistical tools for MPT models. We demonstrate the approach using the 2HTM of recognition memory.

MPT models imply mixtures of latent RT distributions

One core property of MPT models is their explicit assumption that a finite number of discrete processing sequences determines response behavior. In other words, each branch in an MPT model constitutes a different, independent cognitive processing path. It is straightforward to assume that each of these possible processes results in some (unknown) distribution of RTs. Given a category that is reached only by a single branch, the distribution of RTs for the corresponding process is directly observable. In contrast, if at least two branches lead to the same category, the observed distribution of RTs within this category will be a finite mixture distribution (Luce, 1986; Hu, 2001; Townsend & Ashby, 1983). This means that each RT of any response category of the MPT model originates from exactly one of the latent RT distributions corresponding to the branches of the MPT model, with mixture probabilities given by the core MPT structure. Mathematically, an observed RT distribution is a mixture of J latent RT distributions with mixture probabilities α_j if the observed density $f(x)$ is a linear combination of the latent densities $f_j(x)$,

$$f(x) = \sum_{j=1}^J \alpha_j f_j(x). \quad (1)$$

Because MPT models necessarily include more than a single response category, separate mixture distributions are assumed for the RT distributions corresponding to the different response categories, where the proportions α_j are defined by the branch probabilities of the core MPT model.

Fig. 2 The six hypothesized processing paths of the 2HTM imply six separate latent RT distributions. The observed RT distributions of hits and correct rejections are mixture distributions of the corresponding latent RT distributions. Solid and dashed densities represent *old* and *new* responses, respectively

This basic idea is illustrated in Fig. 2 for the 2HTM, where the six branches — and thus the six processing sequences — are assumed to have separate latent RT distributions. We use the term *latent* for these distributions because they cannot directly be observed in general. Consider, for instance, the RT distribution of hits: For a single *old* response to a target and the corresponding RT, it is impossible to decide whether it resulted from target detection (first branch) or guessing *old* in the uncertainty state (second branch). However, we know that a single observed RT must stem from one of two latent RT distributions, either from target detection with probability d_o or from guessing *old* with probability $(1 - d_o)g$. Hence, the RTs for hits follow a two-component mixture distribution where the mixture probabilities are proportional to these two MPT branch probabilities.² In contrast to RTs of hits, the observed RT distributions of misses and false alarms are identical to the latent RT distributions of guessing *new* for targets and guessing *old* for lures, respectively, because only a single branch leads to each of these two categories. Note that, later on, some of these latent RT distributions may be restricted to be identical for theoretical reasons or in order to obtain a model that provides unique parameter estimates.

²Note that in order to obtain proper mixture probabilities that sum up to one, these path probabilities have to be normalized within each response category (e.g., for hits in the 2HTM, $P(\text{detect old} | \text{hit}) = d_o / (d_o + (1 - d_o)g)$ and $P(\text{guess old} | \text{hit}) = (1 - d_o)g / (d_o + (1 - d_o)g)$).

Whereas this mixture structure of latent RT distributions emerges directly as a core property of the model class, MPT models do not specify the temporal order of latent processes, that is, whether processes occur in parallel, in partially overlapping order, or in a strictly serial order (Batchelder & Riefer, 1999). In general, both parallel and serial processes can be represented within an MPT structure (Brown, 1998). Whereas the assumption of a serial order of latent processing stages might be plausible for the 2HTM (guessing occurs after unsuccessful detection attempts; cf. Dube et al., 2012; Erdfelder et al., 2011), other MPT models represent latent processes without the assumption of such a simple time course. Moreover, many MPT models can be reparameterized into equivalent versions with a different order of latent processing stages (Batchelder & Riefer, 1999), which prohibits simple tests of the processing order. In some cases, however, factorial designs allow for such tests using only response frequencies (Schweickert & Chen, 2008).

To avoid auxiliary assumptions about the serial or parallel nature of processing stages, our approach is more general and does not rely on an additive decomposition of RTs into processing times for different stages as proposed by Hu (2001). Instead, we directly estimate the latent RT distributions shown in Fig. 2 separately for each process (i.e., for each branch of the MPT model). Nevertheless, once the latent RT distributions are estimated for each branch, the results can be useful to exclude some of the possible processing sequences. For instance, we can test the assumption that detection occurs strictly before guessing by testing whether detection-responses are stochastically faster than guessing-responses (see below).

Another critical assumption concerns the exact shapes of the hypothesized latent RT distributions. These unobservable distributions could, for instance, be modeled by typical, right-skewed distributions for RTs such as ex-Gaussian, log-normal, or shifted Wald distributions (Luce, 1986; Matzke & Wagenmakers, 2009; Van Zandt & Ratcliff, 1995). However, instead of testing only the core structure of an MPT model, the validity of such a parametric model will also rest on its distributional assumptions. Since MPT models exist for a wide range of experimental paradigms, it is unlikely that a single parameterization of RTs will be appropriate for all applications. Moreover, many substantive theories might only provide predictions about the relative speed of cognitive processes instead of specific predictions about the shape of RTs. Given that most MPT models deal with RT distributions that are not directly observable, it is difficult to test such auxiliary parametric assumptions. Therefore, to avoid restrictions on the exact shape of latent RT distributions, we propose a distribution-free RT model.

Categorizing RTs into bins

Any continuous RT distribution can be approximated in a distribution-free way by a histogram (Van Zandt, 2000). Given some fixed boundaries on the RT scale, the number of responses within each bin is counted and displayed graphically. Besides displaying empirical RT distributions, discrete RT bins can also be used to test hypotheses about RT distributions without parametric assumptions. For instance, Yantis, Meyer, and Smith (1991) used discrete bins to test whether observed RT distributions can be described as mixtures of a finite number of basis distributions corresponding to different cognitive states (e.g., being in a prepared or unprepared cognitive state, Meyer, Yantis, Osman, & Smith, 1985). Yantis et al.'s (1991) method is tailored to situations in which all basis distributions can directly be observed in different experimental conditions. Using discrete RT bins, one can then test whether RT distributions in additional conditions are mixtures of these basis distributions. The method is mathematically tractable because the frequencies across RT bins follow multinomial distributions if identically and independently distributed RTs are assumed. Moreover, likelihood ratio tests allow for testing the mixture structure without the requirement of specifying parametric assumptions.

The model has two sets of parameters, the mixture probabilities α_{ij} that responses in condition i emerge from the cognitive state j , and the probabilities L_{jb} that responses of the state j fall into the b -th RT bin. Importantly, the latency parameters $L_{j1}, \dots, L_{jB}$ model the complete RT distribution of the j -th state without any parametric assumptions. Using this discrete parametrization of the basis RT

distributions, the mixture density in Eq. 1 gives the probability that responses in the i -th condition fall into the b -th RT bin,

$$p_{ib} = \sum_{j=1}^J \alpha_{ij} L_{jb}. \quad (2)$$

Given a sufficient amount of experimental conditions and bins, the model can be fitted and tested within the maximum likelihood framework.

Besides the advantage of avoiding arbitrary distributional assumptions, Yantis et al. (1991) showed that the method has sufficient power to detect deviations from the assumed mixture structure given realistic sample sizes between 40 to 90 responses per condition. However, the method is restricted to RT distributions for a single type of response, and it also requires experimental conditions in which the basis distributions can directly be observed. MPT models, however, usually account for at least two different types of responses and do not necessarily allow for direct observations of the underlying component distributions. For instance, in the 2HTM (Fig. 2), the latent RT distributions of target and lure detection cannot directly be observed, because hits and false alarms also contain guesses. Moreover, MPT models pose strong theoretical constraints on the mixture probabilities α_{ij} , which are assumed to be proportional to the branch probabilities.

We retain the benefits of Yantis et al.'s (1991) method and overcome its limitations regarding MPT models by using

Fig. 3 The RT-extended two-high threshold model uses response categorization from fast to slow (Q1 to Q4) to model latent RT distributions. The gray boxes represent the latent RT distributions for target detection, guessing *old*, and guessing *new* (each parameterized by three latency parameters)

discrete RT bins and probability parameters L_{jb} to model the latent RT distributions in Fig. 2. Thereby, instead of assuming a specific distributional shape, we estimate each latent RT distribution directly. This simple substitution produces a new, RT-extended MPT model. To illustrate this, Fig. 3 shows the RT-extended 2HTM for old items. In this example, each of the original categories is split into four bins using some fixed boundaries on the observed RTs. In the 2HTM, this results in a total of $4 \cdot 4 = 16$ categories for the RT-extended MPT model. Once the RTs have been used to obtain these more fine-grained category frequencies, the new MPT model can easily be fitted, tested, and compared to other models by means of existing software for MPT models (e.g., Moshagen, 2010; Singmann & Kellen, 2013).

Note that the number of RT bins can be adjusted to account for different sample sizes. Whereas experiments with many responses per participant allow for using many RT bins, the method can also be applied with small samples by using only two RT bins. The latter represents a special case in which responses are categorized as either fast or slow based on a single RT boundary. As a result, each latent RT distribution j is described by one latency parameter L_j , the probability that RTs of process j are below the individual RT boundary, which is direct measure of the relative speed of the corresponding process.

Choosing RT boundaries for categorization

If the latent RT distributions are modeled by discrete RT bins, the question remains how to obtain RT boundaries to categorize responses into bins. In the following, we discuss the benefits and limitations of three different strategies. One can use (1) fixed RT boundaries, (2) collect a calibration data set, or (3) rely on data-dependent RT boundaries. We used the last of these strategies in the present paper for reasons that will become clear in the following.

Most obviously, one can use fixed RT boundaries that are chosen independently from the observed data. For instance, to get eight RT bins, we can use seven equally-spaced points on a range typical for the paradigm under consideration, for example, from 500 ms to 2,000 ms in steps of 250 ms. Obviously, however, this strategy can easily result in many zero-count cells if the RTs fall in a different range than expected. This is problematic, because a minimum expected frequency count of five per category is a necessary condition to make use of the asymptotic χ^2 test for the likelihood ratio statistic, which is used to test the proposed mixture structure. For category counts smaller than five, the asymptotic test may produce misleading results (Agresti, 2013). Even more importantly, the resulting frequencies are not comparable across participants because of natural differences

in the speed of responding (Luce, 1986). Hence, frequencies cannot simply be summed up across participants, which further complicates the analysis.

As a remedy, it is necessary to obtain RT boundaries that result in comparable bins across participants. If comparable RT bins are defined for all participants on a-priori grounds, it is in general possible to analyze the RT-extended MPT model for the whole sample at once, either by summing individual frequencies or by using hierarchical MPT extensions (Klauer, 2010; Smith & Batchelder, 2010). Hence, as a second strategy, one could use an independent set of data to calibrate the RT boundaries separately for each participant. Often, participants have to perform a learning task similar to the task of interest. We can then use any summary statistics of the resulting RT distribution to define the RT boundaries for the subsequent MPT analysis. For instance, the empirical 25 %-, 50 %-, and 75 %-quantiles ensure that the resulting four RT bins are a-priori equally likely (for B RT bins, one would use the b/B -quantiles with $b = 1, \dots, B - 1$). Given that the RT boundaries are defined for each participant following this rule, the resulting RT bins share the same interpretation and can be analyzed jointly. Besides minimizing zero-count frequencies, the adjusted RT boundaries for each participant eliminate individual RT differences in the RT-extended MPT model.

