# Systematic Quantitative Parameter Reviews in Cognitive Modeling: Towards Robust and Cumulative Models of Psychological Processes

N.-Han Tran[1*], Leendert van Maanen[3], Andrew Heathcote[4], Dora Matzke[2]

[1]Department of Human Behavioural Ecology, Max Planck Institute for Evolutionary Anthropology

[2]Department of Psychological Methods, University of Amsterdam

[3]Department of Experimental Psychology, Utrecht University

[4]Department of Psychology, University of Tasmania

## Author Note

*Corresponding author: N.-Han Tran, Department of Human Behavior, Ecology and Culture, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, han_tran@eva.mpg.de

# Systematic Quantitative Parameter Reviews in Cognitive Modeling: Towards Robust and Cumulative Models of Psychological Processes


## Abstract

Parametric cognitive models are increasingly popular tools for analysing data obtained from psychological experiments. One of the main goals of such models is to formalize psychological theories using parameters that represent distinct psychological processes. We argue that systematic quantitative reviews of parameter estimates can make an important contribution to robust and cumulative cognitive modeling. Parameter reviews can benefit model development and model assessment by providing valuable information about the expected parameter space, and can facilitate the more efficient design of experiments. Importantly, parameter reviews provide crucial—if not indispensable—information for the specification of informative prior distributions in Bayesian cognitive modeling. From the Bayesian perspective, prior distributions are an integral part of a model, reflecting cumulative theoretical knowledge about plausible values of the model's parameters (Lee, 2018). In this paper we illustrate how systematic parameter reviews can be implemented to generate informed prior distributions for the Diffusion Decision Model (DDM; Ratcliff & McKoon, 2008), the most widely used model of speeded decision making. We surveyed the published literature on empirical applications of the DDM, extracted the reported parameter estimates, and synthesized this information in the form of prior distributions. Our parameter review establishes a comprehensive reference resource for plausible DDM parameter values in various experimental paradigms that can guide future applications of the model. Based on the challenges we faced during the parameter review, we formulate a set of general and DDM-specific suggestions aiming to increase reproducibility and the information gained from the review process.


*Keywords: Bayesian inference, Cognitive Modeling, Cumulative Science, Diffusion Decision Model, Prior Distributions*

## Introduction

With an expanding recent appreciation of the value of quantitative theories that make clear and testable predictions (Lee & Wagenmakers, 2014; Navarro, in press; Oberauer & Lewandowsky, 2019), cognitive models have become increasingly popular. As a consequence, open science and reproducibility reforms have been expanded to include modeling problems. In light of this, Lee et al. (2019) proposed a suite of methods for robust modeling practices largely centred on the pre- and postregistration of models. In the interest of cumulative science, we believe that the development and assessment of cognitive models should also include systematic quantitative reviews of the model parameters. Several model classes, including multinomial processing trees (Riefer & Batchelder, 1988), reinforcement learning models (Busemeyer & Stout, 2002), and evidence-accumulation models (Donkin & Brown, 2018), have now been applied widely enough that sufficient information is available in the literature to arrive at a reliable representation of the distribution of the parameter estimates. In this paper we describe a systematic parameter review focusing on the latter class of models.

A systematic quantitative characterization of model parameters provides knowledge of the likely values of the model parameters and has various benefits. First, it can promote more precise and realistic simulations that help to optimally calibrate and design experiments, avoiding unnecessary experimental costs (Gluth & Jarecki, 2019; Heck & Erdfelder, 2019; Kennedy, Simpson, & Gelman, 2019; Pitt & Myung, 2019; Schad, Betancourt, & Vasishth, 2020). Second, knowledge about the parameter space can be crucial in maximum-likelihood estimation where an informed guess of the starting point of optimization is often key to finding the globally best solution (Myung, 2003). Third—and most important for the present paper—systematic quantitative parameter reviews provide crucial information for the specification of informative prior distributions in Bayesian cognitive modeling.

The *prior distribution* is a key element of Bayesian inference; it provides a quantitative summary of the likely values of the model parameters in the form of a probability distribution. The prior distribution is combined with the incoming data through the likelihood function to form the *posterior distribution*. The prior distribution is an integral part of Bayesian models, and should reflect theoretical assumptions and cumulative knowledge about the relative plausibility of the different parameter values (Lee, 2018; Vanpaemel, 2011; Vanpaemel & Lee, 2012). Prior distributions and prior predictive simulations play a role both in parameter estimation and model selection. Informative prior distributions can improve parameter estimation by assigning relatively more weight to plausible regions of the parameter space. Informative priors are crucial for Bayesian model selection as priors have a strong and lasting effect on Bayes factors. Unfortunately, the theoretical and practical advantages of the prior have been undermined by the common use of vague priors (Gill, 2014; Trafimow, 2005).

The goal of this paper is to illustrate how a systematic quantitative parameter review can facilitate the specification of informative prior distributions. To this end, we first introduce the Diffusion Decision Model (DDM; Ratcliff, 1978; Ratcliff & McKoon, 2008), a popular cognitive model for two-choice response time tasks (see Ratcliff, Smith, Brown, & McKoon, 2016, for a recent review). Using the DDM as a case study, we will then outline how we used a systematic literature review in combination with principled data synthesis and data quantification using distribution functions to construct informative prior distributions. Lastly, based on the challenges we faced during the parameter review, we formulate a set of general and DDM-specific suggestions about how to report cognitive modeling results, and discuss the limitations of our methods and future directions to improve them.

## Case Study: The Diffusion Decision Model

In experimental psychology, inferences about latent cognitive processes from two-choice response time (RT) tasks are traditionally based on separate analyses of mean RT and the proportion of correct responses. However, these measures are inherently related to each other in a speed-accuracy trade-off. That is, individuals can respond faster at the expense of making more errors. Evidence-accumulation models of choice RT and accuracy have provided a solution for this conundrum because they allow for the decomposition of speed-accuracy trade-off effects into latent variables that underlie performance (Donkin, Averell, Brown, & Heathcote, 2009; Ratcliff & Rouder, 1998; van Maanen et al., 2019). In these models, evidence is first extracted from the stimuli and then accumulated over time until a decision boundary is reached and a response initiated. Among the many evidence-accumulation models, the DDM is the most widely applied, not only in psychology, but also in economics and neuroscience, accounting for experiments ranging from decision making under time-pressure (Dutilh, Krypotos, & Wagenmakers, 2011; Leite, Ratcliff, Lette, & Ratcliff, 2010; Voss, Rothermund, & Brandtstädter, 2008), prospective memory (Ball & Aschenbrenner, 2018; Horn, Bayen, & Smith, 2011) to cognitive control (Gomez, Ratcliff, & Perea, 2007; Schmitz & Voss, 2012).

Figure 1 illustrates the DDM. Evidence (i.e., grey line) fluctuates from moment to moment according to a Gaussian distribution with standard deviation $s$, drifting until it reaches one of two boundaries, initiating an associated response. The DDM decomposes decision making in terms of four main parameters corresponding to distinct cognitive processes: (1) the mean rate of evidence accumulation (drift rate $v$), representing subject ability and stimulus difficulty; (2) the separation of the two response boundaries ($a$), representing response caution; (3) the mean starting point of evidence accumulation ($z$), representing response bias; and (4) mean non-decision time ($T_{er}$), which is the sum of times for stimulus encoding and response execution. RT is the sum of non-decision time and the time to diffuse from the starting point to one of the boundaries. A higher drift rate leads to faster and more accurate responses. However, responses can also be faster because a participant chooses to decrease their boundary separation, which will reduce RT but increase errors, causing the speed-accuracy trade-off. Starting accumulation closer to one boundary than the other creates a bias towards the corresponding response. Starting points $z$ is therefore most easily interpreted in relation to boundary separation $a$, where the relative starting point, also known as bias, is given by $z_r = \frac{z}{a}$.

