

Article

Cognitive Modeling in Intelligence Research: Advantages and Recommendations for their Application

Gidon T. Frischkorn^{1,†*} and Anna-Lena Schubert^{2,†}

¹ Heidelberg University, Institute of Psychology; gidon.frischkorn@psychologie.uni-heidelberg.de

² Heidelberg University, Institute of Psychology; anna-lena.schubert@psychologie.uni-heidelberg.de

* Correspondence: gidon.frischkorn@psychologie.uni-heidelberg.de; Tel.: +49-6221-54-7746

† Both authors contributed equally to this work.

Abstract: Mathematical models of cognition measure individual differences in cognitive processes, such as processing speed, working memory capacity, and executive functions, that may underlie general intelligence. As such, cognitive models allow identifying associations between specific cognitive processes and tracking the effect of experimental interventions aimed at the enhancement of intelligence on mediating process parameters. Moreover, cognitive models provide an explicit theoretical formalization of theories regarding specific cognitive process that may help overcoming ambiguities in the interpretation of fuzzy verbal theories. In this paper, we give an overview of the advantages of cognitive modeling in intelligence research and present models in the domains of processing speed, working memory, and selective attention that may be of particular interest for intelligence research. Moreover, we provide guidelines for the application of cognitive models in intelligence research, including data collection, the evaluation of model fit, and statistical analyses.

Keywords: Intelligence, Cognitive Modeling, Methods, Measurement, Practical Guidelines

One of the greatest challenges in intelligence research is the identification of cognitive processes underlying cognitive abilities and the measurement of process parameters giving rise to individual differences in intelligence [1]. The currently most discussed process parameters are the speed of information processing [e.g., 2,3], the capacity of short-term memory [e.g., 4,5] or working memory [e.g., 6–8], and the efficiency of executive functions [e.g., 5,9]. Individual differences in these cognitive processes are usually measured by behavioral indicators such as response times and accuracies in tasks engaging one specific cognitive process. The behavioral performance in these tasks is then used to quantify the relationship of these cognitive processes to performance in intelligence tests.

This approach presumes that a specific task provides a process-pure measure of a single cognitive process - an assumption that is often violated as most cognitive tasks do not measure one specific cognitive process, but rather a combination of several cognitive processes. For example, tasks measuring the efficiency of inhibitory processes such as the Stroop or Flanker task usually use reaction times as performance measures [10,11]. These reaction times arguably reflect not only the efficiency of inhibitory processes, but also basic information-processing speed. Another example are complex cognitive tasks such as complex span tasks measuring working memory capacity that do not only require the storage of information in the face of processing, but may also rely on attentional control processes and speed of information processing [12,13]. In sum, typical measures for a specific cognitive process thus require additional cognitive processes beyond the cognitive process aimed to be measured.

Two approaches are typically pursued to overcome this problem. First, variance decomposition methods may be used to isolate the variance of one latent cognitive process parameter from the influence of other variables [e.g., 4,6]. This method is feasible as long as there are "pure" measurements of the confounding cognitive processes available that can be controlled for. However, this approach may be resource- and time-consuming, as participants have to complete large test batteries including both measures of interest and of possible confounds.

A second approach to this measurement problem is to design experimental tasks that contain a baseline condition requiring the engagement of all confounding processes and an experimental condition that is equal to the baseline condition except for the insertion of one additional processing requirement of interest. Subtracting performance in the baseline condition from performance from the experimental condition is supposed to isolate the efficiency or speed of the added process [14]. However, it is questionable if the resulting difference scores only contain variance that can be attributed to the inserted process or if the insertion of additional processing demands may affect or interact with other task demands that are also reflected in the difference scores [15,16]. Moreover, the low between-subject variability and low reliability of difference scores in typical cognitive tasks renders the isolation of individual differences in experimental effects by means of difference scores virtually impossible [17,18].

In the present paper, we aim to demonstrate how mathematical models of cognition can be used to partially overcome these measurement problems by directly quantifying specific cognitive processes. Moreover, we will provide practical guidelines and recommendations for the use of cognitive models in intelligence research.

1. Advantages of cognitive modeling in intelligence research

Although often not explicitly in mind, each measurement of a cognitive process and more generally any property of a person is based on a model. Most often we use statistical models, such as classical test theory or latent variable models for this measurement procedure [19]. These models typically assume that the measured and observed behavior is the compound of some *true* or *latent* property of a person and of an *error* of measurement [20–22]. Across repeated measurements of the same property, this results in a distribution of observations of which the average or expected value given a person (i.e., the arithmetic mean) is conceptualized as the best estimate of the *true* person property, while deviations from this value (i.e., the standard deviation) correspond to the amount of *error* or uncertainty in the measurement.

Even though statistical models have proven to be very useful in the context of measurement, such models bear serious conceptual problems [19,23] and the selection of an adequate statistical model for measurement is anything but trivial. Apart from these general philosophical and epistemological problems of measurement with statistical models such as the ontological status of true-scores or latent variables and the adoption of a realist or constructionist perspective on science and measurement [19], all of these models have another serious shortcoming: Statistical models do not specify any psychological or cognitive processes underlying the *true* part of the measurement, but rather focus on separating *true* properties of a person from the *error* of measurement.

To overcome this problem, it has been recommended to use other statistical models such as ex-Gaussian- or Wald- distributions for reaction times [24–26], and Binomial-distributions for accuracies or mental test scores [27,28]. Although these distributions correspond more closely to the empirical shape of the distributions of observed variables, the parameters of these distributions do not consistently resemble indicators of distinct cognitive processes [see 29]. In sum, statistical models may be useful to quantify the amount of variance in a measurement attributable to the *true* personality trait (i.e., the reliability), however they do not allow any theoretically founded statements about the cognitive processes underlying the observed behavior or the latent personality trait.

Conversely, cognitive models may provide a mathematically-guided quantification of specific cognitive processes [30]. Specifically, cognitive models translate explicit verbal theories of cognitive processes in specific tasks into mathematical formulations of these theories. In this, the behavioral measures within a task are described as the result of different interacting processes or parameters of the model. The detailed interplay and interaction of these processes is specified within the formal architecture of the model and represents the assumptions the model makes with respect to a specific cognitive process. Thus, a cognitive model represents a formalized theory of a cognitive process that objectively states which parameters of the cognitive process affect differences in observed behavior

across conditions or individuals. The adequacy and validity of this formalization can be evaluated by parameter recovery studies and by testing the selective effects of theoretically-guided experimental manipulations on model parameters [31].

Taken together, cognitive models provide several advantages over statistical models: 1) They provide falsifiable descriptions of the cognitive process underlying behavioral responses in a specific task; 2) Model parameters can be interpreted in an objective and formally described manner; and 3) Model parameters can be used to quantify individual differences in specific cognitive processes based on the underlying model architecture.

2. Selecting cognitive models suitable for intelligence research

The field of cognitive modeling typically distinguishes between explanatory cognitive models or process models and cognitive measurement models [32]. Explanatory cognitive models¹ aim to provide formal explanations for variations across experimental conditions in specific paradigms in terms of cognitive processes. These models formally describe the architecture of a cognitive process and focus on the interplay of different mechanisms that lead to specific experimental results. In contrast, cognitive measurement models typically decompose the observed behavior of a person into meaningful parameters of a latent cognitive process. Thus, instead of explaining differences across individuals or experimental conditions, cognitive measurement models are highly flexible tools that reflect these differences in variations of their estimated parameters [for a comparison of these two model types, see 33]. Several cognitive measurement models rely on a more elaborated explanatory cognitive model. However, there are many cognitive measurement models that have been developed independently of any explanatory cognitive model [e.g., 34].