Often, a separate calibration data set might not be available or costly to obtain. In such a situation, it is possible to use the available data twice: First, to obtain some RT boundaries according to a principled strategy, and second, to analyze the same, categorized data with an RT-extended MPT model. For instance, Yantis et al. (1991) proposed to vary the location and scale of equal-width RT intervals to ensure a minimum of five observations per category. At first view, however, choosing RT boundaries as a function of the same data is statistically problematic. If the boundaries for categorization are fixed as in our first strategy, the multinomial sampling theory is correct and the standard MPT analysis is valid. In contrast, if the RT boundaries are stochastic and depend on the same data, it is not clear whether the resulting statistical inferences are still correct. However, the practical advantages of obtaining RT boundaries from the same data are obviously very large. Therefore, to justify this approach, we provide simulation studies in the [Supplementary Material](#) showing that the following strategy results in correct statistical inferences with respect to point estimates, standard errors, and goodness-of-fit tests.

According to the third strategy, RT boundaries are obtained from the overall RT distribution across all response categories as shown in Fig. 4. For each person, the distribution of RTs across all response categories is approximated by a right-skewed distribution, for instance, a log-normal distribution (Fig. 4a). This approximation serves as a reference distribution to choose reasonable boundaries that result

Fig. 4 A principled strategy to obtain comparable RT bins across individuals. **a** Computation of quantiles from a log-normal approximation to the RT distribution across all response categories. **b** Categorization of hits, misses, false alarms, and correct rejections from fast to slow

in RT bins with a precise interpretation, for example, the b/B -quantiles (the 25 %-, 50 %-, 75 %-quantiles for four RT bins). These individual RT boundaries are then used to categorize responses into B subcategories from fast to slow (Fig. 4b). By applying the same approximation strategy separately for each individual, the new category system has an identical interpretation across individuals. Moreover, the right-skewed approximation results in a-priori approximately equally likely RT bins. Importantly, this approximation is only used to define the RT boundaries. Hence, it is neither necessary that it fits the RT distribution, nor does it constrain the shape of the latent RT distributions because the latency parameters L_{jb} of the RT-extended MPT model can still capture any distribution.

We decided to use a log-normal approximation because it is easy to apply and uses less parameters than other right-skewed distributions such as the ex-Gaussian or the shifted Wald distribution (Van Zandt, 2000). However, in our empirical example below, using these alternative distributions resulted in similar RT boundaries, and hence, in the same substantive conclusions. The detailed steps to derive individual RT boundaries are (1) log-transform all RTs of a person, (2) compute mean and variance of these log-RTs, (3) get b/B -quantiles of a normal distribution with mean and variance of Step 2, (4) re-transform these log-quantiles t_b back to the standard RT scale to

get the boundaries required for categorization (i.e., $RT_b = \exp(t_b)$).³

Note that there is an alternative, completely different approach of using RT boundaries based on the overall RT distribution. For instance, with two RT bins, we can categorize a response as fast or slow if the observed RT is below or above the overall median RT, respectively. However, instead of using an MPT model with latent RT distributions as in Fig. 3, it is possible to use the standard MPT tree in Fig. 1 twice but with separate parameters for fast and slow responses. Hence, this approach allows for testing whether the core parameters (d_o , d_n , and g in case of the 2HTM) are invariant across fast and slow responses. Since our interest here is in direct estimates of the relative speed of the hypothesized processes, we did not pursue this direction any further.

Identifiability: constraints on the maximum number of latent RT distributions

Up to this point, we assumed that each processing path is associated with its own latent RT distribution. However,

³Mathematically more concise, $RT_b = \exp[SD_{\log RT} \cdot (\Phi^{-1}(b/B) + \overline{\log RT})]$, where Φ^{-1} is the inverse cumulative density function of the standard-normal distribution.

assuming a separate latent RT distribution for each processing path will in general cause identifiability problems. For example, Fig. 2 suggests that it is not possible to obtain unique estimates for six latent RT distributions based on only four observed RT distributions (for hits, misses, false alarms, and correct rejections, respectively). Hence, it will usually be necessary to restrict some of these latent distributions in one way or another to obtain an identifiable model (Hu, 2001). For instance, in the 2HTM, one could assume that any type of guessing response is equally fast, thereby reducing the number of latent RT distributions to three. In practice, these restrictions should of course represent predictions based on psychological theory. However, before discussing restrictions on RTs in case of the 2HTM, we will first present a simple strategy how to check that a chosen set of restricted latent RT distributions can actually be estimated.

Before using any statistical model in practice, it has to be shown that fitting the model to data results in unique parameter estimates. A sufficient condition for unique parameter estimates is the global identifiability of the model, that is, identical category probabilities $p(\theta) = p(\theta')$ imply identical parameter values $\theta = \theta'$ for all θ, θ' in the parameter space Ω (Bamber & van Santen, 2000). For larger, more complex MPT models, it can be difficult to proof this one-to-one mapping of parameters to category probabilities analytically (but see Batchelder & Riefer, 1990; Meiser, 2005 for examples of analytical solutions). Therefore, numerical methods based on simulated model identifiability (Moshagen, 2010) or the rank of the Jacobian matrix Schmittmann, Dolan, Raijmakers, & Batchelder, (2010) have been developed to check the local identifiability of MPT models, which ensures the one-to-one relation only in the proximity of a specific set of parameters. All of these methods can readily be applied to RT-extended MPT models.

However, when checking the identifiability of an RT-extended MPT model, the problem can be split into two parts: (1) the identifiability of the core parameters θ from the original MPT model and (2) the identifiability of the additional latency parameters L of the RT-extension given that the original MPT model is identifiable. The first issue has a simple solution. If the original MPT model is globally identifiable, then, by definition, the core parameters θ of the extended MPT model are also globally identifiable. This holds irrespective of the hypothesized structure of latent RT distributions. This simple result emerges from the fact that the original category frequencies are easily recovered by summing up the frequencies across all RT bins within each category (see Appendix A).

Given a globally identifiable original MPT model, the identifiability of the latency parameters L in the RT-

extended model can be checked using a simple strategy. First, all categories are listed that are reached by a single branch. The latency parameters of the respective latent RT distributions of these categories are directly identifiable. Intuitively, the RTs in such a category can be interpreted as process-pure measures of the latencies corresponding to the processing path (similar to the basis distributions of Yantis et al., 1991). Second, for each of the remaining categories, their associated latent RT distributions are listed. If for one of these categories, all but one of the latent RT distributions are identifiable (from the first step), then the remaining latent RT distribution of this category is also identifiable. This second step can be applied repeatedly using the identifiability of RT distributions from previous steps. If all of the hypothesized latent RT distributions are rendered identifiable by this procedure, the whole model is identifiable. Note, however, that this simple strategy might not always provide a definite result. For instance, if one or more latent RT distributions might not be rendered identifiable by this strategy, it is still possible that the model is identifiable. Therefore, the successful application of this simple strategy provides a sufficient but not a necessary condition for the identifiability of an RT-extended MPT model.

In case this simple strategy does not ensure the identifiability of all of the latency parameters, a different, more complex approach can be used. For this purpose, a matrix $P(\theta)$ is defined which has as many rows as categories and as many columns as latent RT distributions. A matrix entry $P(\theta)_{kj}$ is defined as the branch probability $p_{ki}(\theta)$ if the i -th branch of category k is assumed to generate RTs according to the j -th latent RT distribution; otherwise, it is zero (see Appendix A for details). The extended MPT model will be globally identifiable if and only if the rank of this matrix $P(\theta)$ is equal to or larger than the number of latent RT distributions. This general strategy directly implies that an identifiable RT-extended model cannot have more latent RT distributions than observed categories. Otherwise, the rank of the matrix $P(\theta)$ will necessarily be smaller than the number of latent RT distributions. For example, consider the 2HTM as shown in Fig. 2. For this model, the matrix $P(\theta)$ has four rows for the four types of responses and six columns for the latent RT distributions. Hence, this model is not identifiable. Below, we discuss theoretical constraints to solve this issue.

Note that these results about the identifiability of the latency parameters only hold if the core parameters are in the interior of the parameter space, that is, $\theta \in (0, 1)^S$, where S is the number of parameters. In other words, if one of the processing paths can never be reached because some of the core parameters are zero or one, it will be

Fig. 5 **a** The continuous RT distribution F_j stochastically dominates F_i . **b** In an RT-extended MPT model, the latent RT distribution of process j is slower (i.e., RTs are larger) than that of process i

impossible to estimate the corresponding RT distribution if it does not occur in another part of the MPT model. Therefore, in experiments, it is important to ensure that all of the hypothesized cognitive processes actually occur. Otherwise, it is possible that some of the latent RT distributions are empirically not identified (Schmittmann et al., 2010). One solution to this issue is the analysis of the group frequencies aggregated across participants. Summing the individual frequencies of RT-extended MPT models is in general possible because only the relative speed of processing is represented in the data due to the individual log-normal approximation. However, despite the comparable data structure, the assumption of participant homogeneity is questionable in general (Smith & Batchelder, 2008) and might require the use of hierarchical MPT models (e.g., Klauer, 2010; Smith & Batchelder, 2010).

Testing the ordering of latent processes

There are two main types of hypotheses that are of interest with regard to RT-extended MPT models. On the one hand, psychological theory might predict which of the latent RT distributions corresponding to cognitive processes are identical or different. Substantive hypotheses of interest could be, for instance, whether RTs due to target and lure detection follow the same distribution, or whether the speed of detection is affected by response-bias manipulations. These hypotheses about the equality of latent RT distributions i and j can easily be tested by comparing an RT-extended MPT model against a restricted version with equality constraints on the corresponding latency parameters, that is, $L_{ib} = L_{jb}$ for all RT bins $b = 1, \dots, B - 1$. In other words, testing such a hypothesis is similar to the common approach in MPT modeling to test whether an experimental manipulation affects the parameters (Batchelder & Riefer, 1999; Erdfelder et al., 2009).

On the other hand, psychological theory might predict that RTs from one cognitive process are faster than those from another one (Heathcote, Brown, Wagenmakers, & Eidels, 2010). For instance, if guesses occur strictly after unsuccessful detection attempts in the 2HTM (Erdfelder et al., 2011), RTs resulting from detection should be faster than RTs resulting from guessing.⁴ Such a substantive hypothesis about the ordering of cognitive processes translates into the statistical property of stochastic dominance of RT distributions (Luce, 1986; Townsend & Ashby, 1983). The RT distribution of process j stochastically dominates the RT distribution of process i if their cumulative density functions fulfill the inequality

$$F_j(t) \leq F_i(t) \text{ for all } t \in \mathcal{R}, \quad (3)$$

which is illustrated in Fig. 5a. Note that a serial interpretation of the 2HTM directly implies the stochastic dominance of guessing RTs over detection RTs. If the RTs corresponding to the cognitive processes i and j are directly observable, this property can be tested using the empirical cumulative density functions. Note that several nonparametric statistical tests for stochastic dominance use a finite number of RT bins to test Eq. 3 based on the empirical cumulative histograms (see Heathcote et al., 2010, for details).

In RT-extended MPT models, substantive hypotheses concern the ordering of *latent* RT distributions. Moreover, the continuous RTs t are replaced by a finite number of RT bins. However, Eq. 3 can directly be translated to an RT-extended MPT model considering that the term $\sum_{k=1}^b L_{jk}$ gives the probability of responses from process j falling into an RT bin between 1 and b . In other words, it gives the underlying cumulative density function for RTs of process

⁴In terms of serial stage models (Roberts & Sternberg, 1993), we are comparing the random variable of observed detection RTs T_d against the random variable of observed guessing RTs $T_g = T_d + T'$, where T' is a positive random variable resembling pure guessing duration.

j . Hence, the statistical property of stochastic dominance of process j over process i results in the following constraint on the latency parameters of an RT-extended MPT model:

$$\sum_{k=1}^b L_{jk} \leq \sum_{k=1}^b L_{ik} \text{ for all } b = 1, \dots, B-1. \quad (4)$$

This constraint is illustrated in Fig. 5b, where process i is faster than process j .

