Systematic Parameter Reviews in Cognitive Modeling: Towards a Robust and Cumulative Characterization of Psychological Processes in the Diffusion Decision Model

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

[1] Department of Human Behavior, Ecology and Culture, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, [2] Department of Experimental Psychology, Utrecht University, Utrecht, Netherlands, [3] Department of Psychology, University of Tasmania, Hobart, TAS, Australia, [4] Psychological Methods, Department of Psychology, University of Amsterdam, Amsterdam, Netherlands

Parametric cognitive models are increasingly popular tools for analyzing 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 and 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, diffusion decision model, prior distributions, cumulative science

1. INTRODUCTION

With an expanding recent appreciation of the value of quantitative theories that make clear and testable predictions (Lee and Wagenmakers, 2014; Oberauer and Lewandowsky, 2019; Navarro, 2020), cognitive models have become increasingly popular. As a consequence, open science and reproducibility reforms have been expanded to include modeling problems. 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 and Batchelder, 1988), reinforcement learning models (Busemeyer and Stout, 2002), and evidence-accumulation models (Donkin and 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.
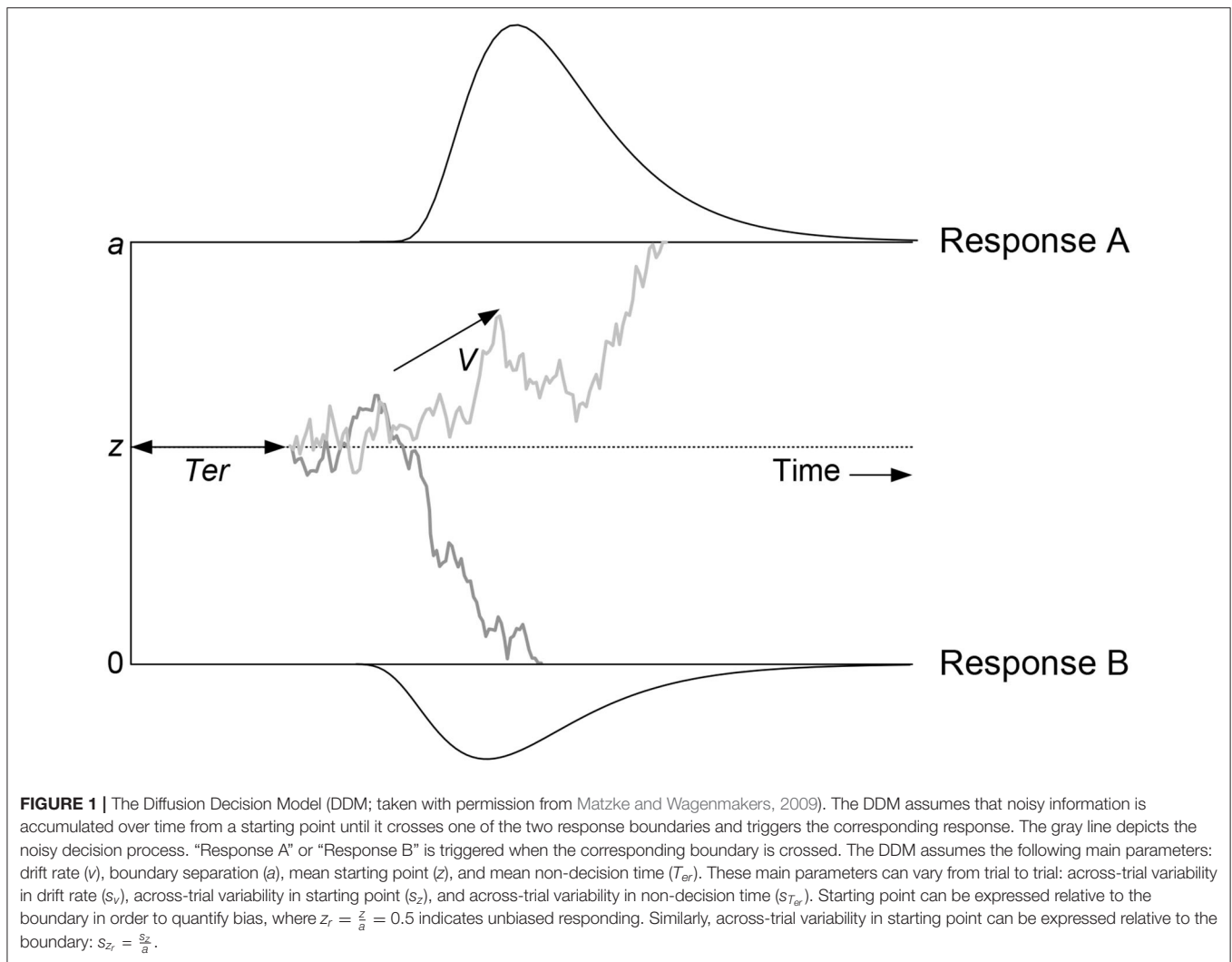
A systematic quantitative characterization of model parameters provides knowledge of the likely values of the model parameters and has various benefits. First, it can promote more precise and realistic simulations that help to optimally calibrate and design experiments, avoiding unnecessary experimental costs (Gluth and Jarecki, 2019; Heck and Erdfelder, 2019; Kennedy et al., 2019; Pitt and Myung, 2019; Schad et al., 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 (Vanpaemel, 2011; Vanpaemel and Lee, 2012; Lee, 2018). Prior distributions play a role both in parameter estimation and model selection. By assigning relatively more weight to plausible regions of the parameter space, informative prior distributions can improve parameter estimation, particularly when the data are not sufficiently informative, for instance due to a small number of observations. Even as the number of observations grows, informative priors remain crucial for Bayesian model selection using Bayes factors (Jeffreys, 1961; Kass and Raftery, 1995). Unfortunately, the theoretical and practical advantages of the prior have been undermined by the common use of vague distributions (Trafimow, 2005; Gill, 2014).