With respect to their application, explanatory models, such as the SOB-CS [35], the Slot-Averaging model [33], or the Interference model of visual working memory [32], resemble highly elaborated model architectures that specify detailed formal models for a cognitive process. These models are often very complex and require high computing power to calculate predictions for a given set of parameters. In contrast, measurement models, such as signal-detection theory [34], the two-high threshold model for recognition [36], or the drift-diffusion model [37], are mostly simplified descriptions of a cognitive process that can be generalized to a broad set of paradigms and observed variables. Beyond that, measurement models are easy to use and parameters of measurement models can either be readily calculated from observed variables or estimated with adequate fitting procedures.

In intelligence research, the use of cognitive measurement models is far more widespread than the use of explanatory cognitive models. Although explanatory cognitive models provide a powerful tool for comparing different theories with respect to their predictions for experimental paradigms and manipulations [see 38], their complexity and especially the lack of estimable parameters renders their application in intelligence research difficult. Still, results from explanatory cognitive models may provide the theoretical foundation for deciding for or against a specific cognitive measurement model.

Furthermore, there have been efforts to formulate explanatory models of intelligence test performance such as the Carpenter *et al.* [39] model for performance in the Raven matrices. In this model, Carpenter *et al.* [39] described different cognitive processes that are used while solving the Raven matrices. Some of these processes such as incremental encoding processes and rule induction for each matrix were used by all participants, while other processes such as the induction of abstract relations of the dynamic management of different goals in memory were specific to participants

¹ Although the term *process* models and *explanatory* models can mostly be used interchangeably, for comprehension and consistency we will use explanatory cognitive models when referring to computational models that describe the architecture of a cognitive process in detail and thereby aim to explain the results of specific manipulations in experimental paradigms.

with above-average performance. Although this model provides a strong theoretical explanation for individual differences in Raven performance, its application remains limited.

Cognitive measurement models may instead provide person- and condition-specific parameters for distinct cognitive processes. These person-specific parameters can be easily used as measures of individual differences in specific aspects of a cognitive processes, which can then be related to performance in intelligence tests. For instance, parameters of the drift-diffusion model, that will be introduced later, have been associated with performance in intelligence test or memory tasks [16,40–42]. In this, parameters from cognitive measurement models may thus provide insights on which cognitive processes are actually linked to intelligence.

While all cognitive measurement models are deliberate simplifications of the cognitive processes within a task and rely on often critically discussed assumptions, there is actually no alternative to the use of a measurement model, may it be statistical or cognitive. While most researches do not explicitly decide for a specific measurement model, by calculating the mean performance for a person in a task (as often done) they implicitly adopt a statistical measurement model that makes no explicit statements about the underlying cognitive processes of the measurement. It may even be argued that not explicitly deciding for a specific measurement model is practically similar to implicitly using the most simple cognitive model at hand: A model assuming that the observed variable directly represents the cognitive processes of interest. As already mentioned earlier, this assumption is almost always false. Therefore, we would argue that using explicit measurement models is always superior to equating observed variables with the cognitive process of interest.

To convey an idea of the benefits of the application of cognitive modeling in intelligence research, we will discuss three examples of cognitive models in the following sections. We selected different models describing cognitive processes of particular interest to intelligence research, such as decision making, working memory, and cognitive control, and demonstrate how they may be used to quantify individual differences in the respective cognitive processes. Following these examples, we then provide guidelines for choosing the appropriate model for a particular research question.

2.1. The drift diffusion model of binary decision making

The drift diffusion model (DDM) describes performance in two-alternative forced choice decisions tasks. The model assumes that evidence is accumulated in a random walk process until one of two decision thresholds is reached, the decision process is terminated, and a motor response (usually a key press) is initiated [see Figure 1 for an illustration; 37]. This evidence accumulation process can be described by a Wiener diffusion process that consists of a systematic component, the drift rate v , and normally distributed random noise with a mean of 0 and a variance of s^2 (this so-called diffusion constant s is usually fixed to a standardized value such as 0.1 or 1 for reasons of identifiability). The drift rate can be considered as a performance measure that directly quantifies the velocity of information uptake. In addition, the DDM quantifies the distance between decision thresholds as a measure of speed-accuracy trade-offs and decision cautiousness in the boundary separation as the parameter a , the starting point of evidence accumulation as the parameter z , and the time of non-decisional processes such as encoding and response preparation and execution as the parameter t_{er} or t_0 . Beyond these basic parameters, intra-individual variability parameters have been added to the DDM (i.e. st_0 , sv , and sz) to account for inter-trial variability within a person [37,43].

The validity of DDM parameters has been demonstrated both by parameter recovery studies [44] and by experimental validation studies [45–47]. Moreover, model parameters have been showing satisfying reliabilities estimated with test-retest correlations given sufficiently large trial numbers [48] and at least drift rates have been shown to exhibit trait-like properties [49]. Hence, it is not surprising that the DDM is the most frequently used cognitive model in intelligence research. By mathematically identifying parameters quantifying the speed of information uptake (v), the decision cautiousness (a), and encoding and movement times (t_{er}), it renders complicated experimental setups that have been used to dissociate these elements of the decision process with little success unnecessary [50].

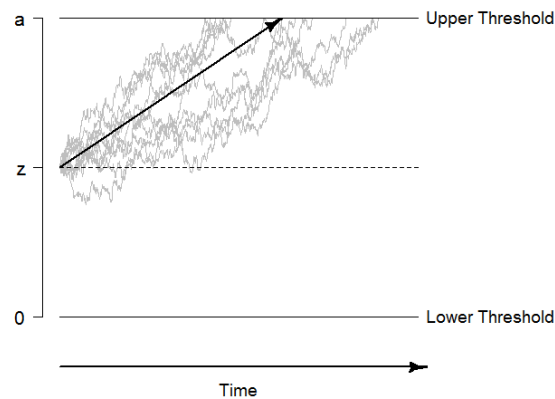


Figure 1. Graphical illustration of the DDM. The decision process starts at the starting point z , and information is accumulated until the boundary a is reached. The systematic part of the accumulation process, the drift rate v , is illustrated with the black arrow. The non-decision time t_0 is not included in this figure.

Several studies have reported positive associations between cognitive abilities and drift rates [e.g., 16,40,42,51–53], whereas the other model parameters have been shown to be largely unrelated to fluid intelligence [16,42,53]. The application of the DDM to data sets is made fairly easy by user-friendly software programs such as *DMAT* [54], *EZ* [55,56], and *fast-dm* [57].

The DDM is part of a larger family of evidence accumulation models that provide a general description of decision processes. Another member of this model family is the linear ballistic accumulator model [LBA; 58], which presumes that a number of independent accumulators race towards a common response threshold. Hence, where the DDM can only be applied to data from two-choice reaction times tasks, the LBA can be applied to data from both two- and multiple-choice reaction time tasks. Another member of this model family is the leaky, competing accumulator model [LCA; 59], which entails a number of stochastic accumulators that compete against each other via mutual inhibition to reach a decision threshold. Both models have not been applied in intelligence research yet, probably because they do not provide a single performance measure such as the drift rate of the DDM, as one drift parameter for each of the accumulators is estimated in LBA and LCA models, resulting in several drift rates.

2.2. The time-based resource-sharing model of working memory

The time-based resource sharing (TBRS) model of working memory started out as a verbal theory explaining the performance in complex span tasks measuring working memory capacity [60,61], but has been extended to verbal and visual WM in general [62–64]. The TBRS model claims that processing and the maintenance of stored information rely on the same attentional resource in working memory. Because of this attentional bottleneck, only one of these two processes can be performed at a given time. In detail, the model assumes that information stored in working memory decays over time, unless this decay is counteracted by an attentional refreshing process or verbal rehearsal. Moreover, additional processing demands as imposed in complex span tasks shift attention towards these secondary processing tasks, resulting in the decay of items stored in working memory (see Fig. 2 for an illustration). Altogether, working memory as conceptualized in the TBRS model continuously shares attentional resources between maintenance and processing in order to counteract decay of memory items and efficiently process information.