When using only two RT bins, this relation directly simplifies to the linear order restriction $L_{j1} \leq L_{i1}$ because each latent RT distribution has a single latency parameter only. In MPT models, such linear order constraints can be reparameterized using auxiliary parameters η , for example, $L_{j1} = \eta L_{i1}$ (Klauer & Kellen, 2015; Knapp & Batchelder, 2004). To test this kind of restrictions, model selection techniques that take the diminished flexibility of order-constrained MPT models into account have become available (e.g., Klauer & Kellen, 2015; Vandekerckhove, Matzke, & Wagenmakers, 2015; Wu, Myung, & Batchelder, 2010). Hence, the ordering of latent processes can directly be tested using existing software and methods as shown in the empirical example below.

Even though the use of only two RT bins is attractive in terms of simplicity, it will also limit the sensitivity to detect violations of stochastic dominance. Specifically, the cumulative densities might intersect at the tails of distributions. In such a case, an analysis based on two RT bins might still find support for stochastic dominance even with large sample sizes (Heathcote et al., 2010). Therefore, it might be desirable to use more than two RT bins if the sample size is sufficiently large. However, when using more than two RT bins, a statistical test of Eq. 4 becomes more complex because the sum on both sides of the inequality allows for trade-offs in the parameter values. Importantly, a test of the set of simple linear order constraints $L_{jb} \leq L_{ib}$ for all bins b captures only a sufficient but not a necessary condition for stochastic dominance. In other words, the set of restrictions $L_{jb} \leq L_{ib}$ can be violated even though two latent RT distributions meet the requirement for stochastic dominance. Note that this problem transfers to any binary reparameterization of an extended MPT model with more than two RT bins (see Appendix B). As a consequence, standard software for MPT models that relies on a binary representations (e.g., Moshagen, 2010; Singmann & Kellen, 2013) can in general not be used to test stochastic dominance of latent RT distributions using more than two bins.

As a solution, Eq. 4 can be tested using Bayes factors based on the encompassing prior approach (Klugkist, Laudy, & Hoijtink, 2005), similarly as in Heathcote et al. (2010).

Essentially, this approach quantifies the evidence in favor of a restriction by the ratio of prior and posterior probability mass over a set of order-constraints on the parameters. To test stochastic dominance of latent RT distribution, the RT-extended MPT model is fitted without any order constraints. Next, the posterior samples are used to estimate the prior and posterior probabilities that Eq. 4 holds. The ratio of these two probabilities is the Bayes factor in favor of stochastic dominance (Klugkist et al., 2005). We show how to apply this approach in the second part of our empirical example below.

The RT-extended two-high-threshold model

In this section, we discuss theoretical constraints on the number of latent RT distributions for the 2HTM. Moreover, we demonstrate how the proposed model can be applied to test hypotheses about latent RT distributions using data on recognition memory.

Theoretical constraints on the latent RT distributions of the 2HTM

As discussed above, the 2HTM with six latent RT distributions in Fig. 2 is not identifiable. However, assuming complete-information loss of the latent RT distributions of guessing, the model becomes identifiable. Here, complete-information loss refers to the assumption that guessing, conditional on unsuccessful detection, is independent of item type. In other words, the guessing probability g is identical for lures and targets because no information about the item type is available in the uncertainty state (Kellen & Klauer, 2015). This core property of discrete-state models should also hold for RTs: Given that an item was not recognized as a target or a lure, the RT distributions for guessing *old* and *new* should not differ between targets and lures. Based on a serial-processing interpretation of the 2HTM (Erdfelder et al., 2011), this actually includes two assumptions: Not only the time required for actual guessing but also the preceding time period from test item presentation until any detection attempts are stopped are identically distributed for targets and lures.⁵ Both assumptions are reasonable. They are in line with the idea that participants do not have additional information about the item type that might affect their motivation to detect a test item. In

⁵As noted by a reviewer, this assumption is mandatory for the 2HTM when applied to *yes-no* recognition tasks. However, it might be dispensable in other models (e.g. the one-high threshold model) or when using other memory tasks (e.g. the two-alternative forced choice task).

sum, complete-information loss implies that RT distributions involving guessing are indistinguishable for targets and lures.

The assumption of complete-information loss restricts the six latent RT distributions in Fig. 2 to only four latent RT distributions; those of target and lure detection, and those of guessing *old* and *new*, respectively. This restriction renders the RT-extended 2HTM identifiable given that the core parameters d_o , d_n , and g are identified. Identification of the core parameters can be achieved, for example, by imposing the equality constraint $d_o = d_n$ (e.g., Bayen et al., 1996; Klauer & Wegener, 1998). To check the identifiability of the RT-extended model, we apply the simple strategy introduced above. First, we observe that misses and false alarms both result from single branches. Hence, the corresponding RT distributions for guessing *new* and *old* are identifiable. Second, because RTs for hits either emerge from the identifiable distribution of guessing *old* or from the process of target detection, the latter RT distribution is also identifiable. The same logic holds for correct rejections and lure detection.

In sum, all of the four latent RT distributions are identifiable within a single experimental condition. Therefore, we can estimate four separate latent RT distributions for several experimental manipulations. Here, we want to test whether the assumptions of complete-information loss and fast detection hold across different memory-strength and base-rate conditions, and whether the speed of detection is affected by the base-rate manipulation.

Sensitivity simulations

Before applying the proposed method to actual data, we show that the approach is sufficiently sensitive to decompose observed mixture distributions. First, we estimate the statistical power to detect a discrepancy between the latent RT distributions of detection and guessing using only two RT bins, and second, we show that our distribution-free approach can recover even atypical RT distributions using eight RT bins.

To assess the statistical power, we simulated data for a single experimental condition with true, underlying detection probabilities of $d_o = d_n = .7$ and a symmetric guessing probability of $g = .5$. To generate RTs from the latent RT distributions of detection and guessing, we used two ex-Gaussian distributions (i.e., RTs were sampled from the sum of a normal and an independent exponential random variable). Both of these distributions shared the standard deviation $\sigma = 100$ ms of the normal component and the mean $\nu = 300$ ms of the exponential component. Differences between the two distributions were induced by manipulating the mean of the normal component. Whereas the normal component for the guessing RTs had a constant mean of $\mu_g = 1,000$ ms, we manipulated this mean in the range of $\mu_d = 1,000$ ms, $\dots$, 700 ms in steps of 50 ms for the detection RTs. Note that the resulting mean RT differences correspond to effect sizes in terms of Cohen's d between $d = 0$ and $d = 0.95$.

Fig. 6 **a** The theoretically expected mixture distributions for observed RTs of hits and correct rejections, depending on the (standardized) difference in means of the latent RT distributions of detection and

guessing. **b** Power of the RT-extended 2HTM to detect the discrepancy between the latent RT distributions of detection and guessing

Figure 6a shows the theoretically expected mixture distributions for the observed RTs of hits and correct rejections for the five simulated conditions. Note that these mixture distributions are all unimodal and are thus statistically more difficult to discriminate than bimodal mixtures. For each of the five conditions, we generated 1,000 data sets, categorized RTs into two bins based on the log-normal approximation explained above, and fitted two RT-extended MPT models. The more general model had two latency parameters L_d and L_g for the relative speed of detection and guessing, respectively. The nested MPT model restricted these two parameters to be identical and was compared to the more general model by a likelihood-ratio test using a significance level of $\alpha = .05$.

Figure 6b shows the results of this simulation. When the latent RT distributions for detection and guessing were identical ($d = 0$), that is, the null hypothesis did hold, the test adhered to the nominal α -level of 5 % for all sample sizes. In this simple setting, medium and large effects were detected with sufficient power using realistic sample sizes. For instance, the power was 85.9 % to detect an effect of $d = 0.71$ based on a sample size of 100 learned items. As can be expected for a model that does not rely on distributional assumptions, the statistical power to detect a small difference in mean RTs ($d = 0.24$) was quite low even for 200 learned items.

In principle, the distribution-free approach allows for modeling nonstandard RT distributions, for instance, those

emerging in experimental paradigms with a fixed time-window for responding. In such a scenario, latent RT distributions could potentially be right-skewed (fast detection), left-skewed (truncation by the response deadline), or even uniform (guessing). To test the sensitivity of the proposed method in such a scenario, we therefore sampled RT values in the fixed interval $[0, 1]$ from beta distributions with these different shapes.

In contrast to the power study, we used a slightly more complex setting with different parameters for target and lure detection ($d_o = .65$, $d_n = .4$), two response bias conditions and parameters ($g_A = .3$, $g_B = .7$), and separate beta distributions to generate RTs for target detection (right-skewed), lure detection (left-skewed), and guessing in condition A and B (uniform and bimodal, respectively). Figure 7 shows the latent, data-generating RT distributions (black curves) along with the mean estimates (black histogram) across 500 replications (light gray), each based on 150 responses per item type and condition. The results clearly indicate a good recovery of all four distributions.

Overall, we conclude that our distribution-free approach is sufficiently sensitive to estimate latent RT distributions even if the observed mixture distributions are unimodal or if the latent distributions differ markedly in shape. Note that, as in all simulation studies, these results depend on the specific model and the scenarios under consideration. They do not necessarily generalize to more complex situations. Nevertheless, our simulations show that, in principle, the

Fig. 7 Latent, data-generating beta distributions are shown by solid curves, whereas individual and mean estimates of the RT-extended 2HTM are shown as light gray and black histograms, respectively

categorization of RTs into bins allows for a distribution-free estimation of latent RT distributions.

Methods

As in previous work introducing novel RT models (e.g., Ollman, 1966; Yantis et al., 1991), we use a small sample with many responses per participant to illustrate the application of the proposed method. The data are from four students of the University of Mannheim (3 female, mean age = 24) who responded to 1,200 test items each. The study followed a 2 (proportion of old items: 30 % vs. 70 %) $\times$ 2 (stimulus presentation time: 1.0 vs. 2.2 seconds) within-subjects design. To obtain a sufficient number of responses while also maintaining a short testing duration per session, the study was split into three sessions that took place on different days at similar times. The four conditions were completed block-wise in a randomized order in each of the three sessions. We selected stimuli from a word list with German nouns by Lahl, Gritz, Pietrowsky, & Rosenberg, (2009). From this list, words shorter than three letters and longer than ten letters were removed. The remaining items were ordered by concreteness to select the 1464 most concrete words for the experiment. From this pool, words were randomly assigned to sessions, conditions, and item types without replacement.

The learning list contained 72 words in each of the four conditions including one word in the beginning and the end of the list to avoid primacy and recency effects, respectively. Each word was presented either for one second in the low memory strength condition or for 2.2 seconds in the high memory strength condition with a blank inter-stimulus interval of 200 ms. After the learning phase, participants worked on a brief distractor task (i.e., two minutes for finding differences in pairs of pictures). In the testing phase, 100 words were presented including either 30 % or 70 % learned items. Participants were told to respond either *old* or *new* as accurate and as fast as possible by pressing the keys ‘A’ or ‘K.’ Directly after each response, a blank screen appeared for a random duration drawn from a uniform distribution between 400 ms and 800 ms. This random inter-test interval was added to prevent the participants from responding in a rhythmic manner that might result in statistically dependent RTs (Luce, 1986). After each block of ten responses, participants received feedback about their current performance to induce base-rate-conform response behavior. The experiment was programmed using the open-source software OpenSesame (Mathôt, Schreijf, & Theeuwes, 2012).