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 and McKoon, 2008), a popular cognitive model for two-choice response time tasks (see Ratcliff et al., 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.

1.1. 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 (Ratcliff and Rouder, 1998; Donkin et al., 2009a; van Maanen et al., 2019). These models assume that 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 (Voss et al., 2008; Leite et al., 2010; Dutilh et al., 2011), prospective memory (Horn et al., 2011; Ball and Aschenbrenner, 2018) to cognitive control (Gomez et al., 2007; Schmitz and Voss, 2012).

Figure 1 illustrates the DDM. Evidence (i.e., gray 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 be less cautious and thus decreases their boundary

FIGURE 1 | The Diffusion Decision Model (DDM; taken with permission from Matzke and 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 gray 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: $s_{z_r} = \frac{s_z}{a}$.

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 toward 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 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: $s_{z_r} = \frac{s_z}{a}$. One parameter of the accumulation process needs to be fixed to establish a scale that makes the other accumulation-related parameters identifiable (Donkin et al., 2009b). 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.

Fitting the DDM and many other evidence-accumulation models to experimental data is difficult because of the complexity of the models and the form of their likelihood resulting in high correlations among the parameters (i.e., "sloppiness"; Gutenkunst et al., 2007; Gershman, 2016). Informative prior distribution can ameliorate some of these problems. The growing popularity of cognitive modeling has led to extensive application of the DDM to empirical data (Theisen et al., 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.

2. MATERIALS 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/.

2.1. Literature Search
The literature search was conducted according to the PRISMA guidelines (Moher et al., 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.

2.1.1. 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: (National Library of Medicine, 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 the **Supplementary Materials** or https://osf.io/9ycu5/ for details). The searches began from the publication year of the seminal paper by Ratcliff (1978). The cited reference searches were based on Ratcliff and McKoon (2008), Wiecki et al. (2013), and Palmer et al. (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 3,000 cited references to the seminal Ratcliff 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.

2.1.2. 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.

2.1.3. 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 across-trial starting point ($s_{z_r} = \frac{s_z}{a}$), 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 or in Bayesian hierarchical applications the group-level estimates. 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 and 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 et al., 2007) or EZ2 (Grasman et al., 2009) methods, or the RWiener R package (Wabersich and Vandekerckhove, 2014), which all fit the simple diffusion model 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.