Over the past years there have been formalizations of the TBRS model as an explanatory model [38] and as a simplified measurement model [65]. Such models may be of great interest for the field of

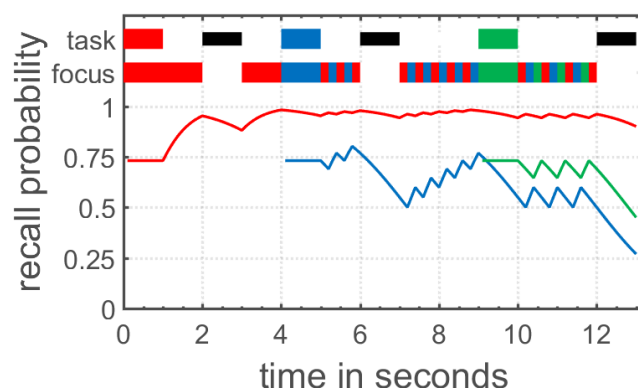


Figure 2. Visualization of the TBRS as implemented in the TBRS2 model by Gauvrit and Mathy [65]. At the top, the current task is displayed. A colored box represents a to be encoded memory item, a black box represents a distractor task, and a white box represents free time. Below, the focus of attention is shown. During free time, participants engage in refreshing of the already encoded memory item; during distractor tasks or encoding of other items, the already encoded memory items decay over time.

intelligence research, not only because intelligence is strongly related to working memory [8,66,67], but because the field is still in debate which specific cognitive processes within working memory - storage or executive processing - underlie its strong relationship with intelligence [4,6]. While the explanatory TBRS* model by Oberauer and Lewandowsky [38] is fairly complex and foremost an in-depth test for the experimental predictions of the TBRS theory, the TBRS2 implementation by Gauvrit and Mathy [65] provides a simplified version of the TBRS model and allows to estimate parameters that are directly linked to specific processes within the TBRS model. Such a model may provide person specific estimates of different processes in working memory, such as the encoding strength when an item is presented (i.e. the baseline β) or the speed of attentional refreshing (i.e. the refreshing rate r). These parameters may provide further information on which specific processes within working memory give rise to the strong relationship between working memory and intelligence.

As the mathematical implementations of the TBRS model have been developed only recently, there have not been any independent, systematic validation studies for the parameters of the model. Moreover, the psychometric properties of the model estimates (i.e., their reliability and validity) have not yet been assessed. Additionally, there still is a controversial debate in cognitive psychology whether decay actually is the core process limiting working memory capacity [68]. Although there are competing explanatory models of working memory questioning the role of decay as a limiting factor for working memory capacity [32,38], these models have not yet been translated into simple measurement models that allow estimating person-specific parameters of cognitive processes within working memory.² Until then, the TBRS2 model may provide a first step for including cognitive measurement models of working memory in intelligence research.

2.3. The shrinking spotlight model of selective attention

The shrinking spotlight model of selective attention describes processing in the Eriksen flanker task, in which participants have to respond according to the orientation of a centrally presented target

² Oberauer and Lewandowsky [69] are working on an alternative measurement model that is more closely connected to interference models of working memory [32,35]

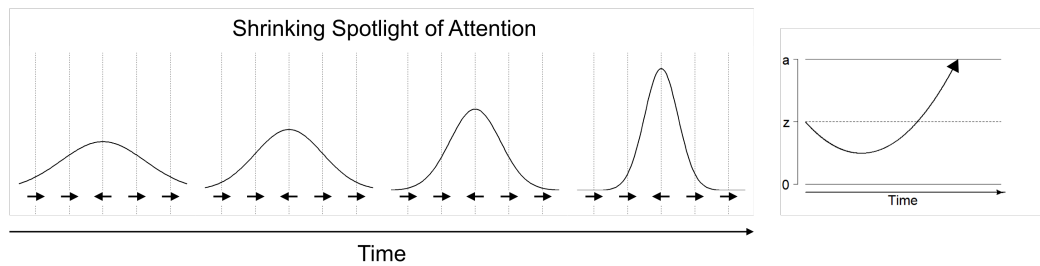


Figure 3. Illustration of the Shrinking Spotlight model for selective attention. The attentional focus narrows to the central arrow over time (left part). This results in a stronger weight of the critical information (i.e. the central stimulus) in the drift-rate of an associated diffusion process (right part).

arrow while ignoring irrelevant arrows flanking the target stimulus [10,70]. The shrinking spotlight model is an extension of the drift diffusion model of sequential processing: It assumes that both target and flanker arrows provide perceptual evidence p for a particular response weighted by the amount of attention a allocated to each of these stimuli. The drift rate consists of the sum of weighted perceptual evidence across all stimuli at a given time. Over time, attention is assumed to zoom in on the central arrow, reflecting a narrowing of the focus of selective attention on the target stimulus. Thus, the target stimulus is weighted more strongly in comparison to the flanker stimuli and therefore affects the drift rate more strongly over time (see Fig. 3). The initial width of attentional distribution is estimated in the attentional spotlight parameter sd_a , which reflects the standard deviation of a Gaussian distribution centered on the target stimulus, whereas the rate of attentional distribution reduction is estimated in the parameter r_d . In addition, the model also allows estimating the encoding and movement times in the t_{er} parameter, and the distance of symmetrical response thresholds from the starting point of evidence accumulation in the parameters A and $B = -A$.

The model has been shown to be able to account for data from a standard flanker task and experimental manipulations of task properties have been shown to specifically affect single model parameters [70]. Moreover, parameter recovery studies have shown that model parameters can be accurately recovered with as few as only 50 experimental trials [71]. However, simulation results have also shown that the model is not able to recover the attentional spotlight and the shrinking rate parameter accurately, because a wide initial spotlight with a high shrinking rate makes the same predictions as a narrow initial spotlight with a low shrinking rate [71]. Therefore, it has been recommended to calculate a composite measure of the duration of interference as the ratio of the two parameters, sd_a/d_r , to account for the trade-off during model estimation. Although there have not yet been any systematic analyses on the psychometric properties of parameter estimates, correlations of $.42 \leq r \leq .80$ between model parameters across different cue conditions of the Attention Network Test suggest at least moderate to good reliabilities, especially for the interference ratio with correlations of about $r \approx .80$ [72]. So far, the shrinking spotlight model has not yet been applied in intelligence research, but it would be promising to relate individual differences in the susceptibility to interference (as reflected in the interference parameter) to individual differences in intelligence test performance and working memory capacity to further explore the role of selective attention in cognitive abilities.

An alternative account of performance in the Eriksen flanker task is given by the dual-stage two phase model [73]. This model proposes two distinct processing stages: In the first processing stage, evidence accumulation is affected both by evidence accumulation towards the response associated with the target stimulus and by evidence accumulation towards the response associated with the flanker stimuli. At the same time, an attention-driven parallel evidence accumulation process selects a single stimulus for further processing. If this stimulus selection process terminates before response selection is finished, response selection enters a second stage with the drift rate being solely

determined by the selected stimulus. As of yet, model comparison studies have not yet decided which of the two models provides the best account of selective attention phenomena [70,71,73,74]. Both models can be fit to data and subsequently be compared using the R package *flankr* [75].