Analyses based on two RT bins

We first tested our three hypotheses (complete-information loss, fast detection, and effects of base-rates on detection

RTs) based on only two RT bins using standard software and methods. In the next step, we reexamined our conclusions regarding stochastic dominance of detection RTs over guessing RTs using more RT bins.

Competing models

We compared six RT-extended MPT models to test our hypotheses about the latent RT distributions. All of these models share the same six core parameters of the basic 2HTM. In line with previous applications of the 2HTM (e.g., Bayen et al., 1996; Bröder & Schütz, 2009; Klauer & Wegener, 1998), we restricted the probability of detecting targets and lures to be identical, $d_o = d_n$, separately for the two memory strength conditions to obtain a testable version of the standard 2HTM. In addition to the two detection parameters, we used four parameters for the guessing probabilities, separately for all four experimental conditions. Note that we included separate response bias parameters for the memory strength conditions because this factor was manipulated between blocks, which can result in different response criteria (Stretch & Wixted, 1998). Whereas all of the six RT-extended MPT models shared these six core parameters, the models differed in their assumptions about the latent RT distributions.

The most complex substantive model (‘CI loss’) assumes four separate latent RT distributions for each of the four experimental conditions. This model allows for any effects of experimental manipulations on the relative speed of the underlying processes. The core assumptions of this model are (a) two-component mixtures for RTs of correct responses to targets and lures and (b) complete-information loss in the uncertainty state. Specifically, the latter assumption implies that guesses are equally fast for targets and lures. This holds for both *old* and *new* guesses, respectively.

The second substantive model (‘Fast detection’) is a sub-model of the ‘CI loss’ model. It tests a necessary condition for a serial interpretation of the 2HTM (Dube et al., 2012; Erdfelder et al., 2011). According to the hypothesis that guesses occur strictly after unsuccessful detection attempts, responses due to detection must be faster than responses due to guessing. Specifically, we assumed that this hypothesis of stochastic dominance holds within each response category, that is, within *old*-responses ($L_{do} \geq L_{go}$) and within *new*-responses ($L_{dn} \geq L_{gn}$). Imposing these two constraints in all four experimental conditions results in a total of eight order constraints. Note that adding the corresponding constraints across response categories (i.e., $L_{do} \geq L_{gn}$ and $L_{dn} \geq L_{go}$) would result in the overall constraint

$$\min(L_{do}, L_{dn}) \geq \max(L_{go}, L_{gn}) \quad (5)$$

within each condition. However, this constraint requires the additional assumption that there is no difference in the overall speed of *old* and *new* responses. Since our interest is only in the core assumptions of the RT-extended 2HTM, we tested stochastic dominance only within response categories.

The third substantive model ('Invariant L_{do} ') is nested in the 'Fast detection' model. It tests the hypothesis that the speed of the actual recognition process is not affected by response bias, or in other words, the restriction that target detection is similarly fast across base rate conditions, $L_{do}^{30\%} = L_{do}^{70\%}$. Since the constraint applies in both memory strength conditions, this reduces the number of free parameters by two. The fourth substantive model ('Invariant L_d ') adds the corresponding constraint for lure detection to the model 'Invariant L_{do} ,' thus implying invariance of both L_{do} and L_{dn} against base-rate manipulations.

These four substantive RT-extended MPT models are tested against two reference models that should be rejected if the theoretical assumptions underlying the RT-extended 2HT model hold. The first of these models ('No mix') does *not* assume RT-mixture distributions for correct responses. Instead, it estimates $4 \cdot 4 = 16$ separate RT distributions, one for each response category of all experimental conditions. This model has no additional restrictions in comparison to the standard 2HTM without RTs because the categorization of responses from fast to slow is perfectly fitted within each response category. The second reference model ('Null') tests whether the proposed method is sufficiently sensitive to detect any differences in latent RT distributions at all. For this purpose, it assumes that all observed RTs stem from a single RT distribution by constraining all latency parameters to be equal, $L_j = L_i$ for all i, j .

Model selection results

Before testing any RT-extended MPT model, it is necessary to ensure that the basic MPT model fits the observed responses. Otherwise, model misfit cannot clearly be attributed either to the assumed structure of latent RT distributions or to the MPT model itself. In our case, the standard 2HTM with the restriction of identical detection probabilities for targets and lures ($d_o^{\text{strong}} = d_n^{\text{strong}}$, $d_o^{\text{weak}} = d_n^{\text{weak}}$) fitted the individual data of all participants well (all $G^2(2) \leq 3.6$, $p \geq .16$, with a statistical power of $1 - \beta = 88.3\%$ to detect a small deviation of $w = 0.1$). Figure 8 shows the fitted ROC curves (gray solid lines) and indicates that both experimental manipulations selectively influenced the core parameters as expected.⁶ This visual impression was confirmed by likelihood-ratio tests

showing that the d -parameters differed between memory strength conditions (all $\Delta G^2(1) > 21.1$, $p < .001$) and that the g -parameters differed between the base-rate conditions (all $\Delta G^2(2) > 11.3$, $p < .004$). Note that we did not constrain the guessing parameters to be identical across memory strength conditions because of possibly different response criteria per block (Stretch & Wixted, 1998). Both Fig. 8 and a likelihood-ratio test showed that this was indeed the case for Participant 3 ($G^2(2) = 15.7$, $p < .001$).

In addition to a good model fit, it is important that the core parameters of the MPT model are not at the boundary of the parameter space. If one of the core parameters is close to zero, the corresponding cognitive process is assumed not to occur, and thus, the corresponding latent RT distribution can neither be observed nor estimated if it does not occur in another part of the tree. Hence, if some of the core parameters are at the boundaries, it is possible that some of the latent RT distributions are empirically not identified. This issue did not arise in our data because all core parameter estimates varied between .10 and .81. Note that core parameter estimates and standard errors were stable across the different RT-extended 2HT models below with mean absolute differences of .013 and .001, respectively (maximum absolute differences of .079 and .008, respectively).

To select between the six competing RT-extended 2HT models, we used the Fisher information approximation (FIA; Rissanen, 1996), an information criterion based on the minimum description length principle (Grünwald, 2007; Rissanen, 1978). Compared to standard goodness-of-fit tests and other model selection measures, FIA has a number of advantages. Resembling the Akaike information criterion (AIC; Akaike, 1973) and the Bayesian information criterion (BIC; Schwarz, 1978), FIA minimizes errors in predicting new data by making a trade-off between goodness of fit and model complexity. However, whereas AIC and BIC penalize the complexity of a model only by the number of free parameters, FIA takes additional information into account such as order constraints on parameters and the functional form of the competing models (Klauer & Kellen, 2011). This is important in the present case, because three of the models do not differ in the number of parameters. Whereas the substantive model 'CI loss' has the same number of parameters as the reference model 'No mix', it has a smaller complexity because of the structural assumption of conditional independence of RTs. The model 'Fast detection' is even less complex since it adds eight order constraints that decrease the flexibility even more without affecting the number of free parameters.

Besides accounting for such differences in functional flexibility, FIA has the benefits that it is consistent (i.e., it asymptotically selects the true model if it is in the set

⁶The same pattern of selective influence tests also emerged if d_o and d_n were not constrained to be identical (see [Supplementary Material](#)).

Fig. 8 Fit of the standard 2HTM (solid gray lines) to the observed frequencies (shown with standard errors). ML parameter estimates (and standard errors) are displayed for each participant

of competing models) and that it can efficiently be computed for MPT models (Wu et al., 2010). Despite these advantages, FIA can be biased in small samples and should therefore only be used if the empirical sample size exceeds a lower bound (Heck, Moshagen, & Erdfelder, 2014). In the present case, we can use FIA for model selection on the individual level because the lower bound of $N' = 190$ is clearly below 1,200, the number of responses per participant.

Table 1 shows the ΔFIA values for all participants and models, that is, the difference in FIA values to the preferred model. Moreover, the likelihood-ratio tests for all models are shown to assess their absolute goodness of fit. For all participants, both reference models were rejected. Despite its larger complexity, the model ‘No mix’ did not outperform the substantive models. Concerning the other extreme, the null model which assumes no RT differences at all

Table 1 Model selection for the RT-extended 2HTM models using two RT bins

Model	df	Participant 1			Participant 2			Participant 3			Participant 4		
		G^2	p	ΔFIA	G^2	p	ΔFIA	G^2	p	ΔFIA	G^2	p	ΔFIA
No mix	2	3.6	.16	13.2	0.3	.88	14.5	2.6	.27	13.0	2.0	.37	12.5
CI loss	2	3.6	.16	6.2	0.3	.88	7.4	2.6	.27	5.9	2.0	.37	5.5
Fast detection	2	3.6	.16	0.7	0.3	.88	1.9	2.6	.27	0.4	2.0	.37	0.0
Invariant L_{do}	4	6.7	.15	0.0	0.8	.93	0.0	8.1	.09	0.9	27.5	<.01	10.5
Invariant L_d	6	23.4	<.01	6.5	6.4	.38	0.9	10.0	.13	0.0	37.3	<.01	13.5
Null	17	128.7	<.01	45.0	111.3	<.01	39.2	90.0	<.01	25.8	103.1	<.01	32.3

Note. 1,200 responses per participant, which exceeds the lower-bound $N' = 190$ for the application of FIA (Heck et al., 2014)

performed worst as indicated by the large ΔFIA values. Hence, the categorization of RTs into two bins preserved information about the relative speed of the underlying processes, which can be used to test our substantive hypotheses.

Concerning the substantive model, detection responses were faster than guessing responses for all participants as indicated by the low ΔFIA values of the model ‘Fast detection.’ Note that this model fitted just as well as the two more complex models ‘No mix’ and ‘CI loss’ and is therefore preferred by FIA due to its lower complexity. Regarding the effect of different base rates on the speed of detection, the results were mixed. For Participant 4, this manipulation clearly affected the detection speed as indicated by the large FIA and G^2 values for the ‘Invariant L_{do} ’ and ‘Invariant L_d ’ models. In contrast, different base rates did not affect the speed of target detection for the other three participants. Moreover, the model ‘Invariant L_d ,’ which additionally restricts the speed of lure detection to be identical across base rates, was selected only for Participant 3.⁷

Overall, our results support the hypotheses of complete-information loss and fast detection. The results were ambiguous with respect to effects of the base-rate manipulation on detection speed. For some participants, the speed of target detection was invariant with respect to different base rates, for others (Participant 4, in particular) detection tended to be faster when the correct response was congruent with the base-rate conditions.

Plotting the relative speed of latent processes

To facilitate comparisons of the relative speed of the assumed cognitive processes, the parameter estimates $\hat{L}_j$ can directly be compared in a single graph for several processes and participants. Since L_j is defined as the probability that responses of process j are faster than the RT boundary, larger estimates directly correspond to faster cognitive processes. Therefore, the latency estimates in Fig. 9, based on the model ‘CI loss,’ represent a parsimonious and efficient way of comparing the relative speed of the hypothesized processes across experimental conditions and participants.

In line with the model-selection results, target and lure detection were always estimated to be faster than guessing, with the exception of Participant 2 in the base-rate conditions with 70 % targets. Note that many of the differences are quite substantial, especially when comparing the relative speed of target detection, which was clearly faster

than guessing across all of our four experimental conditions. Moreover, for Participants 1 and 4, the time required to detect an item and give the correct response was not invariant under different base rates, in line with the rejection of the model ‘Invariant L_d ’: These two participants were faster at detecting *old* items when base rates of targets were high (i.e., 70 % targets), and they were faster at rejecting *new* items when base rates of lures were high (i.e., 30 % targets). However, based on the present experiment, we cannot disentangle whether this effect resulted from actual differences in the speed of memory retrieval or from a generally increased preparedness towards the more prevalent response.