2.2. 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 **Supplementary Materials**.

2.2.1. 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 and 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 and 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 scaling parameter was not reported and even if we assumed the scaling parameter to be s, its value could not be determined.

---

[1]Based on personal communication with Roger Ratcliff.

2.2.2. 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 rule 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 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.

2.2.3. 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]

2.2.4. Drift Rate

There are two ways in which drift rates v can be reported. In the first, positive drift rates indicate a correct response (e.g., "word" response to a "word" stimulus and "non-word" response to a "non-word" stimulus in a lexical decision task) and negative rates indicate an incorrect response. In the second, positive drift rates correspond to one response option (e.g., "word" response) and negative rates to the other option (e.g., "non-word"). Here, we adopt the former—accuracy coding—method in order to avoid ambiguity regarding the arbitrary attribution of boundaries to response options. We do so by taking the absolute values of the reported drift rates to construct the prior distribution. Readers who wish to adopt the latter—response coding—method, should appropriately mirror our accuracy-coded priors around 0.

2.3. Generating Informative Prior Distributions for the DDM Parameters

After post-processing and transforming the parameter estimates, we combined each parameter type across articles and experimental conditions within each study into separate univariate distributions. We then attempted to characterize these empirical parameter distributions with theoretical distributions that provided the best fit to the overall shape of the distributions of parameter estimates.

2.3.1. 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 model-selection procedures. After extracting all parameters from the best fitting models, we identified parameters that were constrained across within- and between-subject manipulations, conditions, or groups within each study. For the purpose of constructing the prior distributions we only considered these fixed parameters once and did not repeatedly include them in the empirical distributions. For instance, a random dot motion task with three difficulty conditions may provide only one estimate for a constrained parameter (i.e., non-decision time), but three parameters for an unconstrained parameter (i.e., drift rate).

2.3.2. Synthesis Across Articles

Most studies reported parameter estimates aggregated across participants, with only eight reporting individual estimates. Before collapsing them with the aggregated estimates, individual estimates were averaged across participants in each study. Parameter estimates were equally weighted when combined across studies as details necessary for weighting them according to their precision were typically not available. We will revisit this decision in the Discussion.

---

[2]Based on personal communication with Roger Ratcliff.

[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.

In the results reported in the main body of this article, we aggregated the parameter estimates across all research domains (e.g., neuroscience, psychology, economics), populations (e.g., low/high socioeconomic status, clinical populations), and tasks (e.g., lexical decision, random dot motion tasks). In the **Supplementary Materials**, we provide examples of prior distributions derived specifically for two of the most common tasks in our database (i.e., lexical decision and random dot motion tasks) and priors restricted to non-clinical populations. Data and code to generate such task and population-specific priors are available in the open repository, so that interested readers can construct priors relevant to their specific research questions.

2.3.3. Distributions

A full characterization of the distribution of model parameters takes into account not only the parameters' average values and variability but also their correlations across participants (e.g., people with lower drift rates may have higher thresholds) and potentially even their correlations across studies or paradigms using multilevel structures. Although multivariate prior distributions would be optimal to represent correlations across participants, they require individual parameter estimates for the estimation of the covariance matrices. As only eight studies reported individual parameter estimates, we were restricted to use univariate distributions.