2.4. Guidelines for model selection

When deciding which cognitive model to use for a specific research question, there are some conceptual and practical issues to be considered in order to select the appropriate model: First, cognitive processes of interest for the current research question have to be identified. Second, an appropriate model providing a description of these cognitive processes has to be chosen. During this step, theoretical reasons for choosing one model over its alternatives should be considered. Third, experimental tasks congruent with the assumptions of the selected model should be selected to allow the valid estimation of model parameters. For an illustration of these decision steps see the upper part of Figure 4 (p. 9).

In general, discussing these issues during project planning aims to strengthen two important points for the conclusions from the modeling results. On the one hand, researchers should clarify which specific cognitive processes they are interested in and select a cognitive model accordingly. On the other hand, researchers should maximize the fit between the measurement or operationalization of a specific cognitive process (i.e. the task used) and the selected cognitive model.

For example, a group of researchers might be interested which cognitive processes in simple decision tasks are related to intelligence. Such tasks may require participants to decide whether a number is odd or even, or whether a letter is a vowel or consonant. They decide to use the drift diffusion model to quantify the different cognitive processes associated with binary decision making. However, one of these tasks has an additional switching demand, requiring participants to switch between the number and the letter decision [for an example, see 76]. Because this task is a binary decision task, the drift-diffusion model may still provide suitable estimates for the cognitive processes in such a task [77,78]. However, this task arguably requires more than one decision: On the one hand the decision which task is to be carried out, and on the other hand the decision corresponding to the task. Thus, this task does not fully fit the conceptualization of the drift-diffusion model as there may not be a single decision process but two. Therefore researchers should either think about using a different task that has a better fit to the basic assumptions of the drift-diffusion model or search for an alternative model that better fits the task they want to use.

This example reiterates the importance of an explicit and critical decision for a specific cognitive measurement model with respect to the measurement and operationalization that has already been pointed out before. As the developers of cognitive models often suggest a specific task suitable for parameter estimation [e.g. 65], the initial model publication is usually a good starting point for finding prototypical tasks that match the model assumptions. For popular cognitive models such as the diffusion model there are review articles summarizing studies in which the diffusion model was successfully applied to data from several different tasks [79]. Although some of these prototypical tasks may not provide the most suitable measures for a specific research question, they nevertheless constitute a meaningful starting point.

3. Guidelines for model application

After identifying an appropriate model based on theoretical considerations as outlined in the previous section, we strongly recommend to further plan the application of mathematical models ahead of data collection to ensure the interpretability and trustworthiness of the estimated model parameters. Specifically, three basic steps should be pursued when applying a cognitive model to a specific research question (see lower part of Figure 4, p. 9):

1. Researchers should plan their data collection to meet requirements for reliable and stable parameter estimates.
2. Model fit should be carefully evaluated after fitting the model to the empirical data.

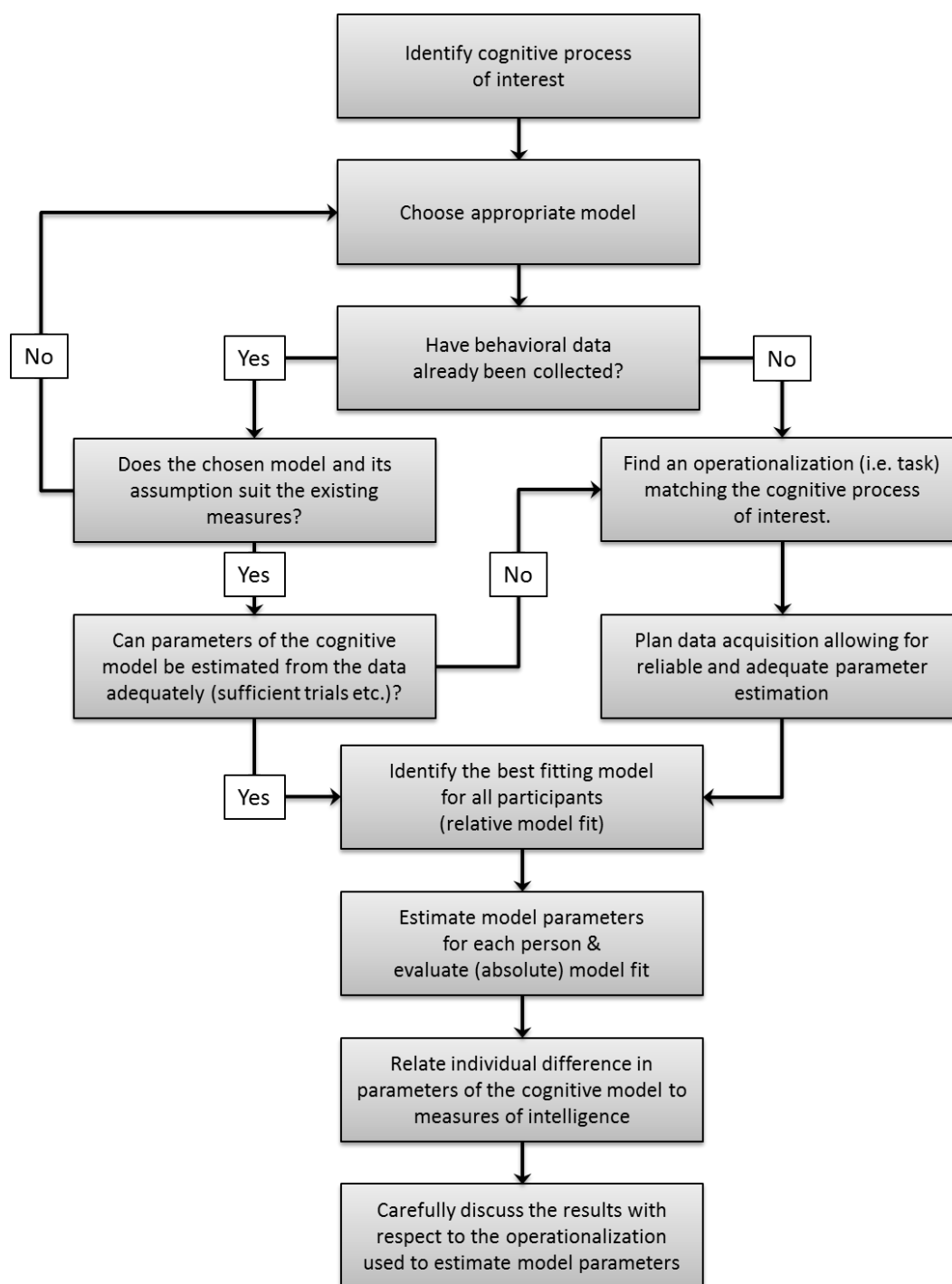


Figure 4. Flowchart illustrating the different planning and decision steps when using cognitive models in intelligence research.

3. Model parameters should be adequately related to other individual differences variables of interest such as intelligence test performances.

In the following section, we will provide step-by-step instructions using examples from the application of diffusion models in intelligence research, which may serve as guidelines when using any kind of cognitive model in individual differences research.

3.1. Design and data collection

3.1.1. Reliability and stability of estimated model parameters

The reliable estimation of model parameters from empirical data usually requires more data points than would be needed if only applying a statistical model to the data. For illustration, compare the description of reaction time distributions in decision tasks by a Gaussian distribution to the description by a diffusion model. When describing performance in a binary choice task by a Gaussian distribution, 20-30 trials are usually sufficient to provide reliable estimates of means and standard errors of the distribution [80]. When describing performance by a diffusion model, however, many more trials are needed because model parameters are not calculated analytically, but are found by fitting them to empirical response time distributions in an iterative process. Hence, a small number of trials will result in an inadequate representation of the full response time distribution and will therefore impair the estimation of model parameters describing distributional elements beyond measures of central tendency [81].