Testing stochastic dominance using more than two RT bins

In the following, we compute the Bayes factor based on the encompassing-prior approach (Hojtink, Klugkist, & Boelen, 2008; Klugkist et al., 2005) to quantify the evidence in favor of the hypothesis that responses due to detection are stochastically faster than those due to guessing. The Bayes factor is defined as the odds of the conditional probabilities of the data y given the models $\mathcal{M}_0$ and $\mathcal{M}_1$ (Kass & Raftery, 1995),

$$B_{01} = \frac{p(y | \mathcal{M}_0)}{p(y | \mathcal{M}_1)}, \quad (6)$$

and represents the factor by which the prior odds are multiplied to obtain posterior odds. Moreover, the Bayes factor has a direct interpretation of the evidence in favor of model $\mathcal{M}_0$ compared to model $\mathcal{M}_1$ and can be understood as a weighted average likelihood ratio (Wagenmakers et al., 2015). Note that model selection based on the Bayes factor takes the reduced functional complexity of order-constrained models into account and is asymptotically identical to model selection by FIA under some conditions (Heck, Wagenmakers, & Morey, 2015). In the present case, the model $\mathcal{M}_0$ is the order-constrained RT-extended 2HTM that assumes stochastic dominance within each response category (the model ‘Fast detection’ above). Moreover, the model $\mathcal{M}_1$ is the unconstrained model that only assumes complete-information loss (‘CI loss’). Hence, in the present case, observed Bayes factors substantially larger than one provide evidence in favor of the hypothesis that detection is faster than guessing.

To compute the Bayes factor, we rely on a theoretical result showing that the Bayes factor for order-constrained models is identical to a simple ratio of posterior to prior probability if the prior distributions of the parameters are proportional on the constrained subspace

⁷However, without the constraint $d_o = d_n$, the model ‘Fast detection’ was selected for Participant 3 (see [Supplementary Material](#)).

Fig. 9 Estimated probabilities $\hat{L}_{j1}$ that responses generated by each of four processes are faster than the individual RT boundaries, based on the model ‘CI loss,’ including 95 % confidence intervals

(Klugkist et al., 2005). For instance, the Bayes factor in favor of the simple order constraint $\theta_1 \leq \theta_2$ is identical to

$$B_{01} = \frac{p(\theta_1 \leq \theta_2 \mid \mathbf{y}, \mathcal{M}_1)}{p(\theta_1 \leq \theta_2 \mid \mathcal{M}_1)}, \quad (7)$$

where the numerator and the denominator are the posterior and prior probabilities, respectively, that the constraint holds within the unconstrained model $\mathcal{M}_1$. In our case, the simple order constraint $\theta_1 \leq \theta_2$ is replaced by the order constraint representing stochastic dominance in Eq. 4. In practice, this result facilitates the computation of the Bayes factor

because one only needs to fit the unconstrained model using standard Markov chain Monte Carlo (MCMC) sampling.

In detail, computing the Bayes factor for the present purpose requires the following steps: (1) obtain RT-extended frequencies using the log-normal approximation or any other strategy as explained above, (2) estimate the full model ‘CI loss’ by sampling from the posterior distribution using MCMC (e.g., using the software JAGS, Plummer, 2003), (3) count the MCMC samples that fulfill the constraint of stochastic dominance in Eq. 4, (4) compute the Bayes factor as the ratio of this posterior probability vs. the prior probability that the constraint holds

Table 2 Bayesian factors in favor of stochastic dominance

Participant	4 RT bins			8 RT bins		
	B_{Ldo}	B_{Ldn}	p_{T1}	B_{Ldo}	B_{Ldn}	p_{T1}
1	179.22	50.00	0.29	696.52	2.48	0.12
2	17.45	1.99	0.42	33.34	0.65	0.18
3	201.52	43.85	0.39	1269.57	94.65	0.22
4	174.79	44.81	0.37	288.45	51.88	0.23

The Bayes factors B_{Ldo} and B_{Ldn} quantify the evidence that target detection is faster than guessing *old* and that lure detection is faster than guessing *new*, respectively. The posterior predictive p -value p_{T1} provides an absolute measure of fit for the encompassing model ‘CI Loss’

Fig. 10 Estimated cumulative density functions of the latent RT distributions of target detection and guessing *old* (including 80 % credibility intervals) for Participant 3

(Klugkist et al., 2005). In the present case, we estimated the prior probability that the constraint holds based on parameters directly sampled from the prior distribution. Note that we used uniform priors on the core parameters of the 2HTM and symmetric, uninformative Dirichlet priors on each set of the multinomial latency parameters, that is, $(L_{j1}, \dots, L_{jB}) \sim \text{Dir}(\alpha, \dots, \alpha)$ with parameters $\alpha = 1/B$, similarly as in Heathcote et al. (2010). Note that the core parameter estimates for d and g differed by less than .02 compared to fitting the RT-extended 2HTM using maximum likelihood.

For testing stochastic dominance, we differentiated between the hypotheses that target detection is faster than guessing *old* and that lure detection is faster than guessing *new*. Table 2 shows the resulting Bayes factors using four and eight RT bins based on 50 million MCMC samples. The table also includes posterior predictive p -values p_{T1} that represent an absolute measure of fit similar to p -values associated with the χ^2 test (Klauer, 2010). The large p_{T1} -values imply that the model did not show substantial misfit for any participant and or number of RT bins. More interestingly, the large Bayes factors B_{Ldo} in Table 2 represent substantial evidence that target detection was faster than guessing *old*. In contrast, the corresponding hypothesis that lure detection was faster than guessing *new* was only supported for

Participants 3 and 4. For the other two participants, the Bayes factors B_{Ldn} was close to one and does therefore neither provide evidence for nor against stochastic dominance.

In addition to computing the Bayes factor, the Bayesian approach allows to directly compute point estimates and posterior credibility intervals for the cumulative density functions of the latent RT distributions sketched in Fig. 5b. To illustrate the substantive meaning of large Bayes factors in favor of stochastic dominance, Fig. 10 shows the mean of the estimated cumulative densities for Participant 3 including 80 % credibility intervals. To facilitate the comparison, the cumulative densities are shown separately for *old* and *new* responses (rows) and for the four experimental conditions (columns). Note that the latency estimates are based on the full model ‘CI loss’ and are therefore not constrained to fulfill stochastic dominance necessarily. However, across all experimental conditions, the cumulative densities of RTs due to detection (dark gray area) are clearly above the cumulative densities of RTs due to guessing (light gray). Moreover, stochastic dominance is more pronounced for target detection and guessing *old* (first row) than for lure detection and guessing *new* (second row). Note that this larger discrepancy for target detection contributes to the larger Bayes factors B_{Ldo} compared to B_{Ldn} .

Conclusion

In sum, the empirical example showed (1) how to test hypotheses about the relative speed of cognitive processes using two RT bins and standard MPT software and (2) how to test stronger hypotheses about the ordering of latent processes in a Bayesian framework using more than two RT bins. With respect to the 2HTM, we found support for the hypothesis that RTs of correct responses emerge as two-component mixtures of underlying detection and guessing processes. Moreover, both modeling approaches supported the hypothesis of slow guesses that occur strictly after unsuccessful detection attempts (Erdfelder et al., 2011). However, the latent RT distributions of detecting targets and lures differed across base-rate conditions for two participants, whereas those of the other two participants were not affected by this manipulation. Hence, further studies are necessary to assess the effects of base-rate manipulations on the speed of detecting targets and lures in recognition memory.

Discussion

We proposed a new method to measure the relative speed of cognitive processes by including information about RTs into MPT models. Basically, the original response categories are split into more fine-grained subcategories from fast to slow. To obtain inter-individually comparable categories, the individual RT boundaries for the categorization into bins are based on a log-normal approximation of the distribution of RTs across all response categories. Using these more informative frequencies that capture both response-type and response-speed information, an RT-extended MPT model accounts for the latent RT distributions of the underlying processes using the latency parameters L_{jb} defined as the probability that a response from the j -th processing path falls into the b -th RT bin. Importantly, this approach does not pose any a-priori restrictions on the unknown shape of the latent RT distributions. Our approach allows for testing whether two or more of the latent RT distributions are identical and whether some latent processes are faster than others. Such hypotheses concerning the ordering of latent processes can be tested by simple order constraints of the form $L_i < L_j$ in the standard maximum-likelihood MPT framework when using two RT bins and within a Bayesian framework when using more RT bins.

Substantively, our empirical example supported the hypothesis of complete-information loss, that is, responses due to guessing are similarly fast for targets and lures.

Moreover, we found strong evidence that responses due to target detection are faster than those due to guessing *old*, but weaker evidence that the same relation holds for the speed of lure detection and guessing *new*. Moreover, RTs due to target detection were less affected by base rates than those due to lure detection. Both of these results indicate an important distinction. Whereas target detection seems to occur relatively fast and largely independent of response bias, lure detection seems to be a slower and response-bias dependent process.

Advantages of RT-extended MPT models

The proposed framework provides simple and robust methods to gain new insights. It allows for estimating and comparing the relative speed of cognitive processes based only on the core assumption of MPT models, that is, a finite number of latent processes can account for observed responses. Most importantly, we showed that using discrete RT bins preserves important information in the data that can be used to test substantive hypotheses. Both the power simulation and the empirical example showed that the approach is sufficiently sensitive to detect differences in latent RT distributions. Importantly, this indicates that categorizing RTs into bins does preserve structural information in the data.

The proposed approach also has several practical benefits. Its application is straightforward and based on a principled strategy to categorize RTs, in which the number of RT bins can be adjusted to account for different sample sizes. In practice, the choice of the number of RT bins will typically depend on the substantive question. If the interest is only in a coarse measure of the relative speed of cognitive processes, two RT bins might suffice. However, if sufficient sample sizes are available, using more than two RT bins might be beneficial to obtain more powerful tests that are more sensitive to violations of stochastic dominance.

An additional advantage of the proposed distribution-free framework is that modeling histograms of RTs reduces the sensitivity of the method towards outliers. Moreover, RT-extended MPT models can be fitted using existing software such as multiTree (Moshagen, 2010) or MPTinR (Singmann & Kellen, 2013). Alternatively, inter-individual differences can directly be modeled by adopting one of the Bayesian hierarchical extensions for MPT models proposed in recent years (Klauer, 2010; Matzke, Dolan, Batchelder, & Wagenmakers, 2015; Smith & Batchelder, 2010). Especially in cases with only a small to medium number of responses per participant, these methods are preferable to fitting the model to individual data separately. Last but not least, MPT

models can easily be reparameterized to test order constraints (Klauer et al., 2015; Knapp & Batchelder, 2004), thereby facilitating tests of stochastic dominance.

Parametric modeling of latent RT distributions

Our approach aims at providing a general method to include RTs in any kind of MPT model in a distribution-free manner. However, for some substantive questions, more specific models that predict exact parametric shapes for RTs might be preferable. For instance, Donkin, Nosofsky, Gold, and Shiffrin (2013) extended the discrete-slots model for visual working memory, which closely resembles the 2HTM of recognition memory, to RTs by assuming a mixture of two evidence-accumulation processes. The discrete-slots model assumes that in each trial of a change-detection task, a probe either occupies one of the discrete memory slots or it does not. Depending on this true, latent state, RTs either emerge from a memory-based evidence-accumulation process or from a guessing-based accumulation process. Specifically, the time course of both accumulation processes is modeled by separate linear ballistic accumulators, which assume a race between two independent linear processes reaching a boundary (see Brown & Heathcote, 2008 for details).⁸ Note that the resulting predicted RT distributions resemble specific instantiations of the latent RT distributions sketched in Fig. 2.