Drift rate can vary from trial to trial according to a Gaussian distribution with

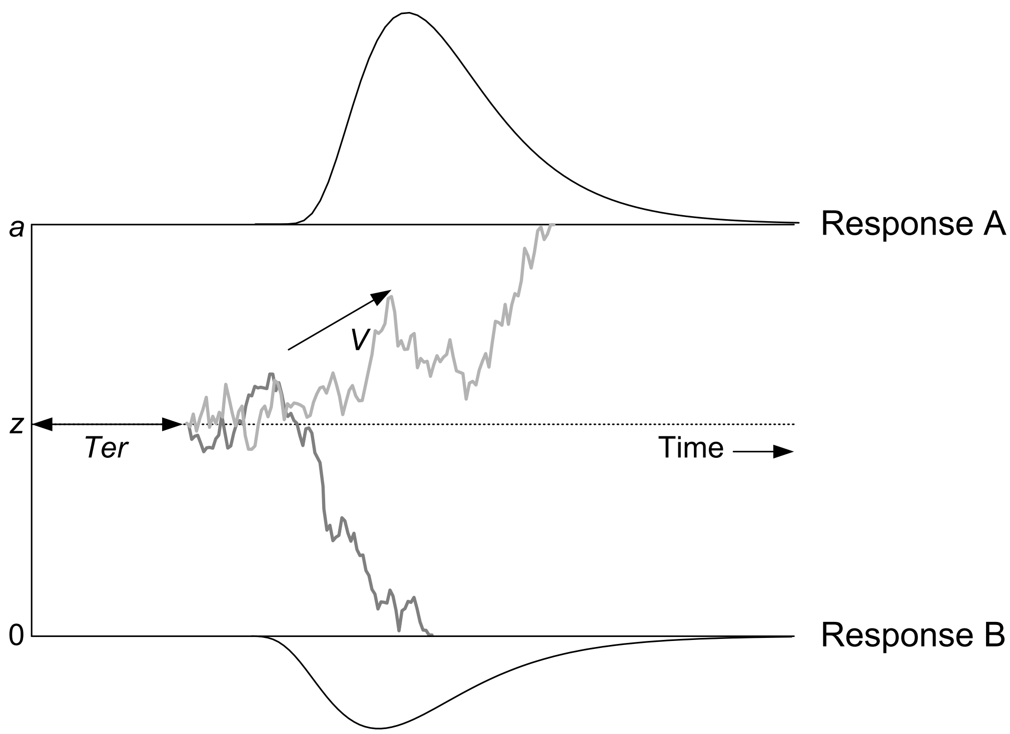


*Figure 1*. The Diffusion Decision Model (DDM; taken with permission from Matzke & Wagenmakers, 2009). The DDM assumes that noisy information is accumulated over time from a starting point until it crosses one of the two response boundaries and triggers the corresponding response. The grey line depicts the noisy decision process. 'Response A' or 'Response B' is triggered when the corresponding boundary is crossed. The DDM assumes the following main parameters: drift rate ($v$), boundary separation ($a$), mean starting point ($z$), and mean non-decision time ($T_{er}$). These main parameters can vary from trial to trial: across-trial variability in drift rate ($s_v$), across-trial variability in starting point ($s_z$), and across-trial variability in non-decision time ($s_{T_{er}}$). Starting point can be expressed relative to the boundary in order to quantify bias, where $z_r = \frac{z}{a} = 0.5$ indicates unbiased responding. Similarly, across-trial variability in starting point can be expressed relative to the boundary: $s_{z_r} = \frac{s_z}{a}$.

standard deviation $s_v$. Both non-decision time and starting point are assumed to be uniformly distributed across trials, with range $s_{T_{er}}$ and $s_z$, respectively, where $s_z$ can be expressed relative to $a$: $s_z = \frac{s_z}{a}$. A parameter of the accumulation process needs to be fixed to establish a scale that makes the other accumulation-related parameters identifiable (Donkin, Brown, & Heathcote, 2009). Most commonly this scaling parameters is the moment-to-moment variability of drift rate ($s$), usually with a value fixed to 0.1 or 1.

The growing popularity of cognitive modeling has led to extensive application of the DDM to empirical data (Theisen, Lerche, von Krause, & Voss, 2020), providing us with a large number of parameter estimates to use for constructing informative prior distributions. In 2009, Matzke and Wagenmakers presented the first quantitative summary of the DDM parameters based on a survey of parameter estimates found in 23 applications. However, their survey is now outdated and was not as extensive or systematic as the approach taken here.

# Material and Methods

All analyses were written in R or R Markdown (Allaire et al., 2018; R Core Team, 2020). The extraced parameter estimates and the analysis code are available on GitHub (`http://github.com/nhtran93/DDM_priors`) and the project's Open Science Framework (OSF) site: `https://osf.io/9ycu5/`.

## Literature Search

The literature search was conducted according to the PRISMA guidelines (Moher, Liberati, Tetzlaff, Altman, & Group, 2009). Every step was recorded and the inclusion as well as rejection of studies adhered strictly to the pre-specified inclusion criteria. Results from different search engines were exported as BibTex files, maintained with reference management software and exported into separate Microsoft Excel spreadsheets.

**Search Queries.** The literature search was commenced and completed in December 2017. It consisted of cited reference searches and independent searches according to pre-specified queries. Searches in all databases were preformed three times in order to ensure reproducibility. Four electronic databases were searched with pre-specified queries: Pubmed (PubMed, 2017), PsycInfo (American Psychological

Association, 2017), Web of Science (WoS, 2017), and Scopus (Elsevier, 2017). A preliminary search of the four databases served to identify relevant search strings, which were different for each database (see Appendix or `https://osf.io/9ycu5/` for details. The searches began from the publication year of Ratcliff (1978) seminal paper. The cited reference searches were based on Ratcliff and McKoon (2008), Wiecki, Sofer, and Frank (2013), and Palmer, Huk, and Shadlen (2005), and were performed in both Scopus and Web of Science. These key DDM papers were selected to circumvent assessing an unfeasible number of over 3000 cited references to the seminal Ratcliff (1978) article, with a potentially high number of false positives (in terms of yielding papers that reported parameter estimates), while still maintaining a wide search covering various areas of psychology and cognitive neuroscience.

**Inclusion and Exclusion Criteria.** All duplicated references were excluded. After obviously irrelevant papers —judged based on title and abstract— were excluded, the full-texts were acquired to determine the inclusion or exclusion of the remaining articles. Articles were included in the literature review if they (i) used the standard DDM according to Ratcliff (1978) and Ratcliff and McKoon (2008) with or without across-trial variability parameters; and (ii) reported parameter estimates based on empirical data from humans. Articles were excluded if (i) they reported reviews; and (ii) the parameter estimates were based on animal or simulation studies. We also excluded articles that did not report parameter estimates (neither in tables nor in graphs) and articles that estimated parameters in the context of a regression model with continuous predictors that resulted in estimates of intercepts and regression slopes instead of single values of the model parameters.

**Data Extraction.** The data extraction spreadsheet was pilot-tested using six articles and adjusted accordingly. The following parameter estimates were extracted: drift rate ($v$), boundary separation ($a$), starting point ($z$) or bias ($z_r = \frac{z}{a}$), non-decision time ($T_{er}$), across-trial variability in drift rate ($s_v$), across-trial range in starting point ($s_z$) or relative starting point ($z_r = \frac{z}{a}$), and across-trial range in non-decision time ($s_{T_{er}}$). Parameter estimates were obtained from tables as well as from graphs using the

GraphClick software (Arizona, 2010). Whenever possible, we extracted parameter estimates for each individual participant; otherwise we extracted the mean across participants. When the DDM was fit multiple times with varying parameterizations to the same data within one article, we used the estimates corresponding to the model identified as best by the authors, with a preference for selections made based on the AIC (Akaike, 1973, 1974), in order to identify the best trade-off between goodness-of-fit and parametric complexity (Myung & Pitt, 1997). When the DDM was applied to the same data across different articles, we extracted the parameter estimates from the first application; if the first application did not report parameter estimates, we used the most recent application that reported parameter estimates. Finally, articles that obtained estimates using the EZ (Wagenmakers, Van Der Maas, & Grasman, 2007) or EZ2 (Grasman, Wagenmakers, & van der Maas, 2009) methods, or the `RWiener` R package (Wabersich & Vandekerckhove, 2014), which all fit the simple diffusion estimating only the four main DDM parameters (Stone, 1960), were excluded due to concerns about potential distortions caused by ignoring across-trial parameter variability (Ratcliff, 2008). Note that we did not automatically exclude all articles without across-trial variability parameters. For articles that did not use EZ, EZ2, or `RWiener`, but reported models without across-trial variability parameters, we assumed that the author's choice of fixing these parameters to zero was motivated by substantive or statistical reasons and not by the limitations of the estimation software, and hence we included them in the parameter review.

## Parameter Transformations

Once extracted, parameter estimates had to be transformed in a way that makes aggregation across articles meaningful. In this section we report issues that arose with respect to these transformations and the solutions that we implemented. A detailed explanation of the transformations can be found in the Appendix.