We attempted to characterize the aggregated results using a range of univariate distribution functions that respected the parameter types' 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 clearly could not be captured by the univariate distributions 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 empirical distribution 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), 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 other parameters. We then used AIC weights (wAIC; Wagenmakers and 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.
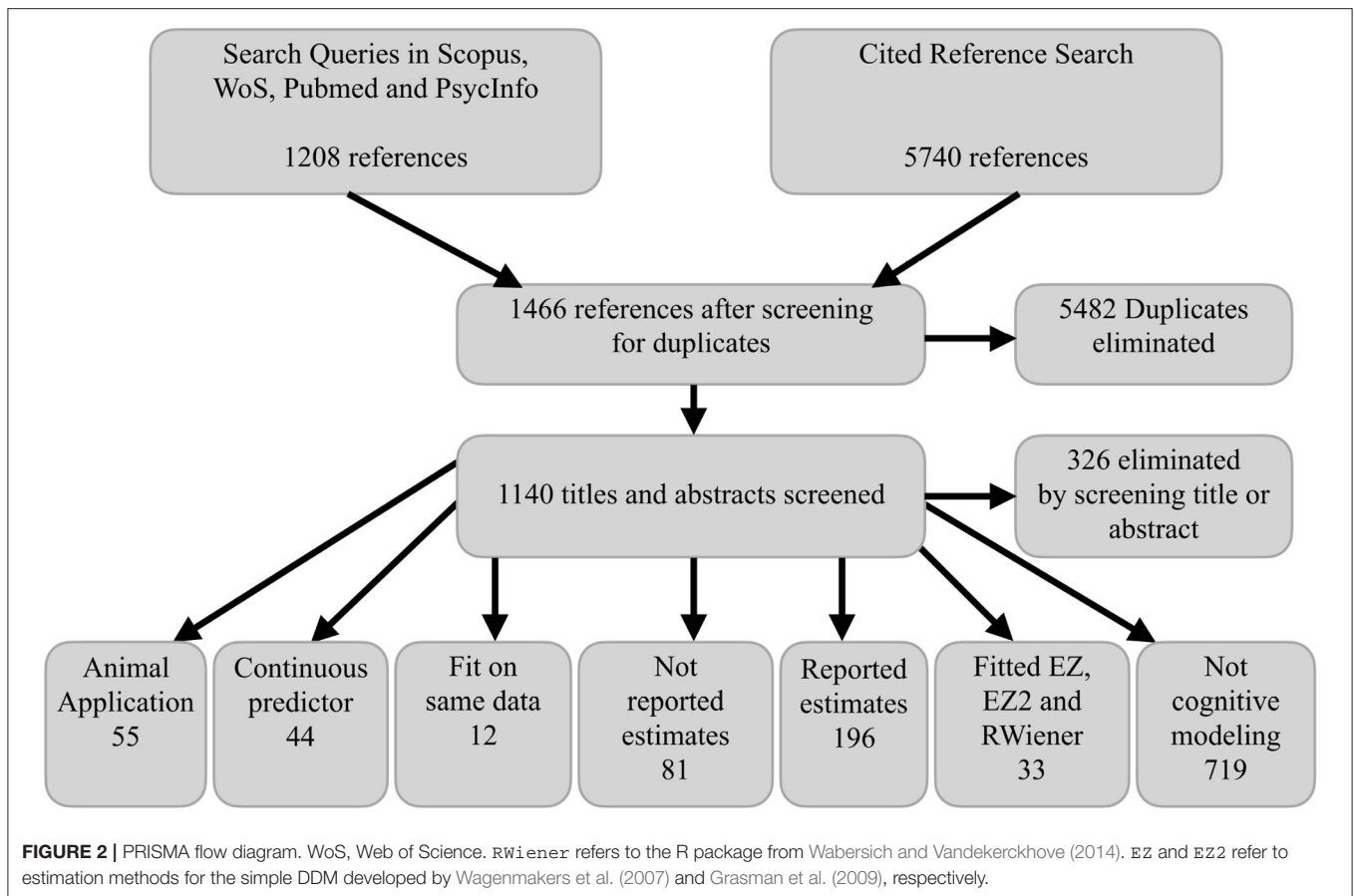
3. RESULTS

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 scaling parameters and we were unable to reverse engineer them 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 (33), and recognition memory tasks (17). 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. The red lines show the best fitting theoretical distributions or the dominant theoretical distribution components with the highest mixture weight (i.e., the proposed informative prior distributions). The black lines show the non-dominant mixture distribution components. Note that in most cases the the mixture served to inflate the distributions' tails while preserving a single mode.