The model by Donkin et al. (2013) makes precise predictions how RT distributions change across different experimental conditions. For instance, the model predicts that the RT distribution of false alarms (i.e., *change*-responses in no-change trials) is invariant with respect to set-size manipulations, that is, how many stimuli are presented. Moreover, the model establishes a functional relationship between the guessing parameter g and the corresponding RT distribution for guessing responses. Given that the response threshold for *change*-guessing is much lower than that for *no-change*-guessing, the linear ballistic accumulator predicts both more and faster *change*-responses. Similarly, the model constrains the structure of RTs across different levels of set size and base rates. Donkin et al. (2013) showed that the relevant patterns in the data followed these predictions of the discrete-slots model. Moreover, the model performed better than competitor models assuming continuous memory strength.

Overall, this example illustrates the advantages of process models that directly transfer psychological theory into precise predictions for responses and RTs. Given

such hypotheses about the mechanisms how RTs emerge from the underlying processes, these hypotheses should of course be included in the corresponding statistical model in the form of exact parametric constraints as in Donkin et al. (2013). If the data are in line with such a precise model for a broad range of context conditions, this provides strong support for the theory and clearly more evidence than fitting a less precise, distribution-free model. Compared to such computational models, our focus on measurement models is more general and useful for other types of applications. For instance, if the substantive questions concern only the relative speed or the ordering of latent processes irrespective of the precise nature of the underlying processes, the proposed distribution-free approach allows inclusion of RTs in any MPT model without the necessity to make auxiliary parametric assumptions.

In sum, this shows that both approaches — parametric computational models and measurement models such as RT-extended MPT models — have their advantages and disadvantages, many of which are typical for parametric and nonparametric models in general. Most importantly, the substantive conclusions concerning the underlying cognitive processes should eventually converge irrespective of the approach used.

RT extensions of more complex MPT models

The 2HTM used as an example in the present paper is one of the most simple MPT models. Of course, more complex MPT models might directly benefit from including RTs as well. For instance, in the field of judgment and decision making, the inclusion of RTs in MPT models might be useful in testing between different theoretical accounts. For instance, Hilbig et al. (2010) proposed and validated the *r*-model, which disentangles use of the recognition heuristic from decisions based on integration of further knowledge. According to the heuristic account, a recognized option is chosen in a noncompensatory manner, that is, further information does not influence the decision at all (Goldstein & Gigerenzer, 1999). As a consequence, decisions based on such a process should be faster compared to responses based on integration of further knowledge. In contrast, global information-integration models assume that a single, underlying process integrates all of the available information in order to make a decision (Glöckner & Betsch, 2008). If further congruent knowledge is used in addition to the actual recognition cue, decisions are predicted to be faster compared to decisions solely based on the recognition cue (Glöckner & Bröder, 2011). This prediction directly contradicts the hypothesis that choice of the recognized option is always faster than information integration. Obviously, an RT-extension of the *r*-model

⁸Note that this model allows for incorrect responses in the detection state (Donkin et al., 2013, p. 895) and thus differs from the model shown in Fig. 2. However, the corresponding predicted probability of incorrect responses in the detection state appeared to be very small due to a large drift rate towards the correct response.

could help in discriminating between these two theoretical accounts.

Social psychology is another field in which MPT models are often used, for instance, to measure implicit attitudes (e.g., Conrey et al., 2005; Nadarevic & Erdfelder, 2011) or effects of processing fluency (e.g., Fazio et al., 2015; Unkelbach & Stahl, 2009). Concerning measures of implicit attitudes, several authors proposed to rely on MPT models to analyze response frequencies instead of analyzing mean RTs in the implicit association test (IAT) and the go/no-go association task (e.g., Conrey et al., 2005; Meissner & Rothermund, 2013; Nadarevic & Erdfelder, 2011). These MPT models disentangle cognitive processes such as activation of associations and stimulus discrimination (Conrey et al., 2005) or recoding of response categories (Meissner & Rothermund, 2013). However, to do so, these models sacrifice information about the speed of processes and only rely on response frequencies. Obviously, the inclusion of RTs would allow for assessing the selective speed of and testing the temporal order of the hypothesized processes. For instance, one could test whether responses due to implicit associations are actually faster than those due to stimulus discrimination as is often assumed.

Conclusion

We proposed a novel and general method to estimate latent RT distribution of the hypothesized processes in MPT models in a distribution-free way. The approach enables researchers to directly test hypotheses about the relative speed of these processes. Thereby, it is possible to better understand the nature of the cognitive processes accounted for by MPT models. In sum, the proposed method follows the spirit of MPT modeling in general (Batchelder & Riefer, 1999): It provides a simple way for psychologists to test substantive hypotheses about the relative speed of cognitive processes without the need to rely on many auxiliary assumptions.

Acknowledgments We thank Chris Donkin, David Kellen, and Christoph Klauer for valuable comments and suggestions.

This work was supported by the University of Mannheim's Graduate School of Economic and Social Sciences funded by the German Research Foundation.

Appendix A: Identifiability of latent RT distributions

In the following, we assume that an MPT model is given with T trees, K_t categories in tree t , and I_{tk} branches that lead to a category C_{tk} . The model is parameterized by the

core parameter vector $\theta \in (0, 1)^S$ to model the expected branch probabilities $p_{tki}(\theta)$ and the resulting category probabilities $p_{tk}(\theta) = \sum_{i=1}^{I_{tk}} p_{tki}(\theta)$. For an exact definition of the branch probabilities and other details of the MPT model class, see Hu and Batchelder (1994).

To model latent RT distributions, each branch is further split into B bins with branch probabilities $q_{tkib}(\theta, L) = p_{tki}(\theta) L_{tkib}$, where $L_{tkib} \in (0, 1)$ are the latency parameters of interest (collected in the vector L). Note that for each branch one of these latency parameters is fixed since $\sum_{b=1}^B L_{tkib} = 1$. For notational convenience, we will use the abbreviation $L_{tkiB} = 1 - \sum_{b=1}^{B-1} L_{tkib}$ and thus list B latency parameters for each branch, even though only $B - 1$ of these are actually defined as free parameters. The expected category probabilities of the extended MPT model are then given by summing across the original I_{tk} branches of category C_{tk} separately for each RT bin, i.e., $q_{tkb}(\theta, L) = \sum_{i=1}^{I_{tk}} q_{tkib}(\theta, L)$.

Moreover, it is assumed that some of the latent RT distributions are restricted to be equal based on theoretical arguments. In Observations 3 and 4, we therefore refer to subsets of J RT distributions to facilitate notation. The j -th of these RT distribution is then modeled by the vector of latency parameters $L_j = (L_{j1}, \dots, L_{jB})'$, which replaces the corresponding latency parameters $(L_{tki1}, \dots, L_{tkiB})'$ in the appropriate branches.

Observations 1 to 4 assume that the basic MPT model is globally identifiable and that the core parameters are in the interior of the parameter space, $\theta \in (0, 1)^S$.

Observation 1 (Core Parameters) If all of the original branches and categories of the basic identifiable MPT model are split into B bins to model RT distributions, the core parameters θ of this extended MPT model are also globally identifiable. This holds independently of the number or exact parameterization of the latent RT distributions.

Proof We have to show that $q_{tkb}(\theta, L) = q_{tkb}(\theta', L')$ for all t, k , and b implies $\theta = \theta'$. From the definition of $q_{tkb}(\theta, L)$, it follows that

$$\sum_{i=1}^{I_{tk}} p_{tki}(\theta) L_{tkib} = \sum_{i=1}^{I_{tk}} p_{tki}(\theta') L'_{tkib}. \quad (8)$$

Summarizing across all RT bins from $b = 1, \dots, B$ within each category C_{tk} yields

$$\sum_{i=1}^{I_{tk}} \left(p_{tki}(\theta) \sum_{b=1}^B L_{tkib} \right) = \sum_{i=1}^{I_{tk}} \left(p_{tki}(\theta') \sum_{b=1}^B L'_{tkib} \right). \quad (9)$$

Since $\sum_{b=1}^B L_{tkib} = 1$ and $p_{tk}(\theta) = \sum_{i=1}^{I_{tk}} p_{tki}(\theta)$, it follows that $p_{tk}(\theta) = p_{tk}(\theta')$. The identifiability of the original model directly implies $\theta = \theta'$. $\square$

Observation 2 (Single-Branch Identifiability) If a category C_{sm} is only reached by a single branch in a globally identifiable MPT model (i.e., $I_{sm} = 1$), the corresponding latency parameters $\mathbf{L}_{sm1} = (L_{sm11}, \dots, L_{sm1B})'$ of this branch in the RT-extended MPT model are globally identifiable.

Proof We have to show that $q_{tkb}(\theta, \mathbf{L}) = q_{tkb}(\theta', \mathbf{L}')$ for all t, k , and b implies $\mathbf{L}_{sm1} = \mathbf{L}'_{sm1}$. Because only one branch leads to category C_{sm} , the expected probability of the b -th RT bin of this category is simply $q_{smb}(\theta, \mathbf{L}) = p_{sm}(\theta)L_{sm1b}$. Hence, one obtains

$$p_{sm}(\theta)L_{sm1} = p_{sm}(\theta')L'_{sm1}. \quad (10)$$

From Observation 1 and $\theta \in (0, 1)^S$, it follows that $p_{sm}(\theta) = p_{sm}(\theta') \neq 0$ and hence $\mathbf{L}_{sm1} = \mathbf{L}'_{sm1}$. $\square$

Observation 3 (Recursive Identifiability) Assume that the RT bins of category C_{sm} with $I_{sm} > 1$ branches are modeled by $J \leq I_{sm}$ latent RT distributions. If $J - 1$ of these J latent RT distributions are identifiable from other parts of the extended MPT Model (i.e., from Observation 2 and repeated application of Observation 3), the latency parameters $\mathbf{L}_{smJ} = (L_{smJ1}, \dots, L_{smJB})'$ of the remaining J -th RT distribution are globally identifiable.

Proof We have to show that $q_{tkb}(\theta, \mathbf{L}) = q_{tkb}(\theta', \mathbf{L}')$ for all t, k , and b implies $\mathbf{L}_{smJ} = \mathbf{L}'_{smJ}$. The category probabilities for the RT bins of category C_{sm} can be split into two parts, the one containing the first $J - 1$ identifiable RT distributions and the remaining one using only the J -th RT distribution:

$$q_{smb}(\theta, \mathbf{L}) = \sum_{i \in A} p_{smi}(\theta)L_{smib} + \sum_{i \in B} p_{smi}(\theta)L_{smib}, \quad (11)$$

with the disjunct indexing sets

$$A := \{i : L_{tki} = L_{tkj} \text{ for any } j = 1, \dots, J - 1\} \quad (12)$$

$$B := \{i : L_{tki} = L_{tkJ}\}. \quad (13)$$

Hence, since $\theta = \theta'$ from Observation 1 and $\mathbf{L}_{smi} = \mathbf{L}'_{smi} \forall i \in A$, the equation $q_{smb}(\theta, \mathbf{L}) = q_{smb}(\theta', \mathbf{L}')$ can be written as

$$\begin{aligned} & \sum_{i \in A} p_{smi}(\theta)L_{smib} + \sum_{i \in B} p_{smi}(\theta)L_{smib} \\ &= \sum_{i \in A} p_{smi}(\theta)L_{smib} + \sum_{i \in B} p_{smi}(\theta)L'_{smib}. \end{aligned} \quad (14)$$

Since $\mathbf{L}_{smi} = \mathbf{L}_{smJ}$ for all $i \in B$, one obtains

$$L_{smJb} \sum_{i \in B} p_{smi}(\theta) = L'_{smJb} \sum_{i \in B} p_{smi}(\theta). \quad (15)$$

Because of $\theta \in (0, 1)^S$, it follows that $p_{smi}(\theta) \neq 0$ for all i and hence that $\mathbf{L}_{smJ} = \mathbf{L}'_{smJ}$. $\square$

Assume that the recursive application of Observations 2 and 3 did not ensure the identifiability of at least one of the hypothesized latent RT distributions. In such a case, the identifiability of the extended MPT model can be shown as follows. Note that this is the most general observation and that it implies Observations 2 and 3.