**Within-Trial Variability of Drift Rate.** In all of the studies we examined, the accumulation-related parameters were scaled relative to a fixed value of the moment-to-moment variability in drift rate (typically $s = 0.1$ or $s = 1$). This decision influences the magnitude of all parameter estimates except those related to non-decision time. Once we determined $s$ for each article, we re-scaled the affected parameter estimates to $s = 1$. Articles that used the DMAT software (Vandekerckhove & Tuerlinckx, 2008) for parameter estimation were assumed to use the DMAT default of $s = 0.1$, and articles that used HDDM (Wiecki et al., 2013) or fast-DM (Voss & Voss, 2007) were assumed to use the default setting of 1. Articles (co-) authored by Roger Ratcliff were assigned $s = 0.1$.[1] We excluded 25 articles because the value of $s$ could not be determined.

**Measurement (RT) Scale.** Although the measurement (i.e., RT) scale influences the magnitude of the parameter estimates, none of the articles mentioned explicitly whether the data were fit on the seconds or milliseconds scale. Moreover, researchers did not necessarily report all estimates on the same RT scale. For instance, $T_{er}$ or $s_{T_{er}}$ were sometimes reported in milliseconds, whereas the other parameters were reported in seconds. Whenever possible, we used axis labels, captions and descriptions in figures and tables, or the default setting of the estimation software to determine the RT scale. Articles that used the DMAT, HDDM, or fast-DM were assumed to use the default setting of seconds and we assigned an RT scale of seconds to papers authored by Roger Ratcliff [2] even if $T_{er}$ was reported in milliseconds. We also evaluated the plausibility of the reported estimates with respect to the second or millisecond scale by computing a rough estimate of the expected RT for each experimental condition as $E(RT) = (a - z)/v$. We then used the following two-step decision rules to determine the RT scale of each parameter:

1. Determine the RT scale of $T_{er}$: If estimated $T_{er}$ was smaller than 5, we assumed that $T_{er}$ was reported in seconds; otherwise we assumed that $T_{er}$ was reported in milliseconds.

2. Determine RT scale of remaining parameters: If $E(RT)$ was smaller than 10, we

[1] Based on personal communication with Roger Ratcliff.

[2] Based on personal communication with Roger Ratcliff.

assumed that the remaining parameters were reported in seconds; otherwise we assumed that the remaining parameters were reported in milliseconds.

Once we determined the RT scale for each parameter, we re-scaled the parameter estimates to the seconds scale. Individual parameters estimates that were considered implausible after the transformation (i.e., outside of the parameter bounds, such as a negative $a$) were checked manually. In particular, we checked for 1) inconsistencies in the magnitude across the parameter estimates within articles (e.g., a value of $a$ indicative of seconds vs. a value of $T_{er}$ indicative of milliseconds); 2) reporting or typographic errors; 3) extraction errors; and 4) errors in determining the measurement scale, which typically reflected the use of non-standard experiments or special populations. In a number of cases we also revisited and whenever necessary reconsidered the assigned value of $s$. We removed all parameter estimates from 13 articles that reported implausible estimates reflecting ambiguous or inconsistent RT scale descriptions or clear reporting errors.

**Starting Point and Bias.** We expressed all starting point $z$ and starting point variability $s_z$ estimates relative to $a$. As the attributions of the response options to the two response boundaries is arbitrary, the direction of the bias (i.e., whether $z_r$ is greater or less than 0.5) is arbitrary. As these attributions cannot be made commensurate over articles with different response options, values of $z_r$ cannot be meaningfully aggregated over articles. As a consequence, bias, $z_r$, and its complement, $1 - z_r$, are exchangeable for the purpose of our summary. We therefore used both values in order to create a single "mirrored" distribution. This distribution is necessarily symmetric with a mean of 0.5, but retains information about variability in bias.[3]

**Parameter Constraints.** In many applications of the DDM, researchers impose constraints on the parameter estimates across experimental manipulations, conditions, or groups, either based on theoretical grounds or the results of

[3] The bias $z_r$ parameters estimated using the HDDM software (Wiecki et al., 2013) are coded as $1 - z_r$ in our parameter review. Note that this has no influence on the resulting prior distribution as we used both $z_r$ and $1 - z_r$ to create the prior.

model-selection procedures. We only considered unique parameter estimates and did not repeatedly include estimates in our parameter review that were fixed across manipulations, conditions, or groups within the same study in the same article.

## Generating Informed Prior Distributions for the DDM Parameters

After post-processing and transforming the parameter estimates we collapsed each parameter type separately across studies and articles into single distributions. We then attempted to characterize the aggregated results using a range of univariate distribution functions that respected the parameter type's bounds (e.g., non-decision time $T_{er}$ must be positive) and provided the best fit to the overall shape of the empirical distributions. We first considered (truncated) normal, lognormal, gamma, Weibull, and (truncated) Student's t distribution functions. However, in some cases the empirical distributions were clearly multi-modal and were contaminated by outliers due to non-standard tasks, special populations, and possible reporting errors that we not identified during the post-processing steps. We therefore also considered characterizing the aggregated data using mixture distributions. Mixtures were chosen from the exponential family of distributions that respected the theoretical bounds of the parameter estimates. In particular we used mixtures of two gamma distributions, and (truncated) normals mixed with either a gamma, lognormal, or another (truncated) normal distribution. Specifically, we focused on normal mixtures because we assume a finite variance for the parameters and thus the Gaussian distributions represents the most conservative probability distribution to assign to the parameter distributions (for further information see the principles of maximum entropy; Jaynes, 1988).

The univariate and mixture distributions were fit to the empirical distributions using maximum-likelihood estimation (Myung, 2003), in most cases with additional constraints on upper and/or lower bounds. For (mirrored) bias $z_r$ and $s_{z_r}$, which are bounded between 0 and 1, we used univariate truncated normal and truncated $t$ distributions on $[0, 1]$. A lower bound of zero was imposed on all remaining parameters except drift rate $v$ which was unbounded. We then used AIC weights (wAIC;

Wagenmakers & Farrell, 2004) to select the theoretical distributions that struck the best balance between goodness-of-fit and simplicity. A table of the AIC and wAIC values for all fitted univariate and mixture distributions and the code to reproduce this table, can be found in the open repository on GitHub or the OSF.

We propose that the wAIC-selected distributions can be used as informative prior distributions for the Bayesian estimation of the DDM parameters. For simplicity, for parameters where a mixture was the best-fitting distribution, we propose as prior the distribution component that best captures the bulk of the parameter estimates as indicated by the highest mixture weight. We will revisit this choice in the Discussion.

## Results

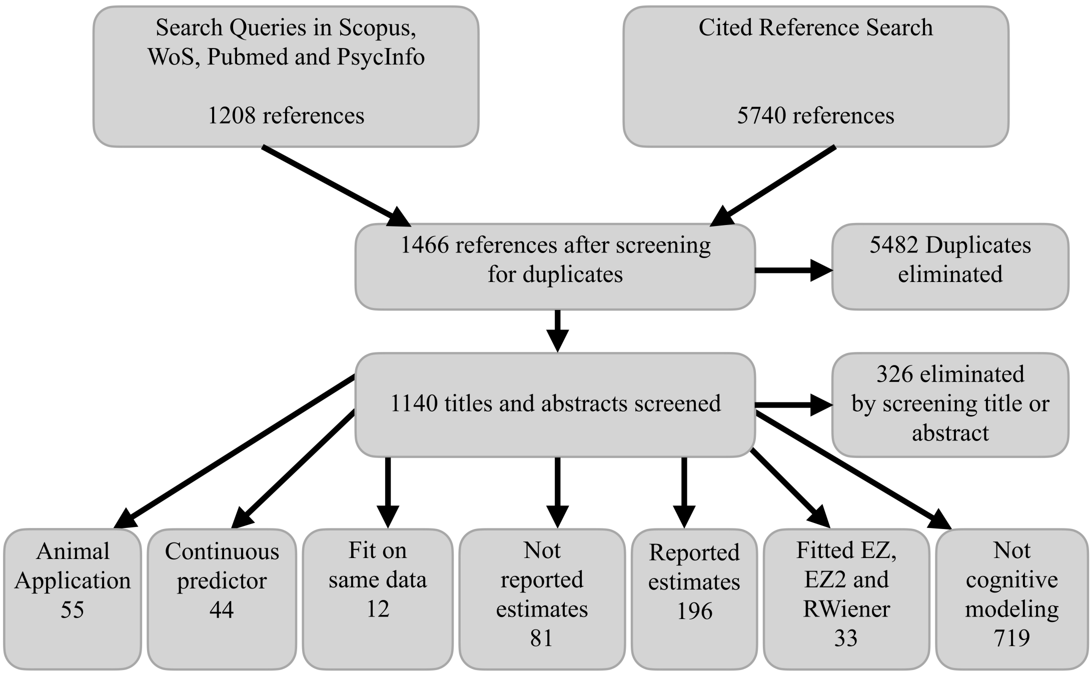


*Figure 2*. *PRISMA flow diagram.* WoS = Web of Science. `RWiener` refers to the R package from Wabersich and Vandekerckhove (2014). EZ and EZ2 refer to estimation methods for the simple DDM developed by Wagenmakers et al. (2007) and Grasman et al. (2009), respectively.