For the basic DDM (with the four parameters drift rate, boundary separation, starting point, and non-decision time), simulation studies have shown that 100 trials are sufficient to produce relatively reliable estimates of drift rates and that no further increases in parameter reliabilities are gained by increasing trial numbers beyond 500 trials [81]. For other measurement models less prominently used in individual differences research, such systematic simulation studies have not yet been conducted. Therefore, we urge researchers interested in applying less frequently used models to run a simulation study before starting data collection to determine how many experimental trials are needed for reliable parameter estimates.

3.1.2. Trait, situation, and task characteristics of model parameters

In addition, it is important to consider to what degree individual differences in model parameters reflect individuals' personality traits or abilities, and to to what degree they reflect task-specific characteristics, state-specific characteristics, and unsystematic measurement error. Imagine applying a model of verbal working memory to complex span data: Model parameters such as the individual rate of verbal refreshing or the ability to resist interference from distracting stimuli would reflect both individuals' *general abilities* in verbal refreshing and inhibition of interference as well as their abilities to maintain memory stimuli *in this specific task*. Depending on the research question, researchers may be more interested in the general ability to maintain information in working memory as reflected in those parameters across different working memory tasks, or they may be interested in the specific ability to maintain information in working memory in precisely this task.

Usually, intelligence research questions are more likely to concern abilities generalized across specific operationalizations and situations than abilities in specific operationalizations or situations. However, model parameters estimated in a specific task are always going to contain both trait-, state- and task-specific amounts of variance [82,83]. For example, a latent state-trait analysis of DDM parameters in elementary cognitive tasks revealed that only about 45 percent of the variance in task-specific drift rates was accounted for by the common trait, and that only about 30 to 35 percent of the variance in task-specific boundary separation and non-decision time parameters were accounted for by their respective common traits [49]. Therefore, if a research question using cognitive models in intelligence research concerns performance in certain cognitive processes that is generalizable across

specific operations, it may be worthwhile to design a test battery consisting of three or more tasks to which the cognitive model can be applied. Averaged or latent performance in process parameters across tasks will then allow a more precise estimate of individuals' performance in model parameters that is independent of specific task or situation characteristics.

3.2. Evaluation of model fit

3.2.1. Relative model fit: Which model provides the best account for the data?

After finishing data collection, but before relating model parameters to intelligence tests or other covariates, it is necessary to evaluate how well a chosen model describes the empirical data and to possibly adjust model specifications to increase model fit. When a model is fitted to empirical data, it has to be decided how many and which model parameters will be estimated and which model parameters will be fixed, because they are not expected to be affected by task characteristics or not of interest for the current research question. Moreover, if experimental tasks contain several conditions, it may be necessary to decide which (if any) parameters are allowed to vary between conditions. It may even be desirable to split data from different conditions into separate data sets for separate model estimations to be able to subsequently model these separately estimated model parameters as latent variables. For this purpose, it may be helpful to reflect on the relationship between model complexity and the stability of parameter estimates: The more parameters of a model are estimated, the more likely it is to provide an accurate account of the data. However, if too many model parameters are estimated relative to the number of experimental trials, the stability of parameter estimates will be impaired [81,84].

Therefore, we suggest fitting several models to the empirical data containing different combinations of estimated or fixed parameters that are consistent with the current research question, unless there are strong theoretical reasons to decide on a specific model instantiation a priori. These models can then be compared based on parsimonious fit indices such as the Akaike Information Criterion [AIC; 85] or the Bayesian Information Criterion [BIC; 86], which take into account both model fit and model parsimony, to identify the model making the best trade-off between model fit and model complexity. As mentioned before, this model comparison step may not be necessary when a priori deciding for a specific instantiation of the model.

However, this model comparison approach only addresses one element of model fit evaluation, *relative model fit*. By identifying the best-fitting specification of model out of a number of alternative specifications, it is possible to identify the model providing the best description of the empirical data. However, this does not guarantee that the best-fitting model provides a *good* description of the data.

3.2.2. Absolute model fit: How well does the selected model describe the data?

Therefore, in the next step the *absolute model fit* has to be evaluated to decide if the model can be accepted for all data sets. Absolute model fit is typically ascertained by either (a) statistical tests of model fit, (b) goodness-of-fit (GOF) indices, or (c) graphical inspections of model fit.

Statistical tests of model fit quantify the discrepancy between the empirical data and model predictions by means of a test statistic that is then tested for significance. However, this null hypothesis-testing of model fit contains several problems, as the power of statistical tests is closely tied to the amount of data available. When only few trials are available, statistical tests may not be capable of rejecting the null hypothesis due to a lack of power, whereas when the trial number is large, statistical tests tend to become overly sensitive and detect even irrelevant deviations between the empirical data and model predictions [87]. To overcome some of the problems associated with null hypothesis testing, it has been suggested to simulate a large number of data sets based on the estimated model parameters, fit the model to each of the simulated data sets, and derive the 95 percent or 90 percent quantile of the resulting distribution of *p*-values as a critical value for the statistical tests

of the originally estimated models [88,89]. However, models will still be accepted with an unknown error probability.

Goodness-of-fit indices are much more common in individual differences research, where they are used to evaluate the model fit of structural equation models [90], than in cognitive modeling. GOF indices standardize test statistics and take into account both model complexity and the number of data points. Typically, GOF indices have a fixed value range from 0 to 1 with certain cut-off values that indicate acceptable or good model fit. GOF indices are less frequently used in cognitive modeling, probably because several GOF indices used in structural equation modeling require the comparison of the actual model to a minimally plausible baseline model, which cannot be easily specified for most cognitive models. However, it has been recently suggested to adapt the root mean square error of approximation for the evaluation of cognitive models than can be fitted with a χ^2 -distribution, such as the diffusion model [87].

Finally, a third and widespread approach to the evaluation of absolute model fit is to graphically compare the empirical data to model predictions. To graphically inspect model fit, empirical data can be plotted against or overlaid by model predictions separately for each participant or aggregated over participants. This process can be rather time-consuming in larger samples if each participant is inspected individually. Moreover, it is important to be aware of the fact that graphical evaluations of model fit are inherently subjective and may therefore lead to spurious conclusions [91]. Having two independent raters evaluate model fit and discuss their conclusions may therefore increase the objectivity of the evaluation process.

If individual data sets can be identified that do not provide a satisfying model fit, raw data should be inspected for coding errors or outliers that may need to be removed (e.g., extremely fast reaction times with decision behavior close to guessing in a decision task). If model fit remains unacceptable, individual data sets may then need to be removed from further analyses, as it cannot be ascertained that the model parameters characterize the cognitive processes in the task accurately.

3.3. Relating model parameters to intelligence test performance

Finally, after reliable estimates for the best fitting model have been obtained, the model parameters should be related to measures of intelligence. While this seems straightforward, there is actually one major methodological concern. Specifically, researchers usually obtain one or more person-specific estimates for each model parameter of interest across different tasks or experimental conditions, just like they do when using aggregated performance measures such as accuracies or mean reaction times. Then the relationship of these model parameters with intelligence test scores is estimated by means of correlations or structural equation modeling. However, this approach represents a sequential analysis plan that treats the estimated parameters as manifest variables when quantifying the relationship between parameters of the cognitive model and the intelligence measures.

Treating estimated model parameters as manifest variables ignores the uncertainty that these parameters inherit from estimation and leads to an underestimation of standard errors in the second analysis step [92]. In fact, this is the case both for behavioral aggregates, such as mean reaction times or proportion correct, and for model parameters that are estimated from behavioral data or calculated from aggregate performance measures. Although this does not necessarily affect the estimated size of the relation between parameters obtained from cognitive models and intelligence measures, a sequential analysis plan always leads to an overestimation of the statistical significance of the estimated relationships [93].