Observation 4 (General Identifiability) The extended MPT model is globally identifiable if and only if the rank of the following $(\sum_{t=1}^T K_t) \times J$ matrix exceeds J , the total number of hypothesized latent RT distributions parameterized by $L_{1b}, \dots, L_{Jb}$:

$$P(\theta) = [p_{tkj}(\theta)]_{\substack{tk=11, \dots, 1K_1, 21, \dots, TK_T \\ j=1, \dots, J}}. \quad (16)$$

The rows of this matrix contain all observable categories across all trees. The entries of $P(\theta)$ are defined as the branch probabilities of the original MPT model if a category C_{tk} includes a branch associated with the latent RT distribution j and are zero otherwise,

$$p_{tkj}(\theta) = \begin{cases} p_{tki}(\theta), & \text{if } L_{tkib} = L_{jb} \text{ for all } b = 1, \dots, B \\ 0, & \text{otherwise.} \end{cases} \quad (17)$$

Proof First, we have to show that $q_{tkb}(\theta, \mathbf{L}) = q_{tkb}(\theta', \mathbf{L}')$ for all t, k , and b implies $\mathbf{L} = \mathbf{L}'$. Using the definition of the matrix $P(\theta)$ and $q_{tkb}(\theta, \mathbf{L}) = \sum_{i=1}^{I_{tk}} p_{tki}(\theta)L_{tkib}$, the system of equations $q_{tkb}(\theta, \mathbf{L}) = q_{tkb}(\theta', \mathbf{L}') \forall t, k, b$ can be written in matrix notation as

$$P(\theta)\mathbf{L}_b = P(\theta')\mathbf{L}'_b \quad \forall b = 1, \dots, B, \quad (18)$$

where $\mathbf{L}_b = (L_{1b}, \dots, L_{Jb})'$ (note that this is a vector for the b -th bin across all distributions in contrast to the previous notation). Observation 1 implies $\theta = \theta'$ and hence $P(\theta) = P(\theta')$. If the rank of $P(\theta)$ is equal or larger than J , the null space of $P(\theta)$ is trivial (i.e., only contains the null vector $\mathbf{0}$). Hence,

$$P(\theta)(\mathbf{L}_b - \mathbf{L}'_b) = \mathbf{0} \quad \forall b = 1, \dots, B \quad (19)$$

implies $\mathbf{L}_b - \mathbf{L}'_b = \mathbf{0}$, or equivalently $\mathbf{L}_b = \mathbf{L}'_b$ for all bins $b = 1, \dots, B$.

Second, we have to show that the models' identifiability implies that $P(\theta)$ has full rank. We proof this by

contradiction and assume that $P(\theta)$ does *not* have full rank. As before, the definition of $P(\theta)$ and the identifiability of θ implies

$$P(\theta)(L_b - L'_b) = \mathbf{0} \quad \forall b = 1, \dots, B. \quad (20)$$

However, because the rank of $P(\theta)$ is smaller than the dimensionality of the vector L , its null space is not trivial. Hence, a solution $\mathbf{a} \neq \mathbf{0} \in \mathbb{R}^J$ exists with $L_b - L'_b = \mathbf{a}$. This implies $L_b \neq L'_b$ and contradicts the models' identifiability. $\square$

Note that computing the rank of the matrix $P(\theta)$ in Observation 4 is less complex than examining the rank of the whole system of model equations of the extended MPT model. Moreover, computer algebra software is available facilitating the search for a set of restrictions on the latent RT distributions that ensures identifiability (cf. Schmittmann et al., 2010).

As an example for the application of Observation 4, the 2HTM with the assumption of conditional independence of latent RTs for guessing results in the matrix

$$P(\theta) = \begin{pmatrix} L_{do} & L_{g-old} & L_{g-new} & L_{dn} \\ d_o & (1-d_o)g & 0 & 0 \\ 0 & 0 & (1-d_o)(1-g) & 0 \\ 0 & (1-d_n)g & 0 & 0 \\ 0 & 0 & (1-d_n)(1-g) & d_n \end{pmatrix} \begin{matrix} Hit \\ Miss \\ FA \\ CR \end{matrix},$$

which has full rank if $d_o, d_n, g \in (0, 1)$ and thus ensures the identifiability of the model (given that the core parameters d_o, d_n , and g are identified). In contrast, the following (non-theoretical) assignment of four latent RT distributions is not identifiable if $d_o = d_n$ and $g = .5$, because the first and last row are then linearly dependent,

$$P(\theta) = \begin{pmatrix} L_d & L_{g-correct} & L_{Miss} & L_{FA} \\ d_o & (1-d_o)g & 0 & 0 \\ 0 & 0 & (1-d_o)(1-g) & 0 \\ 0 & 0 & 0 & (1-d_n)g \\ d_n & (1-d_n)(1-g) & 0 & 0 \end{pmatrix} \begin{matrix} Hit \\ Miss \\ FA \\ CR \end{matrix}.$$

Appendix B: Testing stochastic dominance of latent RT distributions

In the following, we illustrate why stochastic dominance cannot be tested directly by linear inequality constraints in a binary MPT representation if more than two RT bins are used. Consider the binary reparameterization for four

Fig. 11 A possible reparameterization of an RT-extended MPT model with 4 RT bins into a binary MPT model

RT bins in Fig. 11. In this model, the new parameters L'_{ib} of the binary MPT represent the conditional probabilities that a response from process i falls into one of the bins $\{1, \dots, b\}$, given that it is not in a bin $\{b+2, \dots, B\}$. Using this reparameterization, the L'_{jb} are mapped to the original parameters L_{ib} by

$$\sum_{k=1}^b L_{jk} = \prod_{k=b}^{B-1} L'_{ik} \quad (21)$$

Hence, testing stochastic dominance (Eq. 4) is equivalent to testing

$$\prod_{k=b}^{B-1} L'_{jk} \leq \prod_{k=b}^{B-1} L'_{ik} \quad \forall b = 1, \dots, B-1 \quad (22)$$

in the binary MPT model. Note that this is a set of non-linear order constraints, which cannot directly implemented with the standard MPT methods for linear order constraints (Klauer et al., 2015; Knapp & Batchelder, 2004). Importantly, testing the order of the binary latency parameters directly (i.e., testing $L'_{jb} \leq L'_{ib}$ for all b) will result in an overly conservative test. More precisely, it tests only a sufficient but not a necessary condition for Eq. 22. In other words, the property of stochastic dominance in Eq. 22 can perfectly hold, whereas the order restrictions on the individual binary parameters are violated. For instance, with three RT bins, if $L'_{j2} = .2 \leq L'_{i2} = .8$ (Eq. 22 for $b = 2$), stochastic dominance still holds if $.2L'_{j1} \leq .8L'_{i1}$ (Eq. 22 for $b = 1$), for example, for $L'_{j1} = .8 > L'_{i1} = .4$.

The binary model in Fig. 11 can be reparameterized into statistically equivalent binary MPT models that differ in the definition of the L' parameters. However, it can be expected that the set of nonlinear order restrictions in Eq. 22 will also constrain the parameters of the new binary MPT model in a nonlinear way.

References

- Agresti, A. (2013). *Categorical data analysis*. Hoboken: Wiley.
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Second International Symposium on Information Theory*, pp. 267–281.
- Ashby, F.G., Prinzmetal, W., Ivry, R., & Maddox, W.T. (1996). A formal theory of feature binding in object perception. *Psychological Review*, 103, 165–192. doi:[10.1037/0033-295X.103.1.165](https://doi.org/10.1037/0033-295X.103.1.165)
- Bamber, D., & van Santen, J.P.H. (2000). How to assess a model's testability and identifiability. *Journal of Mathematical Psychology*, 44, 20–40. doi:[10.1006/jmps.1999.1275](https://doi.org/10.1006/jmps.1999.1275)
- Batchelder, W.H., & Riefer, D.M. (1986). The statistical analysis of a model for storage and retrieval processes in human memory. *British Journal of Mathematical and Statistical Psychology*, 39, 129–149. doi:[10.1111/j.2044-8317.1986.tb00852.x](https://doi.org/10.1111/j.2044-8317.1986.tb00852.x)
- 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](https://doi.org/10.1037/0033-295X.97.4.548)
- Batchelder, W.H., & Riefer, D.M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, 6, 57–86. doi:[10.3758/BF03210812](https://doi.org/10.3758/BF03210812)
- Bayen, U.J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 197–215. doi:[10.1037/0278-7393.22.1.197](https://doi.org/10.1037/0278-7393.22.1.197)
- 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](https://doi.org/10.1037/a0015279)
- Brown, S., & 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](https://doi.org/10.1016/j.cogpsych.2007.12.002)
- Brown, V. (1998). Recent progress in mathematical psychology: Psychophysics, knowledge, representation, cognition, and measurement. C. E. Dowling, F. S. Roberts, & P. Theuns (Eds.), (pp. 253–284). NJ: Mahwah.
- Buchner, A., Erdfelder, E., Steffens, M.C., & Martensen, H. (1997). The nature of memory processes underlying recognition judgments in the process dissociation procedure. *Memory & Cognition*, 25, 508–517. doi:[10.3758/BF03201126](https://doi.org/10.3758/BF03201126)
- Conrey, F.R., Sherman, J.W., Gawronski, B., Hugenberg, K., & Groom, C.J. (2005). Separating multiple processes in implicit social cognition: The quad model of implicit task performance. *Journal of Personality and Social Psychology*, 89, 469–487. doi:[10.1037/0022-3514.89.4.469](https://doi.org/10.1037/0022-3514.89.4.469)
- Donkin, C., Nosofsky, R.M., Gold, J.M., & Shiffrin, R.M. (2013). Discrete-slots models of visual working-memory response times. *Psychological Review*, 120, 873–902. doi:[10.1037/a0034247](https://doi.org/10.1037/a0034247)
- 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](https://doi.org/10.1016/j.jml.2012.06.002)
- Erdfelder, E., Auer, T.-S., Hilbig, B.E., Assfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Zeitschrift für Psychologie / Journal of Psychology*, 217, 108–124. doi:[10.1027/0044-3409.217.3.108](https://doi.org/10.1027/0044-3409.217.3.108)
- Erdfelder, E., Castela, M., Michalkiewicz, M., & Heck, D.W. (2015). The advantages of model fitting compared to model simulation in research on preference construction. *Frontiers in Psychology*, 6, 140. doi:[10.3389/fpsyg.2015.00140](https://doi.org/10.3389/fpsyg.2015.00140)
- Erdfelder, E., Küpper-Tetzl, C.E., & Mattern, S.D. (2011). Threshold models of recognition and the recognition heuristic. *Judgment and Decision Making*, 6, 7–22.
- Fazio, L.K., Brashier, N.M., Payne, B.K., & Marsh, E.J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, 144, 993–1002. doi:[10.1037/xge0000098](https://doi.org/10.1037/xge0000098)
- Glöckner, A., & Betsch, T. (2008). Multiple-reason decision making based on automatic processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1055–1075. doi:[10.1037/0278-7393.34.5.1055](https://doi.org/10.1037/0278-7393.34.5.1055)
- Glöckner, A., & Bröder, A. (2011). Processing of recognition information and additional cues: A model-based analysis of choice, confidence, and response time. *Judgment and Decision Making*, 6, 23–42.
- Goldstein, D.G., & Gigerenzer, G. (1999). The recognition heuristic: How ignorance makes us smart. G. Gigerenzer, P. M. Todd, & ABC Research Group (Eds.) *Simple heuristics that make us smart*, (pp. 37–58). New York: Oxford University Press.
- Grünwald, P. (2007). *The minimum description length principle*. Cambridge: MIT Press.
- Heathcote, A., Brown, S., Wagenmakers, E.J., & Eidels, A. (2010). Distribution-free tests of stochastic dominance for small samples. *Journal of Mathematical Psychology*, 54, 454–463. doi:[10.1016/j.jmp.2010.06.005](https://doi.org/10.1016/j.jmp.2010.06.005)
- Heck, D.W., Moshagen, M., & Erdfelder, E. (2014). Model selection by minimum description length: Lower-bound sample sizes for the Fisher information approximation. *Journal of Mathematical Psychology*, 60, 29–34. doi:[10.1016/j.jmp.2014.06.002](https://doi.org/10.1016/j.jmp.2014.06.002)
- Heck, D.W., Wagenmakers, E.-J., & Morey, R.D. (2015). Testing order constraints: Qualitative differences between Bayes factors and normalized maximum likelihood. *Statistics & Probability Letters*, 105, 157–162. doi:[10.1016/j.spl.2015.06.014](https://doi.org/10.1016/j.spl.2015.06.014)
- Hilbig, B.E., Erdfelder, E., & Pohl, R.F. (2010). One-reason decision making unveiled: A measurement model of the recognition heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 123–134. doi:[10.1037/a0017518](https://doi.org/10.1037/a0017518)
- Hilbig, B.E., & Pohl, R.F. (2009). Ignorance-versus evidence-based decision making: A decision time analysis of the recognition heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 1296–1305. doi:[10.1037/a0016565](https://doi.org/10.1037/a0016565)
- Hojitink, H., Klugkist, I., & Boelen, P.A. (Eds.) (2008). *Bayesian evaluation of informative hypotheses*. New York: Springer.
- Hu, X. (2001). Extending general processing tree models to analyze reaction time experiments. *Journal of Mathematical Psychology*, 45, 603–634. doi:[10.1006/jmps.2000.1340](https://doi.org/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](https://doi.org/10.1007/bf02294263)
- 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](https://doi.org/10.1016/0749-596X(91)90025-F)
- Kass, R.E., & Raftery, A.E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90, 773–795. doi:[10.1080/01621459.1995.10476572](https://doi.org/10.1080/01621459.1995.10476572)
- 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.
- 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](https://doi.org/10.3758/s13423-013-0407-2)