Figure 2 shows the PRISMA flow diagram corresponding to our literature search. The total of 196 relevant articles (i.e., "Reported estimates" in Figure 2) covered a wide range of research areas from psychology and neuroscience to medicine and economics. We excluded 38 references because they did not report the fixed value of the scaling parameter $s$ and we were unable to reverse engineer the values or because of inconsistent RT scale descriptions or clear reporting errors. Thus, we extracted parameter estimates from a total of 158 references. The most common paradigms were various perceptual decision-making tasks (e.g., random dot motion task; 37 references), lexical decision tasks (25), and recognition memory tasks (10). A total of 29 references included clinical groups and 26 references used Bayesian estimation methods.

The histograms in Figure 3 show the empirical distributions of the parameter estimates and the red lines show the theoretical distributions component with the highest mixture weight (i.e., the informative prior distributions). Table 1 gives an overview of the informative prior distributions and the corresponding upper and lower bounds (see column "T-LB" and "T-UB"). The table also shows the upper and lower bounds of the parameter estimates collected from the literature (see column "E-LB" and "E-UB"); these bounds can be used to further constrain parameter estimation by providing bounds for prior distributions and bounded optimization methods.

The results of the model comparisons are available at `https://osf.io/9ycu5/`. For drift rate $v$, the selected model was a $t$ distribution (wAIC $= 1.0$), with degrees of freedom, and location and scale parameters shown in the first row of Table 1. For boundary separation $a$, the selected model was a mixture of gamma distributions (wAIC $= 0.76$), with the shape and scale parameters of the dominant gamma component shown in the second row of Table 1. For non-decision time $T_{er}$ and the across-trial variability in non-decision time $s_{T_{er}}$, the selected model was a zero-bounded truncated $t$ distribution (wAIC $= 1$ for both $T_{er}$ and $s_{T_{er}}$). For mirrored bias $z_r$, the selected model was a truncated $t$ distribution on $[0, 1]$ (wAIC $= 1.0$). For across-trial variability in drift rate $s_v$, the selected model was a mixture of a gamma and a zero-bounded truncated normal distribution (wAIC $= 0.35$), where the truncated

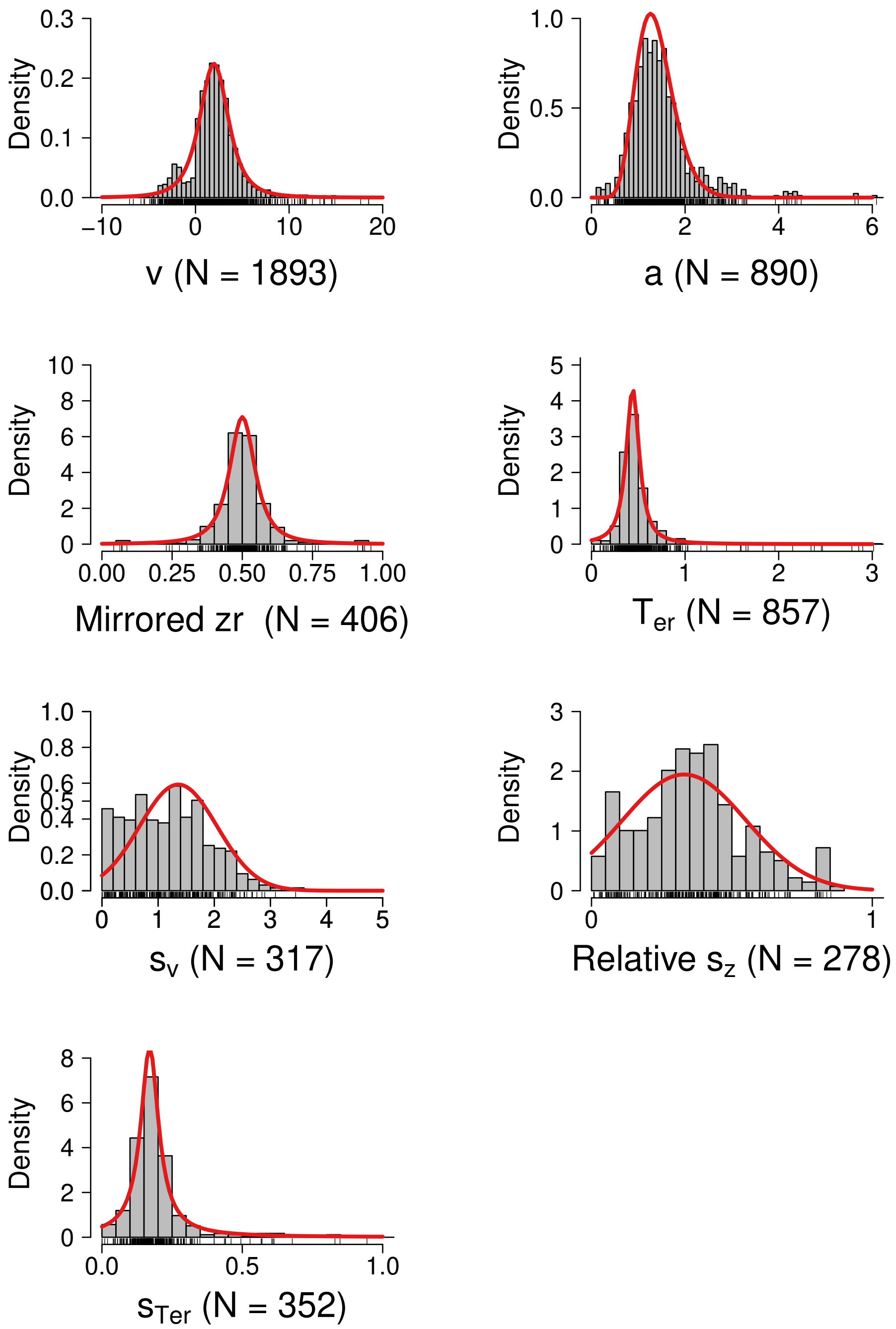


*Figure 3*. *Empirical distributions of the DDM parameter estimates retrieved from the literature.* The red lines represent the theoretical distributions that best capture the empirical distributions (i.e., informative prior distributions). The red lines only show the distribution component with the highest mixture weight. *N*: number of unique estimates.

normal had the highest mixture weight. Lastly, for $s_{z_r}$, the selected model was a truncated normal distribution on $[0, 1]$ (wAIC = 0.74).

Table 1

*Informative Prior Distributions*

| DDM Parameter | N | (Dominant) Distribution | Location/Shape | Scale | df | T-LB | T-UB | E-LB | E-UB |
|---|---|---|---|---|---|---|---|---|---|
| $v$ | 1893 | t | 1.97 | 1.63 | 2.80 | - Inf | + Inf | -26.90 | 18.51 |
| $a$ | 890 | gamma | 11.69 | 0.12 | | 0 | + Inf | 0.11 | 7.47 |
| Mirrored $z_r$ | 406 | truncated t | 0.5 | 0.05 | 1.85 | 0 | 1 | 0.04 | 0.96 |
| $T_{er}$ | 857 | truncated t | 0.44 | 0.08 | 1.32 | 0 | + Inf | 0 | 3.69 |
| $s_v$ | 317 | truncated normal | 1.36 | 0.69 | | 0 | + Inf | 0 | 3.45 |
| $s_{z_r}$ | 278 | truncated normal | 0.33 | 0.22 | | 0 | 1 | 0.01 | 0.85 |
| $s_{T_{er}}$ | 352 | truncated t | 0.17 | 0.04 | 0.88 | 0 | + Inf | 0 | 4.75 |

*Note.* N: The number of unique estimates; df: degrees of freedom; T-LB: theoretical lower bound of the prior distribution; T-UB: theoretical upper bound of the prior distribution.; E-LB: lower bound of the empirical parameter estimates; E-UB: upper bound of the empirical parameter estimates.