Table 1 gives an overview of the informative prior distributions, the corresponding upper and lower bounds (see column "T-LB" and "T-UB"), and whenever appropriate also the mixture weight of the dominant distribution component. 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 limits 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 mixture of a zero-bounded truncated normal and a lognormal distribution (wAIC = 0.4), with the mixture weight, and the location and scale of the dominant truncated normal component shown in the first row of **Table 1**.[4] 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}

---

[4]The location and scale parameters of the truncated normal distribution refer to μ and σ and not to its expected value and variance.

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.

and 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 zero-bounded truncated normal and a gamma distribution (wAIC = 0.35), where the truncated 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).

4. 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 the parameters of cognitive models. We believe that for cognitive models where sufficient information are available in the literature, a systematic quantitative characterization of model parameters can be a very useful addition to existing 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 and Jarecki, 2019; Heck and Erdfelder, 2019; Pitt and 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. Although, for simplicity, here we suggested single-component distributions as priors, the full mixture distributions that we selected could also be used. 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, including mixtures. In most cases the mixture served to inflate the distributions' tails while preserving a single mode. However, aggregation over heterogeneous studies naturally carries with it the possibility of creating multi-modal prior distributions, as illustrated by the results for s_v in **Figure 3**. If the data proved sufficiently uninformative that such multi-modality carried through to the posterior, caution should be exercised in reporting and interpreting measures of central tendency.

Inferring the parameters of complex cognitive models like the DDM from experimental data is challenging because their parameters are often highly correlated. The cumulative
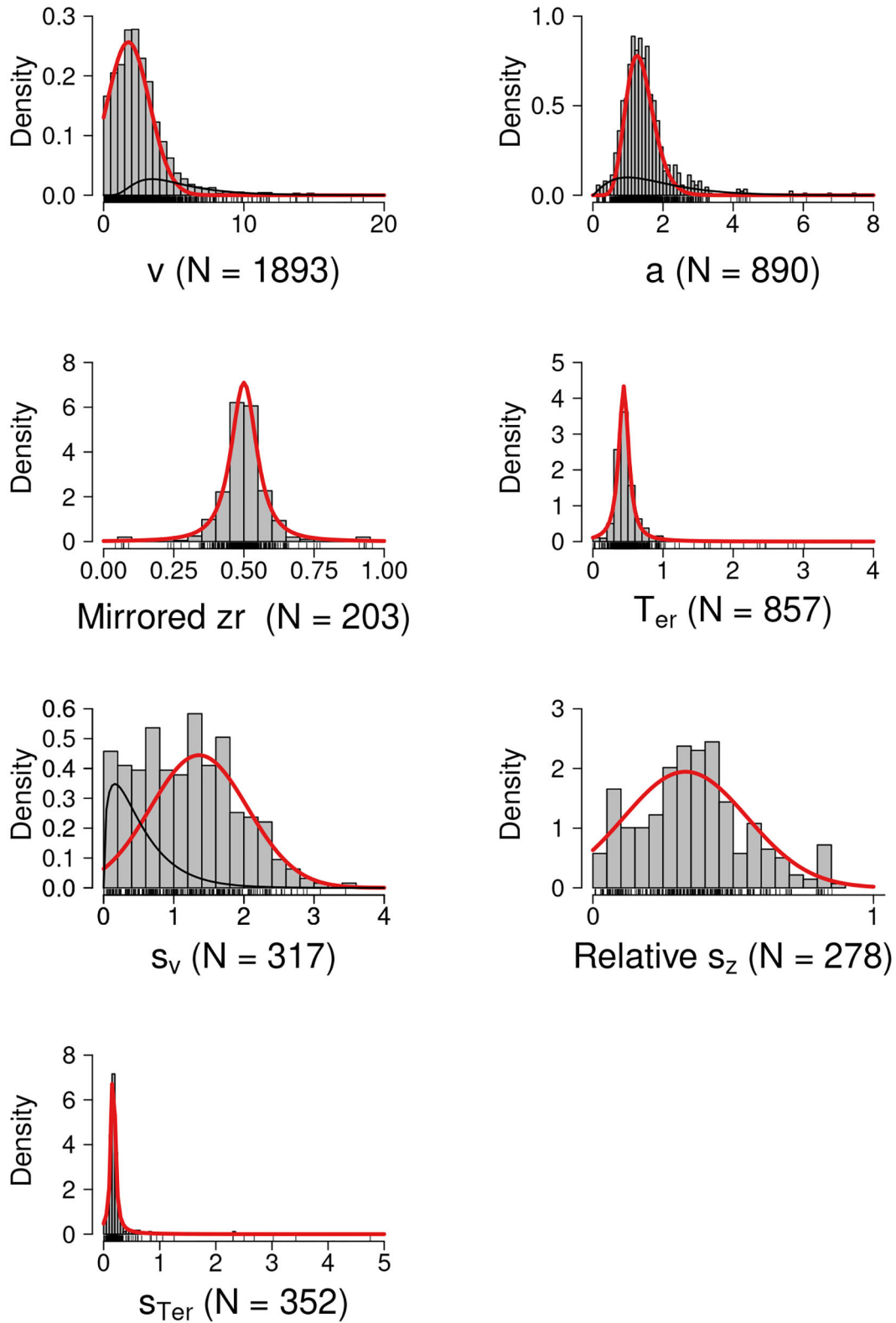
FIGURE 3 | Prior Distributions for the DDM Parameters. The red lines show the best fitting theoretical distributions or the dominant theoretical distribution components with the highest mixture weight (i.e., the proposed informative prior distributions). The black lines show the non-dominant distribution components. N, number of unique estimates.