- 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](https://doi.org/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](https://doi.org/10.1007/s11336-009-9141-0)
- Klauer, K.C., Singmann, H., & Kellen, D. (2015). Parametric order constraints in multinomial processing tree models: An extension of Knapp and Batchelder. *Journal of Mathematical Psychology*, 64, 1–7. doi:[10.1016/j.jmp.2014.11.001](https://doi.org/10.1016/j.jmp.2014.11.001)
- Klauer, K.C., Stahl, C., & Erdfelder, E. (2007). The abstract selection task: New data and an almost comprehensive model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 680–703. doi:[10.1037/0278-7393.33.4.680](https://doi.org/10.1037/0278-7393.33.4.680)
- Klauer, K.C., & Kellen, D. (2011). The flexibility of models of recognition memory: An analysis by the minimum-description length principle. *Journal of Mathematical Psychology*, 55, 430–450. doi:[10.1016/j.jmp.2011.09.002](https://doi.org/10.1016/j.jmp.2011.09.002)
- 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](https://doi.org/10.1016/j.jmp.2015.05.002)
- Klauer, K.C., & Wegener, I. (1998). Unraveling social categorization in the ‘Who said what?’ paradigm. *Journal of Personality and Social Psychology*, 75, 1155–1178. doi:[10.1037/0022-3514.75.5.1155](https://doi.org/10.1037/0022-3514.75.5.1155)
- Klugkist, I., Laudy, O., & Hoijtink, H. (2005). Inequality constrained analysis of variance: A Bayesian approach. *Psychological Methods*, 10, 477.
- Knapp, B.R., & Batchelder, W.H. (2004). Representing parametric order constraints in multi-trial applications of multinomial processing tree models. *Journal of Mathematical Psychology*, 48, 215–229. doi:[10.1016/j.jmp.2004.03.002](https://doi.org/10.1016/j.jmp.2004.03.002)
- Lahl, O., Göritz, A.S., Pietrowsky, R., & Rosenberg, J. (2009). Using the World-Wide Web to obtain large-scale word norms: 190,212 ratings on a set of 2,654 German nouns. *Behavior Research Methods*, 41, 13–19. doi:[10.3758/BRM.41.1.13](https://doi.org/10.3758/BRM.41.1.13)
- Link, S.W. (1982). Correcting response measures for guessing and partial information. *Psychological Bulletin*, 92, 469–486. doi:[10.1037/0033-2909.92.2.469](https://doi.org/10.1037/0033-2909.92.2.469)
- Luce, R.D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.
- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, 44, 314–324. doi:[10.3758/s13428-011-0168-7](https://doi.org/10.3758/s13428-011-0168-7)
- 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, 205–235. doi:[10.1007/s11336-013-9374-9](https://doi.org/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](https://doi.org/10.3758/pbr.16.5.798)
- Meiser, T. (2005). A hierarchy of multinomial models for multidimensional source monitoring. *Methodology*, 1, 2–17. doi:[10.1027/1614-1881.1.1.2](https://doi.org/10.1027/1614-1881.1.1.2)
- Meiser, T., & Bröder, A. (2002). Memory for multidimensional source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 116–137. doi:[10.1037/0278-7393.28.1.116](https://doi.org/10.1037/0278-7393.28.1.116)
- 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](https://doi.org/10.1037/a0030734)
- Meyer, D.E., Yantis, S., Osman, A.M., & Smith, J.K. (1985). Temporal properties of human information processing: Tests of discrete versus continuous models. *Cognitive Psychology*, 17, 445–518. doi:[10.1016/0010-0285\(85\)90016-7](https://doi.org/10.1016/0010-0285(85)90016-7)
- Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, 42, 42–54. doi:[10.3758/BRM.42.1.42](https://doi.org/10.3758/BRM.42.1.42)
- Nadarevic, L., & Erdfelder, E. (2011). Cognitive processes in implicit attitude tasks: An experimental validation of the Trip Model. *European Journal of Social Psychology*, 41, 254–268. doi:[10.1002/ejsp.776](https://doi.org/10.1002/ejsp.776)
- Ollman, R. (1966). Fast guesses in choice reaction time. *Psychonomic Science*, 6, 155–156. doi:[10.3758/BF03328004](https://doi.org/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](https://doi.org/10.1073/pnas.1103880109)
- Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, 14, 465–471. doi:[10.1016/0005-1098\(78\)90005-5](https://doi.org/10.1016/0005-1098(78)90005-5)
- Rissanen, J. (1996). Fisher information and stochastic complexity. *IEEE Transactions on Information Theory*, 42, 40–47. doi:[10.1109/18.481776](https://doi.org/10.1109/18.481776)
- Roberts, S., & Sternberg, S. (1993). The meaning of additive reaction-time effects: Tests of three alternatives. D. E. Meyer, & S. Kornblum (Eds.) *Attention and performance 14: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience*, (pp. 611–653). Cambridge: The MIT Press.
- Schmittmann, V.D., Dolan, C.V., Raijmakers, M.E., & Batchelder, W.H. (2010). Parameter identification in multinomial processing tree models. *Behavior Research Methods*, 42, 836–846. doi:[10.3758/BRM.42.3.836](https://doi.org/10.3758/BRM.42.3.836)
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461–464.
- Schweickert, R., & Chen, S. (2008). Tree inference with factors selectively influencing processes in a processing tree. *Journal of Mathematical Psychology*, 52, 158–183. doi:[10.1016/j.jmp.2008.01.004](https://doi.org/10.1016/j.jmp.2008.01.004)
- Singmann, H., & Kellen, D. (2013). MPTinR: Analysis of multinomial processing tree models in R. *Behavior Research Methods*, 45, 560–575. doi:[10.3758/s13428-012-0259-0](https://doi.org/10.3758/s13428-012-0259-0)
- Smith, J.B., & Batchelder, W.H. (2008). Assessing individual differences in categorical data. *Psychonomic Bulletin & Review*, 15, 713–731. doi:[10.3758/PBR.15.4.713](https://doi.org/10.3758/PBR.15.4.713)
- Smith, J.B., & Batchelder, W.H. (2010). Beta-MPT: Multinomial processing tree models for addressing individual differences. *Journal of Mathematical Psychology*, 54, 167–183. doi:[10.1016/j.jmp.2009.06.007](https://doi.org/10.1016/j.jmp.2009.06.007)
- 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](https://doi.org/10.1037/0096-3445.117.1.34)
- Steffens, M.C., Buchner, A., Martensen, H., & Erdfelder, E. (2000). Further evidence on the similarity of memory processes in the process dissociation procedure and in source monitoring. *Memory & Cognition*, 28, 1152–1164. doi:[10.3758/BF03211816](https://doi.org/10.3758/BF03211816)
- Stretch, V., & Wixted, J.T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 1379–1396. doi:[10.1037/0278-7393.24.6.1379](https://doi.org/10.1037/0278-7393.24.6.1379)

- Townsend, J.T., & Ashby, F.G. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge: Cambridge University Press.
- Unkelbach, C., & Stahl, C. (2009). A multinomial modeling approach to dissociate different components of the truth effect. *Consciousness and Cognition*, 18, 22–38. doi:[10.1016/j.concog.2008.09.006](https://doi.org/10.1016/j.concog.2008.09.006)
- Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review*, 7, 424–465. doi:[10.3758/BF03214357](https://doi.org/10.3758/BF03214357)
- Van Zandt, T., & Ratcliff, R. (1995). Statistical mimicking of reaction time data: Single-process models, parameter variability, and mixtures. *Psychonomic Bulletin & Review*, 2, 20–54. doi:[10.3758/BF03214411](https://doi.org/10.3758/BF03214411)
- Vandekerckhove, J.S., Matzke, D., & Wagenmakers, E. (2015). Model comparison and the principle of parsimony. In *Oxford Handbook of Computational and Mathematical Psychology* (pp. 300–319). New York: Oxford University Press.
- Wagenmakers, E.-J., Verhagen, J., Ly, A., Bakker, M., Lee, M.D., Matzke, D., & Morey, R.D. (2015). A power fallacy. *Behavior Research Methods*, 47, 913–917. doi:[10.3758/s13428-014-0517-4](https://doi.org/10.3758/s13428-014-0517-4)
- 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](https://doi.org/10.3758/PBR.17.3.275)
- Yantis, S., Meyer, D.E., & Smith, J.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](https://doi.org/10.1037/0033-2909.110.2.350)