## Discussion

The increasing popularity of cognitive modeling has led to extensive applications of models like the Diffusion Decision Model (DDM) across a range of disciplines. These applications have the potential to provide substantial information about the plausible values of parameter estimates in cognitive models. We believe that for cognitive models where sufficient information are available in the literature, a systematic quantitative characterization of model parameters should be part of modeling practices. Parameter reviews can benefit modeling practices in various ways, from facilitating parameter estimation to enabling more precise and realistic simulations to improve study design and calibrate future experiments (Gluth & Jarecki, 2019; Heck & Erdfelder, 2019; Pitt & Myung, 2019). Here, we used the DDM as example case of how a systematic quantitative parameter review can be incorporated into modeling practices to provide informative prior distributions for the model parameters. Our empirical distributions of the parameter estimates were largely consistent with those of Matzke and Wagenmakers (2009), but because our sample was much larger we were better able to capture the tails of the parameter distributions.

Inferring the parameters of complex cognitive models like the DDM from experimental data is challenging because their parameters are often highly correlated. The cumulative knowledge distilled into parameter estimates from past research can practically benefit both traditional optimization-based methods (e.g., maximum likelihood) and Bayesian estimation. In the former case, parameter reviews can provide informed guesses for optimization starting points as well as guidance for configuring bounded optimization methods. Even when powerful and robust optimization algorithms (e.g., particle swarm methods) are used, reasonable initial values and bounds can increase time efficiency and are often helpful for avoiding false convergence on sub-optimal solutions. In the latter —Bayesian case—parameter reviews can facilitate the use of informative prior distributions, which benefits both Bayesian model selection and parameter estimation.

Informative priors are absolutely crucial for Bayesian model selection using Bayes factors (Jeffreys, 1961; Kass & Raftery, 1995). Unlike other model-selection methods like the Deviance Information Criteria (DIC; Spiegelhalter, Best, Carlin, & van der Linde, 2014) and Widely Applicable Information Criterion (WAIC; Vehtari, Gelman, & Gabry, 2017) that depend only on posterior samples, Bayes factors depend crucially on the prior distribution even when large amounts of data are available. This is because the marginal likelihood of the competing models is obtained by taking a weighted average of the probability of the data across all possible parameter settings with the weights given by the parameters' prior density. The workflow outlined here may therefore facilitate the more principled use of prior information in Bayesian model selection in the context of evidence-accumulation models (for recent developments, see Evans & Annis, 2019; Gronau, Heathcote, & Matzke, 2020).

In terms of Bayesian estimation, the extra constraint provided by informative priors can benefit some parameters more than others. In the DDM, for example, the across-trial variability parameters are notoriously difficult to estimate (Boehm et al., 2018; Dutilh et al., 2019). This has led to calls for these parameters to be fixed to zero (i.e., use the simple diffusion model; Stone, 1960) to improve the detection of effects on the remaining parameters (van Ravenzwaaij, Donkin, & Vandekerckhove, 2017). Informative priors may provide an alternative solution that avoids the potential systematic distortion caused by ignoring the variability parameters (Ratcliff & McKoon, 2008) and enables the study of effects that cannot be accommodated by the simple diffusion model, such as differences between correct and error RTs (Damaso, Williams, & Heathcote, submitted). Although, for simplicity, here we suggested single-component distributions as priors, there is no reason in principle that the full mixture distributions that we selected could not be used. In practice, Bayesian DDM software, such as the Dynamic Models of Choice software (DMC; Heathcote et al., 2019), can be easily adapted to use any form of univariate prior, and so better capture the tails of the empirical parameter distributions.

Information about the empirical distribution of parameter estimates, both in terms of the main body and the tails of the distributions, can especially benefit design optimization and parameter estimation in non-standard and difficult to access populations (e.g., Matzke, Hughes, Badcock, Michie, & Heathcote, 2017; Shankle et al., 2013). For example, in clinical populations long experimental sessions are often impossible due to exhaustion or attention lapses. Expenses can also be constraining, such as with studies using costly fMRI methods. Therefore, data are often scarce, with a total number of trials as low as 100 reported in some DDM applications (e.g., O'Callaghan et al., 2017). In these cases, experimental designs can be optimized, and parameter estimation improved, with the aid of informative parameter distributions that put weight on plausible parts of the parameter space. Moreover, informative priors can also increase sampling efficiency and speed up the convergence of MCMC routines.

However, systematic quantitative parameter reviews have their pitfalls. Using available cumulative knowledge from past literature always has to be viewed in light of the file drawer problem (Rosenthal, 1979). Many researchers have not published their non-significant results, therefore the literature is biased, and thus the parameter estimates retrieved from the literature might be biased towards specific model settings that converged or led to significant results. Furthermore, some cognitive models are too new and have not been widely applied to empirical data, so past literature might not provide researchers with a sufficiently reliable representation of the distribution of the parameter estimates. Therefore, cognitive modelers may not always be able to incorporate our proposed quantitative parameter review into their workflow, and should carefully weigh out the feasibility and benefits of such an endeavour.

## Recommendations for Reporting Cognitive Modeling Results

Our literature review revealed a wide variety of reporting practices, both in terms of *what* researcher report and *how* they report their modeling results. The diversity of reporting practices is likely to reflect differences between disciplines and is in itself not problematic. However, we believe that the full potential of cumulative science can only be realized if authors provide sufficient information for others to interpret and reproduce their results. We endorse code and data sharing, and —following Lee et al. (2019)— we strongly urge researchers to provide sufficiently precise mathematical and statistical descriptions of their models, and to post-register exploratory model developments. In what follows, we reflect on the challenges we faced in performing the systematic parameter review, and formulate a set of general and DDM-specific suggestions that aim to increase computational reproducibility and the expected information gain from parameter reviews. Although our recommendations are certainly not exhaustive and do not apply to all model classes, we hope that they provide food for thought for cognitive modelers in general and RT modelers in particular.

**Model Parameterization and Scaling.** The following recommendations are aimed at supporting well-informed choices about which model and which model parameters to include in a parameter review. Most parametric cognitive models can be parameterized in various ways. First, some cognitive models require fixing one (or more) parameters to make the model identifiable (Donkin, Brown, & Heathcote, 2009). In the

DDM, modelers typically fix the moment-to-moment variability of drift rate $s$ to 0.1 or 1 for scaling purposes. Note, however, that the exact value of the scaling parameter is arbitrary, and —depending on the application— one may chose to estimate $s$ from the data and use other parameters for scaling. We stress the importance of explicitly reporting which parameters are used for scaling purposes and the value of the scaling parameter(s) because the chosen setting can influence the magnitude of the other parameter estimates. Another scaling issue relates to the measurement units of the data. For example, RTs are commonly measured in both seconds and milliseconds. Although the measurement scale influences the magnitude of the parameter estimates, none of the articles included in the present parameter review explicitly reported the measurement unit of their data. Further, articles did not consistently reported all parameter estimates on the same RT scale (i.e., all parameter estimates reported in seconds, but $T_{er}$ reported in milliseconds). Hence, we urge researchers to make an explicit statement on this matter and whenever possible stick to the same measurement unit throughout an article to avoid any ambiguity. Second, in cognitive models one parameter is sometimes expressed as a function of one or more other parameters. The DDM, for instance, can be parameterized in terms of *absolute* starting point $z$ or *relative* starting point $z_r = \frac{z}{a}$ (i.e., bias). The choice between $z$ and $z_r$ depends on the application but can also reflect default software settings. Although the two parametrizations are mathematically identical and have no consequences for the magnitude of the other parameters, it is clearly important to communicate which parameterization is used in a given application. Third, in many applications, researchers impose constraints on the model parameters across experimental manipulations, conditions, or groups. Such constraints sometimes reflect practical or computational considerations, but preferably they are based on a priori theoretical rationale (e.g., threshold parameters cannot vary based on stimulus properties that are unknown before a trial commences; Donkin, Averell, et al., 2009) or the results of model-selection procedures (e.g., Heathcote, Loft, & Remington, 2015; Strickland, Loft, Remington, & Heathcote, 2018). Regardless of the specific reasons for parameter constraints, we urge modelers to clearly communicate which parameters are hypothesized to reflect the effect(s) of interest, and so which are fixed and which are free to vary across the design. Moreover, we recommend researchers to report the competing models (including the parametrization) that were entertained to explain the data, and indicate the grounds on which a given model was chosen as best, such as AIC (Akaike, 1981), BIC (Schwarz, 1978), DIC (Spiegelhalter, Best, Carlin, & Van Der Linde, 2002; Spiegelhalter et al., 2014), WAIC (Watanabe, 2010), or Bayes factors (Kass & Raftery, 1995). We note that parameter reviews are also compatible with cases where there is uncertainty about which is the best model, through the use of Bayesian model averaging (Hoeting, Madigan, Raftery, & Volinsky, 1999). In this approach, the parameter estimates used in the review are averaged across the models in which they occur, weighted by the posterior probability of the models.