A solution to this problem is hierarchical modeling [94,95]³. In hierarchical modeling approaches, parameters of a cognitive model can be estimated not only for all participants but across various tasks simultaneously. Additionally, relationships with third variables can be estimated in the same step. On the one hand, such models avoid underestimating the standard errors of the relationship between model parameters and third variables such as intelligence measures by simultaneously estimating the model parameters and their relationship to intelligence [for an example of hierarchical models of the worst performance rule, see 93]. On the other hand, by assuming that the distribution of model parameters across individuals follows a higher order distribution⁴ (so called hyper-priors), hierarchical models do not estimate parameters for each individual independently, but instead estimate model parameter for each individual informed by the parameter estimates from all other individuals. Not only does this render the parameter estimation more robust, but it also allows obtaining reliable estimates for the parameters of a cognitive model for each individual with fewer trials [for an example, see the hierarchical diffusion model: 97].

A serious complication of hierarchical modeling is that these models typically have to be explicitly specified and translated into code for each application, and that software solutions for parameter estimation are still rare. Nevertheless, hierarchical models do provide the mathematically accurate and sound solution for estimating the relationship between estimated model parameters and intelligence measures. Still, the sequential estimation of model parameters and their relationship to intelligence test scores seems to yield results comparable to hierarchical approaches [93]. In conclusion, while sequential approaches may overestimate the statistical significance of the relationship between model parameters and covariates (biasing inference), they nevertheless provide reasonable and unbiased estimates of the effect size of this relationship. For the future, it would be desirable that the application of cognitive modeling in the field of intelligence research or individual differences in general leads to the development of further simple software solutions or R packages [98,99] that simplify the use of hierarchical models.

4. Critical interpretation of the results

Regardless how the relationship between parameters from a cognitive model and intelligence measures is estimated, ultimately this relationship has to be interpreted on a conceptual level. Although parameters of a cognitive model provide more specific information about the cognitive process underlying the behavioral responses, these parameters still have to be interpreted with respect to the operationalization of the cognitive process. For instance, the diffusion model can be estimated in a broad set of tasks, ranging from perceptual judgment tasks (e.g., a random-dot motion task), over elementary cognitive tasks (e.g., Posner or Sternberg task), to even more complex memory tasks. In all of these different tasks, the diffusion model estimates the same set of parameters (i.e., drift rates, boundary separations, and non-decision times). However, this alone does not imply that model parameters estimated in the different tasks can be interpreted the same way. Specifically, the drift rate estimated in a random-dot motion task may represent the speed of perceptual information accumulation towards one response alternative. In a memory recognition task, however, the drift rate would rather be interpreted as the signal-to-noise ratio of the representation in memory. In conclusion, the interpretation of parameters of a cognitive model always relies on the specific experimental task.

In general, a cognitive model always represents a structural description of the behavioral measures from a specific task. The semantic meaning of the parameters of a model, however, can only be obtained with respect to the context (i.e., the task or materials) they are estimated in. Consider

³ These two references focus on Bayesian hierarchical modeling. While Bayesian parameter estimation might have additional advantages over frequentist estimation approaches [96], the benefits of hierarchical modeling apply to both Bayesian and frequentist methods.

⁴ Typically a Gaussian distribution with a mean and standard deviation is assumed.

the following equation: $v = x/t$. On its own, this equation is merely a structural description how v can be obtained from x and t . In contrast, if the context of the observations of x as a distance between two points, and t as the time taken to get from one point to the other is known, then v can reasonably be interpreted as the average speed of travel. It is just the same with parameters from any cognitive model: Without the context of their estimation they are merely transformations or estimated simplifications of the observed variables. Adding the semantical context of the observations however allows to interpret the parameters in a meaningful way.

All in all, the matter of adequately interpreting parameters of a cognitive model relates to a broader issue, namely validity. On the one hand, there is the question in how far a cognitive model provides a valid description of the cognitive process underlying the behavioral responses in a task. On the other hand, there is the question in how far individual differences in these parameters can be generalized across different task and assumed to represent between person variation in a more general and task-unspecific cognitive process. These are hardly problems that can be solved within a single study, but there is a combined effort needed to establish which parameters of cognitive models provide meaningful representations of individual differences in specific aspects of cognitive processing. However, following certain guidelines and carefully discussing the underlying assumptions and the operationalization when using a cognitive model provides a more explicit approach to measuring individual differences in cognitive processes, and thus represents a decisive improvement compared to the prevailing methods.

5. Conclusion

Altogether, incorporating cognitive models in intelligence research provides numerous advantages. On the one hand, cognitive models provide explicit theoretical descriptions of cognitive processes that may underlie individual differences in intelligence. On the other hand they allow to estimate person specific parameters for each individual that can be related to measures of intelligence. Therefore, cognitive models allow to relate theoretically founded measures of individual differences in parameters of cognitive processes to individual differences in intelligence and to overcome the fuzzy theoretical interpretation of behavioral indicators such as reaction times or accuracies.

Beyond that, cognitive models may allow identifying the effects of experimental or pharmacological interventions and training interventions on specific cognitive processes. For example, the Shrinking Spotlight Model of Selective Attention might be used to test if a training intervention aimed at improving selective attention actually affects interference parameters of the model or if the intervention only reduces non-decision times or response thresholds. In a similar vein, the drift diffusion model might be used to characterize experimental effects of a pharmacological intervention on mental speed by distinguishing an increase in the velocity of evidence accumulation from an increase in motor response times. Last but not least, cognitive process parameters could not only be related to intelligence differences, but also to individual differences in neural measures related to cognitive abilities, and may thus provide a different and possibly more complete perspective on the neuro-cognitive processes giving rise to individual differences in intelligence. Taken together, the application of cognitive models provides an exciting new avenue for research on the neurocognitive processes underlying intelligence.

Acknowledgments: We thank Debora Fieberg and Ben Riemenschneider for providing Matlab and R code for the illustration of the DDM and the TBRS2 model.

Author Contributions: Gidon T. Frischkorn and Anna-Lena Schubert both wrote the manuscript, and revised and commented on the parts of the other author respectively. Since both authors were equally involved in the structural and conceptual design of the manuscript, the first authorship was shared between the two.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SOB-CS: Serial-Order in Box Model for Complex Span Tasks

DDM: Drift-Diffusion Model

LBA: Linear Ballistic Accumulator Model

LCA: Leaky Competing Accumulator Model

TBRS: Time-base resource sharing theory/model

AIC: Akaike Information Criterion

BIC: Bayesian Information Criterion

GOF: Goodness-of-fit

Bibliography

1. de Boeck, P. Intelligence, Where to Look, Where to Go? *Journal of Intelligence* **2013**, *1*, 5–24.
2. Jensen, A.R. *Clocking the Mind: Mental Chronometry and Individual Differences*; Elsevier: Amsterdam, 2006.
3. Kyllonen, P.C.; Zu, J. Use of Response Time for Measuring Cognitive Ability. *Journal of Intelligence* **2016**, *4*.
4. Colom, R.; Abad, F.J.; Ángeles Quiroga, M.; Shih, P.C.; Flores-Mendoza, C. Working memory and intelligence are highly related constructs, but why? *Intelligence* **2008**, *36*, 584 – 606.
5. Wongupparaj, P.; Kumari, V.; Morris, R.G. The relation between a multicomponent working memory and intelligence: The roles of central executive and short-term storage functions. *Intelligence* **2015**, *53*, 166 – 180.
6. Conway, A.R.; Cowan, N.; Bunting, M.F.; Theriault, D.J.; Minkoff, S.R. A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence* **2002**, *30*, 163 – 183.
7. Engle, R.W.; Tuholski, S.W.; Laughlin, J.E.; Conway, A.R. Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of experimental psychology. General* **1999**, *128*, 309–331.
8. Kyllonen, P.C.; Christal, R.E. Reasoning ability is (little more than) working-memory capacity?! *Intelligence* **1990**, *14*, 389 – 433.
9. Kovacs, K.; Conway, A.R.A. Process Overlap Theory: A Unified Account of the General Factor of Intelligence. *Psychological Inquiry* **2016**, *27*, 151–177.
10. Eriksen, B.A.; Eriksen, C.W. Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics* **1974**, *16*, 143–149.
11. Stroop, J.R. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology* **1935**, *18*, 643 – 662.
12. Kane, M.J.; Meier, M.E.; Smeekens, B.A.; Gross, G.M.; Chun, C.A.; Silvia, P.J.; Kwapil, T.R. Individual differences in the executive control of attention, memory, and thought, and their associations with schizotypy. *Journal of Experimental Psychology: General* **2016**, *145*, 1017 – 1048.
13. McVay, J.C.; Kane, M.J. Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. *Journal of Experimental Psychology: General* **2012**, *141*, 302 – 320.
14. Donders, F. On the speed of mental processes. *Acta Psychologica* **1969**, *30*, 412 – 431.
15. Friston, K.J.; Price, C.J.; Fletcher, P.; Moore, C.; Frackowiak, R.S.; Dolan, R.J. The trouble with cognitive subtraction. *NeuroImage* **1996**, *4*, 97–104.
16. Schubert, A.L.; Hagemann, D.; Voss, A.; Schankin, A.; Bergmann, K. Decomposing the relationship between mental speed and mental abilities. *Intelligence* **2015**, *51*, 28 – 46.
17. Cronbach, L.J.; Furby, L. How we should measure change: Or should we? *Psychological Bulletin* **1970**, *74*, 68–80.
18. Hedge, C.; Powell, G.; Sumner, P. The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods* **2017**.
19. Borsboom, D. *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics.*; Cambridge University Press, 2005.

20. Lord, F.; Novick, M.; Birnbaum, A. *Statistical theories of mental test scores.*; Addison-Wesley, 1968.
21. Schmidt, F.L.; Hunter, J.E. Theory testing and measurement error. *Intelligence* **1999**, *27*, 183 – 198.
22. Borsboom, D. Latent variable theory. *Measurement: Interdisciplinary Research and Perspectives* **2008**, *6*, 25 – 53.
23. Borsboom, D.; Mellenbergh, G.J. True scores, latent variables and constructs: A comment on Schmidt and Hunter. *Intelligence* **2002**, *30*, 505 – 514.
24. Schwarz, W. The ex-Wald distribution as a descriptive model of response times. *Behavior Research Methods, Instruments, & Computers* **2001**, *33*, 457–469.
25. Schwarz, W. On the Convolution of inverse Gaussian and exponential Random Variables. *Communications in Statistics - Theory and Methods* **2002**, *31*, 2113–2121.
26. Miller, R.; Scherbaum, S.; Heck, D.W.; Goschke, T.; Enge, S. On the Relation Between the (Censored) Shifted Wald and the Wiener Distribution as Measurement Models for Choice Response Times. *Applied Psychological Measurement* **2018**, *42*, 116–135.
27. Keats, J.A.; Lord, F.M. A theoretical distribution for mental test scores. *Psychometrika* **1962**, *27*, 59 – 72.
28. Wilcox, R.R. Estimating true score in the compound binomial error model. *Psychometrika* **1978**, *43*, 245 – 258.
29. Matzke, D.; Wagenmakers, E.J. Psychological interpretation of the ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review* **2009**, *16*, 798–817.
30. Farrell, S.; Lewandowsky, S. *Computational Modeling of Cognition and Behavior*; Cambridge University Press, 2018.
31. Heathcote, A.; Brown, S.; Wagenmakers, E. An Introduction to Good Practices in Cognitive Modeling. In *An Introduction to Model-Based Cognitive Neuroscience*; Forstmann, B.; Wagenmakers, E., Eds.; Springer: New York and NY, 2015.
32. Oberauer, K.; Lin, H.Y. An interference model of visual working memory. *Psychological Review* **2017**, *124*, 21 – 59.
33. Zhang, W.; Luck, S.J. Discrete fixed-resolution representations in visual working memory. *Nature* **2008**, *453*, 233 EP –.
34. Banks, W.P. Signal detection theory and human memory. *Psychological Bulletin* **1970**, *74*, 81 – 99.
35. Oberauer, K.; Lewandowsky, S.; Farrell, S.; Jarrold, C.; Greaves, M. Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review* **2012**, *19*, 779–819.
36. Bröder, A.; Schütz, J. 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* **2009**, *35*, 587 – 606.
37. Ratcliff, R. A theory of memory retrieval. *Psychological Review* **1978**, *85*, 59 – 108.
38. Oberauer, K.; Lewandowsky, S. Modeling working memory: a computational implementation of the Time-Based Resource-Sharing theory. *Psychonomic Bulletin & Review* **2011**, *18*, 10–45.
39. Carpenter, P.A.; Just, M.A.; Shell, P. What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review* **1990**, *97*, 404 – 431.
40. Schmiedek, F.; Oberauer, K.; Wilhelm, O.; Süß, H.M.; Wittmann, W.W. Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General* **2007**, *136*, 414 – 429.
41. Ratcliff, R.; Schmiedek, F.; McKoon, G. A diffusion model explanation of the worst performance rule for reaction time and IQ. *Intelligence* **2008**, *36*, 10 – 17.
42. Schmitz, F.; Wilhelm, O. Modeling Mental Speed: Decomposing Response Time Distributions in Elementary Cognitive Tasks and Correlations with Working Memory Capacity and Fluid Intelligence. *Journal of Intelligence* **2016**, *4*.
43. Ratcliff, R.; Tuerlinckx, F. Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review* **2002**, *9*, 438–481.
44. van Ravenzwaaij, D.; Oberauer, K. How to use the diffusion model: Parameter recovery of three methods: EZ, fast-dm, and DMAT. *Journal of Mathematical Psychology* **2009**, *53*, 463 – 473.
45. Ratcliff, R.; McKoon, G. The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation* **2008**, *20*, 873–922.