**Model Estimation.** In the face of the large number of computational tools available to implement cognitive models and the associated complex analysis pipelines, researchers have numerous choices on how to estimate model parameters. For instance, a variety of DDM software is available, such as fast-DM (Voss & Voss, 2007), HDDM (Wiecki et al., 2013), DMC (Heathcote et al., 2019), DMAT (Vandekerckhove & Tuerlinckx, 2008), using a variety of estimation methods, such as maximum likelihood, Kolmogorov-Smirnov, chi-squared minimization (Voss & Voss, 2007), quantile maximum probability (Heathcote & Brown, 2004), or Bayesian Markov chain Monte Carlo (MCMC; e.g., Turner, Sederberg, Brown, & Steyvers, 2013) techniques. We encourage researchers to report the software they used, and whenever possible, share their commented code to enable computational reproducibility(Cohen-Boulakia et al., 2017; McDougal, Bulanova, & Lytton, 2016). Knowledge about the estimation software can also provide valuable information about the parametrization and scaling issues described above.

**Parameter Estimates, Uncertainty and Correlations.** We recommend researchers to report all parameter estimates from their chosen model and not only the ones that are related to the experimental manipulation or the psychological effect of interest. In the DDM in particular, modelers should also report the across-trial variability parameters, and not only the main parameters (i.e., drift rate, boundary separation, starting point, and non-decision time), even if only a subset of parameters is the focus of the study. We also urge researchers to report measures of uncertainty associated with the parameter estimates, let these be frequentist (bootstrapped; Stine, 1989) standard errors and confidence intervals (e.g., Visser & Poessé, 2017), or Bayesian credible intervals and full posterior distributions (Eberly & Casella, 2003; Jeffreys, 1961; Lindley, 1965). Ideally, in the process of aggregation used to create prior distributions, parameter estimates should be weighted by their relative uncertainty. Unfortunately, because of the paucity of uncertainty measures in the literature we surveyed, our aggregation had to give equal weights to all parameter estimates. Ideally, supplementary materials should provide an overview of parameter estimates for each individual participant, or links to such overviews in an electronic format. In the vast majority of the studies examined here, only parameters averaged over participants were available. This means that we were unable to evaluate correlations among parameter estimates reflecting individual differences. Such correlations are likely quite marked. For example, in the DDM a participant with a higher drift rate, which promotes accuracy, is more likely to be able to afford to set a lower boundary and still maintain good performance, so a negative correlation between rates and boundaries might be expected. As we discuss below, the failure to report individual estimates brings with it important limitations on what can be achieved with the results of systematic parameter reviews.

## Limitations and Future Directions

The approach to parameter reviews taken here —obtaining values from texts, tables, and graphs from published papers— has the advantage of sampling estimates that are representative of a wide variety of laboratories, paradigms, and estimation methods. However, it also has a number of limitations beyond those related to the vagaries of incomplete reporting practices just discussed.

The first limitation is related to the aggregation of parameter estimates over different designs even when they all come from the same model. In our DDM application, the direction of response bias provides an example where the meaning of a parameter is design specific, and so it is difficult to form useful aggregates over different paradigms. To take a concrete example, a bias towards "word" responses over "non-word" responses in a lexical-decision paradigm cannot be made commensurate with a bias favoring "left" over "right" responses in a random-dot motion paradigm. Our approach —forming an aggregate with maximum uncertainty by assuming either direction is equally likely (i.e., mirroring the values)— removes any information about the average direction while at least providing some information about variability in bias. Although this approach likely overestimates the variability of the bias estimates, we believe that overestimation is preferable to underestimation which might result in an overly influential prior distribution.

Of course, no issues arise when aggregation is over designs with commensurate responses. Our online data repository reports raw starting point and bias estimates, which combined with the design descriptions from the original papers could be used to perform such an aggregation. We note, however, that in some papers it was unclear which response was mapped to which DDM boundary, so we would add a reporting guideline that this choice be spelled out clearly. We also note that similar problems with aggregation are likely to occur for other parameter types and also beyond the DDM, for instance in evidence-accumulation models such as the Linear Ballistic Accumulator (Brown & Heathcote, 2008). For instance, if one decomposes drift rates in the DDM into the average over stimuli and "stimulus bias" (i.e., the difference in rates between the two stimulus classes; White & Poldrack, 2014), then the same issue applies, but now with respect stimuli rather than responses. The second limitation —which is related to incomplete reporting, but is harder to address within a traditional journal format— concerns obtaining a full multivariate characterization of the prior distribution of parameters that takes into account correlations among parameters as well as their average values and variability. Because most estimates reported in the literature are averages over participants, we were restricted to providing separate univariate characterizations of prior distributions for each parameter. To the degree that the implicit independence assumption of this approach is violated [4] problems can arise. Continuing the example of negatively correlated rates and boundaries, although a higher value of both separately may be quite probable, both occurring together may be much less likely that the product of their individual probabilities that would be implied by independence.

Problems would arise, for example, if in planning a new experiment one were to produce synthetic data by drawing parameter combinations independently from the univariate priors in Figure 3, potentially producing simulated participants with parameter values that are unlikely in a real experiment. Or consider Bayesian methods, the problem of ignoring the correlations among parameters can compromise the efficiency of MCMC samplers and complicate the interpretation of Bayes factors because the resulting uni-variate priors will assign mass to implausible regions of the parameter space. Although standard Bayesian MCMC samplers used for evidence-accumulation models have not taken account of these population correlations, a new generation of samplers is appearing that does (Gunawan, Hawkins, Tran, Kohn, & Brown, 2020). This development underscores the need for future systematic parameter reviews to move in the direction of multivariate characterizations. This may be achieved by revisiting the original data sets, which due to open science practices are becoming increasingly available, refitting the DDM, and then using the resulting individual parameter estimates to form multivariate priors. This future direction will be time consuming and computationally challenging, and will no doubt bring with it new methodological problems that we have not addressed here. Nevertheless, we believe that the long-term gains for cognitive modeling will be worthwhile.

[4] To be clear, we are not talking about correlations among parameters within a participant, which are a consequence of the mathematical form of the model's likelihood and the particular parameterization adopted for the design. Rather, we are addressing correlations at the population level, i.e., across participants. Although the two types of correlations can be related, they are not the same and in our experience can sometimes differ very markedly.

## Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Author Contributions

NHT designed the study, collected the data, performed the analyses, and wrote the manuscript. NHT, LvM and DM designed the study and the analyses. DM and AH provided critical feedback and helped shape the manuscript.

## Funding

This work was supported by the Utrecht University and the Max Planck Institute for Evolutionary Anthropology. AH was supported by the ARC Discovery Project grant no. DP200100655. DM was supported by a Veni grant (451-15-010) from the Netherlands Organization of Scientific Research (NWO).

## Data Availability Statement

The dataset and the analyses can be found at `https://osf.io/9ycu5/` and `https://github.com/nhtran93/DDM_priors`.

# Appendix

## Search Queries for Literature Search

The following search strings were used:

- PsycInfo: ((drift diffusion model* or diffusion decision model or ratcliff diffusion or diffusion model) and (response time* or reaction time*)).mp. and Reaction time.sh. limit 1 to yr="1977 - 2017"
- Pubmed: ((((diffusion[Text Word] OR drift*[Text Word] OR *DDM[Text Word]) AND model*[Text Word]) AND ("1978"[Date - Publication] : "2017"[Date - Publication])) AND "Reaction time"[MeSH Terms])
- Web of Science: (TS=("drift diffusion model") OR TS=("diffusion decision model") OR TS=("ratcliff diffusion") OR TS=("diffusion model")) AND (TS=(response time* OR reaction time*)) AND (SU=(Life Sciences Biomedicine) OR SU=(Social Sciences)) Indexes=SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan=1977-2017
- Scopus: (ALL ( "response* time*" ) OR ALL ( "reaction* time*")) AND (TITLE-ABSKEY ( "diffusion model" ) OR TITLE-ABS-KEY ( "drift diffusion model" ) OR TITLE-ABS-KEY ( "*DDM" ) OR TITLE-ABS-KEY ( "diffusion decision model" ) OR TITLE-ABS-KEY ( "Ratcliff diffusion" )) AND PUBYEAR > 1977 AND SUBJAREA ( psyc OR neur OR medi OR soci OR deci OR econ OR mult )

## Parameter Transformations

The reported parameter estimates were re-scaled to an RT scale in seconds and moment-to-moment variability in drift rate $s$ of 1 as follows:

$$v = v \times \frac{\sqrt{\text{scaling}}}{\text{scaling}}$$
$$a = a \times \sqrt{\text{scaling}}$$
$$z = z \times \sqrt{\text{scaling}}$$
$$T_{er} = T_{er} \times \text{RT scaling}$$
$$s_v = s_v \times \frac{\sqrt{\text{scaling}}}{\text{scaling}}$$
$$s_z = s_z \times \sqrt{\text{scaling}}$$
$$s_{T_{er}} = s_{T_{er}} \times \text{RT scale},$$

where scaling = RT scaling × s scaling. For instance, suppose we want to re-scale parameter estimates from milliseconds to seconds and from $s = 0.1$ to $s = 1$. The scaling factor would be computed as scaling $= \frac{1}{1000} \times 10$. Bias $z_r = \frac{z}{a}$ and $s_{z_r} = \frac{s_z}{a}$ require no scaling. The value of non-decision time $T_{er}$ and $s_{T_{er}}$ are not affected by the value of $s$, therefore, only the RT scaling is necessary.

## References

Akaike, H. (1973). Information theory and an extension of maximum likelihood principle. In *Proceedings of the second international symposium on information theory* (pp. 267–281).

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.

Akaike, H. (1981). Likelihood of a model and information criteria. *Journal of Econometrics*, *16*(1), 3 - 14. doi: https://doi.org/10.1016/0304-4076(81)90071-3

Allaire, J., Cheng, J., Xie, Y., McPherson, J., Chang, W., Allen, J., & Hyndman, R. (2018). rmarkdown: Dynamic documents for R (R package version 1.0) [Computer software manual].

American Psychological Association. (2017). *Psycinfo.* `https://www.apa.org/pubs/databases/psycinfo/`. (Accessed: 2017-12-26)

Arizona. (2010). GraphClick [Computer software manual]. `http://www.arizona-software.ch/graphclick/`. Author. (Accessed: 2018-01-26)

Ball, B. H., & Aschenbrenner, A. J. (2018). The importance of age-related differences in prospective memory: Evidence from diffusion model analyses. *Psychonomic Bulletin & Review*, *25*(3), 1114–1122. doi: 10.3758/s13423-017-1318-4

Boehm, U., Annis, J., Frank, M. J., Hawkins, G. E., Heathcote, A., Kellen, D., . . . Wagenmakers, E.-J. (2018). Estimating across-trial variability parameters of the diffusion decision model: Expert advice and recommendations. *Journal of Mathematical Psychology*, *87*, 46 - 75. doi: https://doi.org/10.1016/j.jmp.2018.09.004

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*(3), 153–178. doi: 10.1016/j.cogpsych.2007.12.002

Busemeyer, J. R., & Stout, J. C. (2002). A contribution of cognitive decision models to clinical assessment: Decomposing performance on the Bechara gambling task.

*Psychological Assessment*, *14*(3), 253–262. doi: https://doi.org/10.1037/1040-3590.14.3.253

Cohen-Boulakia, S., Belhajjame, K., Collin, O., Chopard, J., Froidevaux, C., Gaignard, A., ... Blanchet, C. (2017). Scientific workflows for computational reproducibility in the life sciences: Status, challenges and opportunities. *Future Generation Computer Systems*, *75*, 284–298. doi: https://doi.org/10.1016/j.future.2017.01.012

Damaso, K., Williams, P., & Heathcote, A. (submitted). What does a (hu)man do after (s)he makes a fast versus slow error, and why?

Donkin, C., Averell, L., Brown, S., & Heathcote, A. (2009). Getting more from accuracy and response time data: Methods for fitting the linear ballistic accumulator. *Behavior Research Methods*, *41*(4), 1095–1110. doi: 10.3758/BRM.41.4.1095

Donkin, C., & Brown, S. D. (2018). Response times and decision-making. In E.-J. Wagenmakers & J. T. Wixted (Eds.), *Stevens' handbook of experimental psychology and cognitive neuroscience, Volume 5: Methodology (4th ed.)* (p. 349-382). John Wiley & Sons, Inc.

Donkin, C., Brown, S. D., & Heathcote, A. (2009). The overconstraint of response time models: Rethinking the scaling problem. *Psychonomic Bulletin & Review*, *16*(6), 1129–1135.

Dutilh, G., Annis, J., Brown, S. D., Cassey, P., Evans, N. J., Grasman, R. P. P. P., ... Donkin, C. (2019). The quality of response time data inference: A blinded, collaborative sssessment of the validity of cognitive models. *Psychonomic Bulletin & Review*, *26*(4), 1051–1069. doi: 10.3758/s13423-017-1417-2

Dutilh, G., Krypotos, A.-M., & Wagenmakers, E.-J. (2011). Task-related versus stimulus-specific practice. *Experimental Psychology*, *58*(6), 434–442. doi: 10.1027/1618-3169/a000111

Eberly, L. E., & Casella, G. (2003). Estimating Bayesian credible intervals. *Journal of Statistical Planning and Inference*, *112*(1), 115–132. doi: https://doi.org/10.1016/S0378-3758(02)00327-0

Elsevier. (2017). *Scopus.* `https://www.scopus.com/home.uri`. (Accessed: 2017-12-26)

Evans, N. J., & Annis, J. (2019). Thermodynamic integration via differential evolution: A method for estimating marginal likelihoods. *Behavior Research Methods*, *51*, 930–947.

Gill, J. (2014). *Bayesian methods: A social and behavioral sciences approach.* Chapman & Hall.

Gluth, S., & Jarecki, J. B. (2019). On the importance of power analyses for cognitive modeling. *Computational Brain & Behavior*, *2*(3), 266–270. doi: 10.1007/s42113-019-00039-w

Gomez, P., Ratcliff, R., & Perea, M. (2007). A model of the Go/No-Go task. *Journal of Experimental Psychology: General*, *136*, 389–413.

Grasman, R. P., Wagenmakers, E.-J., & van der Maas, H. L. (2009). On the mean and variance of response times under the diffusion model with an application to parameter estimation. *Journal of Mathematical Psychology*, *53*(2), 55 - 68. doi: https://doi.org/10.1016/j.jmp.2009.01.006

Gronau, Q. F., Heathcote, A., & Matzke, D. (2020). Computing bayes factors for evidence-accumulation models using Warp-III bridge sampling. *Behavior Research Methods*, *52*, 918-937.

Gunawan, D., Hawkins, G. E., Tran, M. N., Kohn, R., & Brown, S. D. (2020). New estimation approaches for the hierarchical Linear Ballistic Accumulator model. *Journal of Mathematical Psychology*, *96*, 102368.

Heathcote, A., & Brown, S. (2004). Reply to Speckman and Rouder: A theoretical basis for QML. *Psychonomic Bulletin & Review*, *11*(3), 577–578.

Heathcote, A., Lin, Y.-S., Reynolds, A., Strickland, L., Gretton, M., & Matzke, D. (2019, 01). Dynamic models of choice. *Behavior Research Methods*, *51*(2), 961–985. doi: 10.3758/s13428-018-1067-y

Heathcote, A., Loft, S., & Remington, R. W. (2015). Slow down and remember to remember! A delay theory of prospective memory costs. *Psychological Review*, *122*, 376–410.

Heck, D. W., & Erdfelder, E. (2019, 01). Maximizing the expected information gain of cognitive modeling via design optimization. *Computational Brain & Behavior*, *2*(3), 202–209. doi: 10.1007/s42113-019-00035-0

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, *14*, 382–401.

Horn, S. S., Bayen, U. J., & Smith, R. E. (2011). What can the diffusion model tell Us about prospective memory? *Canadian Journal of Experimental Psychology - Revue Canadienne de Psychologie Experimentale*, *65*(1, SI), 69–75. doi: 10.1037/a0022808

Jaynes, E. T. (1988). The relation of Bayesian and maximum entropy methods. In G. J. Erickson & C. R. Smith (Eds.), *Maximum entropy and bayesian methods in science and engineering* (1st ed., pp. 25–29). Kluwer Academic Publishers.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Oxford University Press.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773–795.