46. Voss, A.; Rothermund, K.; Voss, J. Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition* **2004**, *32*, 1206–1220.
47. Lerche, V.; Voss, A. Experimental validation of the diffusion model based on a slow response time paradigm. *Psychological Research* **2017**.
48. Lerche, V.; Voss, A. Retest reliability of the parameters of the Ratcliff diffusion model. *Psychological research* **2017**, *81*, 629–652.
49. Schubert, A.L.; Frischkorn, G.T.; Hagemann, D.; Voss, A. Trait Characteristics of Diffusion Model Parameters. *Journal of Intelligence* **2016**, *4*.
50. Longstreth, L.E. Jensen's reaction-time investigations of intelligence: A critique. *Intelligence* **1984**, *8*, 139 – 160.
51. Ratcliff, R.; Thapar, A.; McKoon, G. Individual differences, aging, and IQ in two-choice tasks. *Cognitive Psychology* **2010**, *60*, 127–157.
52. Ratcliff, R.; Thapar, A.; McKoon, G. Effects of aging and IQ on item and associative memory. *Journal of experimental psychology* **2011**, *140*, 464–487.
53. Schulz-Zhecheva, Y.; Voelkle, M.C.; Beauducel, A.; Biscaldi, M.; Klein, C. Predicting Fluid Intelligence by Components of Reaction Time Distributions from Simple Choice Reaction Time Tasks. *Journal of Intelligence* **2016**, *4*.
54. Vandekerckhove, J.; Tuerlinckx, F. Diffusion model analysis with MATLAB: A DMAT primer. *Behavior Research Methods* **2008**, *40*, 61–72.
55. Wagenmakers, E.J.; Van Der Maas, H.L.J.; Grasman, R.P.P.P. An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review* **2007**, *14*, 3–22.
56. Wagenmakers, E.J.; van der Maas, H.L.J.; Dolan, C.V.; Grasman, R.P.P.P. EZ does it! Extensions of the EZ-diffusion model. *Psychonomic Bulletin & Review* **2008**, *15*, 1229–1235.
57. Voss, A.; Voss, J. Fast-dm: A free program for efficient diffusion model analysis. *Behavior Research Methods* **2007**, *39*, 767–775.
58. Brown, S.D.; Heathcote, A. The simplest complete model of choice response time: linear ballistic accumulation. *Cognitive psychology* **2008**, *57*, 153–178.
59. Usher, M.; McClelland, J.L. The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review* **2001**, *108*, 550–592.
60. Barrouillet, P.; Bernardin, S.; Camos, V. Time Constraints and Resource Sharing in Adults' Working Memory Spans. *Journal of Experimental Psychology: General* **2004**, *133*, 83 – 100.
61. Barrouillet, P.; Bernardin, S.; Portrat, S.; Vergauwe, E.; Camos, V. Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **2007**, *33*, 570 – 585.
62. Vergauwe, E.; Barrouillet, P.; Camos, V. Visual and spatial working memory are not that dissociated after all: A time-based resource-sharing account. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **2009**, *35*, 1012 – 1028.
63. Vergauwe, E.; Camos, V.; Barrouillet, P. The impact of storage on processing: How is information maintained in working memory?. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **2014**, *40*, 1072 – 1095.
64. Barrouillet, P.; Portrat, S.; Camos, V. On the law relating processing to storage in working memory. *Psychological Review* **2011**, *118*, 175 – 192.
65. Gauvrit, N.; Mathy, F. Mathematical transcription of the 'timebased resource sharing' theory of working memory. *British Journal of Mathematical and Statistical Psychology* **2018**, *71*, 146 – 166.
66. Colom, R.; Rebollo, I.; Palacios, A.; Juan-Espinosa, M.; Kyllonen, P.C. Working memory is (almost) perfectly predicted by g. *Intelligence* **2004**, *32*, 277 – 296.
67. Gignac, G.E. Working memory and fluid intelligence are both identical to g?! Reanalyses and critical evaluation. *Psychology Science* **2007**, *49*, 187 – 207.
68. Oberauer, K.; Farrell, S.; Jarrold, C.; Lewandowsky, S. What limits working memory capacity?. *Psychological Bulletin* **2016**, *142*, 758 – 799.
69. Oberauer, K.; Lewandowsky, S. Simple Measurement Models for Complex Working-Memory Tasks in prep.
70. White, C.N.; Ratcliff, R.; Starns, J.J. Diffusion models of the flanker task: discrete versus gradual attentional selection. *Cognitive psychology* **2011**, *63*, 210–238.

71. White, C.N.; Servant, M.; Logan, G.D. Testing the validity of conflict drift-diffusion models for use in estimating cognitive processes: A parameter-recovery study. *Psychonomic bulletin & review* **2017**.
72. White, C.N.; Curl, R. A Spotlight Diffusion Model Analysis of the Attentional Networks Task, 2017.
73. Huebner, R.; Steinhauser, M.; Lehle, C. A dual-stage two-phase model of selective attention. *Psychological Review* **2010**, *117*, 759–784.
74. Huebner, R.; Tobel, L. Does attentional selectivity in the flanker task improve discretely or gradually? *Frontiers in psychology* **2012**, *3*, 434.
75. Grange, J.A. flankr: An R package implementing computational models of attentional selectivity. *Behavior research methods* **2016**, *48*, 528–541.
76. Rogers, R.D.; Monsell, S. Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General* **1995**, *124*, 207 – 231.
77. Schmitz, F.; Voss, A. Decomposing task-switching costs with the diffusion model. *Journal of Experimental Psychology: Human Perception and Performance* **2012**, *38*, 222 – 250.
78. Schmitz, F.; Voss, A. Components of task switching: A closer look at task switching and cue switching. *Acta Psychologica* **2014**, *151*, 184 – 196.
79. Ratcliff, R.; McKoon, G. The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation* **2008**, *20*, 873 – 922.
80. Miller, J.; Ulrich, R. Mental chronometry and individual differences: Modeling reliabilities and correlations of reaction time means and effect sizes. *Psychonomic Bulletin & Review* **2013**, *20*, 819 – 858.
81. Lerche, V.; Voss, A.; Nagler, M. How many trials are required for parameter estimation in diffusion modeling? A comparison of different optimization criteria. *Behavior Research Methods* **2017**, *49*, 513–537.
82. Steyer, R.; Schmitt, M.; Eid, M. Latent state–trait theory and research in personality and individual differences. *European Journal of Personality* **1999**, *13*, 389 – 408.
83. Eid, M. A multitrait-multimethod model with minimal assumptions. *Psychometrika* **2000**, *65*, 241 – 261.
84. van Ravenzwaaij, D.; Donkin, C.; Vandekerckhove, J. The EZ diffusion model provides a powerful test of simple empirical effects. *Psychonomic Bulletin & Review* **2017**, *24*, 547 – 556.
85. Akaike, H. Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory, Tsahkadsor, Armenia, USSR, September 2-8, 1971*; Petrov, B.; Csáki, F., Eds.; Akadémiai Kiadó: Budapest, 1973.
86. Schwarz, G. Estimating the Dimension of a Model. *Ann. Statist.* **1978**, *6*, 461–464.
87. Schubert, A.L.; Hagemann, D.; Voss, A.; Bergmann, K. Evaluating the model fit of diffusion models with the root mean square error of approximation. *Journal of Mathematical Psychology* **2017**, *77*, 29 – 45.
88. Clauset, A.; Shalizi, C.R.; Newman, M.E.J. Power-Law Distributions in Empirical Data. *SIAM Review* **2009**, *51*, 661–703.
89. Voss, A.; Nagler, M.; Lerche, V. Diffusion Models in Experimental Psychology. *Experimental Psychology* **2013**, *60*, 385–402. PMID: 23895923.
90. Jackson, D.L.; Gillaspay, J.A.J.; Purc-Stephenson, R. Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods* **2009**, *14*, 6–23.
91. D’Agostino, R.B. Graphical analyses. In *Goodness-of-fit techniques*; D’Agostino, R.B.; Stephens, M.A., Eds.; Marcel Dekker: New York and NY, 1986; pp. 7–62.
92. Skrandal, A.; Laake, P. Regression among factor scores. *Psychometrika* **2001**, *66*, 563 – 575.
93. Frischkorn, G.T.; Schubert, A.L.; Neubauer, A.B.; Hagemann, D. The Worst Performance Rule as Moderation: New Methods for Worst Performance Analysis. *Journal of Intelligence* **2016**, *4*.
94. Lee, M.D. How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology* **2011**, *55*, 1 – 7.
95. Lee, M.D.; Wagenmakers, E.J. *Bayesian cognitive modeling: A practical course.*; Cambridge University Press, 2013.
96. Kruschke, J.K.; Liddell, T.M. The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review* **2017**.
97. Vandekerckhove, J.; Tuerlinckx, F.; Lee, M.D. Hierarchical diffusion models for two-choice response times. *Psychological Methods* **2011**, *16*, 44 – 62.
98. Heck, D.W.; Arnold, N.R.; Arnold, D. TreeBUGS: An R Package for Hierarchical Multinomial-Processing-Tree Modeling. *Behavior Research Methods* **2017**.

99. Bürkner, P.C. brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software* **2017**, *80*, 1–28.