Kennedy, L., Simpson, D., & Gelman, A. (2019, 01). The experiment is just as important as the likelihood in understanding the prior: A cautionary note on robust cognitive modeling. *Computational Brain & Behavior*, *2*(3), 210–217. doi: 10.1007/s42113-019-00051-0

Lee, M. D. (2018). Bayesian methods in cognitive modeling. In E.-J. Wagenmakers & J. T. Wixted (Eds.), *Stevens' handbook of experimental psychology and cognitive neuroscience, Volume 5: Methodology (4th ed.)* (p. 37-84). John Wiley & Sons, Inc.

Lee, M. D., Criss, A. H., Devezer, B., Donkin, C., Etz, A., Leite, F. P., ... Vandekerckhove, J. (2019). Robust modeling in cognitive science. *Computational Brain & Behavior*, *2*, 141–153. doi: 10.1007/s42113-019-00029-y

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course.* Cambridge University Press.

Leite, F. P., Ratcliff, R., Lette, F. P., & Ratcliff, R. (2010). Modeling reaction time and accuracy of multiple-alternative decisions. *Attention, Perception, and Psychophysics*, *72*(1), 246–273. doi: 10.3758/APP.72.1.246

Lindley, D. V. (1965). *Introduction to probability theory and statistics from a Bayesian point of view.* Cambridge University Press, Cambridge.

Matzke, D., Hughes, M., Badcock, J. C., Michie, P., & Heathcote, A. (2017). Failures of cognitive control or attention? the case of stop-signal deficits in schizophrenia. *Attention, Perception, & Psychophysics*, *79*, 1078–1086.

Matzke, D., & Wagenmakers, E.-J. (2009). Psychological interpretation of the ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review*, *16*(5), 798–817. doi: 10.3758/PBR.16.5.798

McDougal, R. A., Bulanova, A. S., & Lytton, W. W. (2016). Reproducibility in computational neuroscience models and simulations. *IEEE transactions on bio-medical engineering*, *63*(10), 2021–2035. doi: 10.1109/TBME.2016.2539602

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, T. P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The prisma statement. *PLOS Medicine*, *6*(7), 1-6. doi: 10.1371/journal.pmed.1000097

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, *47*(1), 90–100. doi: 10.1016/S0022-2496(02)00028-7

Myung, I. J., & Pitt, M. A. (1997). Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin & Review*, *4*(1), 79–95.

Navarro, D. J. (in press). If mathematical psychology did not exist we would need to invent it: A case study in cumulative theoretical development. *Perspectives on Psychological Science.*

Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. *Psychonomic Bulletin & Review*, *26*(5), 1596–1618.

O'Callaghan, C., Hall, J. M., Tomassini, A., Muller, A. J., Walpola, I. C., Moustafa, A. A., … Lewis, S. J. (2017). Visual hallucinations are characterized by impaired sensory evidence accumulation: Insights from hierarchical drift diffusion modeling in Parkinson's disease. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *2*(8), 680 - 688. doi: https://doi.org/10.1016/j.bpsc.2017.04.007

Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision*, *5*(5), 376–404. doi: http://dx.doi.org/10.1167/5.5.1

Pitt, M. A., & Myung, J. I. (2019, 01). Robust modeling through design optimization. *Computational Brain & Behavior*, *2*(3), 200–201. doi: 10.1007/s42113-019-00050-1

PubMed. (2017). *PubMed.* `https://pubmed.ncbi.nlm.nih.gov/`. (Accessed: 2017-12-26)

R Core Team. (2020). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria: R. F. for S. Computing, Ed.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. doi: 10.1037/0033-295X.85.2.59

Ratcliff, R. (2008). The EZ diffusion method: Too EZ? *Psychonomic Bulletin & Review*, *15*(6), 1218–1228.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922. doi: 10.1162/neco.2008.12-06-420

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*(5), 347–356. doi: 10.1111/1467-9280.00067

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, *20*(4), 260–281. doi: 10.1016/j.tics.2016.01.007

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, *95*(3), 318–339. doi: 10.1037/0033-295X.95.3.318

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. doi: 10.1037/0033-2909.86.3.638

Schad, D. J., Betancourt, M., & Vasishth, S. (2020). Toward a principled Bayesian workflow in cognitive science. *Psychological Methods*. doi: 10.1037/met0000275

Schmitz, F., & Voss, A. (2012). Decomposing task-switching costs With the diffusion model. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(1), 222–250. doi: 10.1037/a0026003

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461–464.

Shankle, W. R., Hara, J., Mangrola, T., Hendrix, S., Alva, G., & Lee, M. D. (2013). Hierarchical Bayesian cognitive processing models to analyze clinical trial data. *Alzheimer's & Dementia*, *9*(4), 422-428. doi: 10.1016/j.jalz.2012.01.016

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *64*(4), 583–616. doi: 10.1111/1467-9868.00353

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(3), 485–493.

Stine, R. (1989). An introduction to bootstrap methods: Examples and ideas. *Sociological Methods & Research*, *18*(2-3), 243–291. doi: 10.1177/0049124189018002003

Stone, M. (1960). Models for choice-reaction time. *Psychometrika*, *25*(3), 251–260.

Strickland, L., Loft, S., Remington, R. W., & Heathcote, A. (2018). Racing to remember: A theory of decision control in event-based prospective memory. *Psychological Review*, *125*(6), 851–887.

Theisen, M., Lerche, V., von Krause, M., & Voss, A. (2020). Age differences in diffusion model parameters: A meta-analysis. *Psychological Research*. doi: 10.1007/s00426-020-01371-8

Trafimow, D. (2005). The ubiquitous Laplacian assumption: Reply to Lee and Wagenmakers (2005). *Psychological Review*, *112*(3), 669–674. doi: 10.1037/0033-295X.112.3.669

Turner, B. M., Sederberg, P. B., Brown, S. D., & Steyvers, M. (2013). A method for efficiently sampling from distributions with correlated dimensions. *Psychological Methods*, *18*, 368-384.

Vandekerckhove, J., & Tuerlinckx, F. (2008). Diffusion model analysis with MATLAB: A DMAT primer. *Behavior Research Methods*, *40*(1), 61–72.

van Maanen, L., van der Mijn, R., van Beurden, M. H. P. H., Roijendijk, L. M. M., Kingma, B. R. M., Miletić, S., & van Rijn, H. (2019). Core body temperature speeds up temporal processing and choice behavior under deadlines. *Scientific Reports*, *9*(1), 10053. doi: 10.1038/s41598-019-46073-3

Vanpaemel, W. (2011). Constructing informative model priors using hierarchical methods. *Journal of Mathematical Psychology*, *55*(1), 106–117. doi: https://doi.org/10.1016/j.jmp.2010.08.005

Vanpaemel, W., & Lee, M. D. (2012). Using priors to formalize theory: Optimal attention and the generalized context model. *Psychonomic Bulletin & Review*, *19*(6), 1047–1056. doi: 10.3758/s13423-012-0300-4

van Ravenzwaaij, D., Donkin, C., & Vandekerckhove, J. (2017). The EZ diffusion model provides a powerful test of simple empirical effects. *Psychonomic Bulletin & Review*, *24*, 547–556.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*, 1413–1432.

Visser, I., & Poessé, R. (2017). Parameter recovery, bias and standard errors in the linear ballistic accumulator model. *British Journal of Mathematical and Statistical Psychology*, *70*(2), 280–296. doi: 10.1111/bmsp.12100

Voss, A., Rothermund, K., & Brandtstädter, J. (2008). Interpreting ambiguous stimuli: Separating perceptual and judgmental biases. *Journal of Experimental Social Psychology*, *44*(4), 1048–1056. doi: 10.1016/j.jesp.2007.10.009

Voss, A., & Voss, J. (2007). Fast-dm: A free program for efficient diffusion model analysis. *Behavior Research Methods*, *39*(4), 767–775.

Wabersich, D., & Vandekerckhove, J. (2014). The RWiener package: An R package providing distribution functions for the Wiener diffusion model. *R Journal*, *6*(1), 49–56.

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*(1), 192–196. doi: 10.3758/BF03206482

Wagenmakers, E.-J., Van Der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, *14*(1), 3–22. doi: 10.3758/BF03194023

Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*, 3571–3594.

White, C. N., & Poldrack, R. A. (2014). Decomposing bias in different types of simple decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 385–398.

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in Neuroinformatics*, *7*, 14. doi: 10.3389/fninf.2013.00014

WoS. (2017). *Web of Science.* `http://www.webofknowledge.com/`. (Accessed: 2017-12-26)