TABLE 1 | Informative prior distributions.

DDM parameter	N	Distribution	Weight	Location/Shape	Scale	df	T-LB	T-UB	E-LB	E-UB
v	1,893	Truncated normal* & lognormal	0.85	1.76	1.51		0	+ Inf	0.01	18.51
a	890	Gamma* & gamma	0.76	11.69	0.12		0	+ Inf	0.11	7.47
Mirrored z_r	203	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* & gamma	0.75	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

N, The number of unique estimates; Weight, The mixture weight of the dominant distribution component; 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; *, Dominant distribution component.

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 essential for Bayesian model selection using Bayes factors (Jeffreys, 1961; Kass and Raftery, 1995). Unlike other model-selection methods like the Deviance Information Criteria (DIC; Spiegelhalter et al., 2014) and Widely Applicable Information Criterion (WAIC; Vehtari et al., 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 and Annis, 2019; Gronau et al., 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 et al., 2017). Informative priors may provide an alternative solution that avoids the potential systematic distortion caused by ignoring the variability parameters (Ratcliff and 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 et al., submitted). Of course, in extreme cases, the central tendency of informative prior distributions may provide guidelines for fixing difficult-to-estimate model parameters to a constant (e.g., Matzke et al., 2020).

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., Shankle et al., 2013; Matzke et al., 2017). 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.

Ideally, informative prior distributions for cognitive models should be based on prior information extracted from experimental paradigms (or classes of paradigms) and participant populations relevant to the research question at hand, although care should be exercised in the latter case where group members fall along a continuum (e.g., age or the severity of a clinical diagnosis). In reality, constructing such highly specific priors might not always be feasible, either because of a paucity of relevant parameter values reported in the literature, or when new paradigms or populations are studied. However, we believe that using informative priors based on a range of broadly similar paradigms and heterogeneous populations is better than using vague priors, as long as appropriate caution is exercised in cases when the data are not sufficiently informative and hence the prior dominates inferences about the model parameters. Here, we presented informative prior distributions for parameter estimates aggregated across paradigms and populations and also provided paradigm-specific priors for the two most popular tasks in our database (e.g., lexical decision and random dot motion task) and priors for non-clinical populations.

The variance of the paradigm/population-specific priors showed a general decreasing tendency. The decrease in variance was relatively small for the priors based on non-clinical populations, and was restricted to the main DDM parameters. For the lexical decision task, the variance of all of the priors decreased relative to the overall priors. For the random dot motion task, with the exception of v, all variances decreased, albeit the decrease was negligible for a. To summarize, for the tasks and groups we examined here, paradigm and population heterogeneity appears to introduce additional variability in the parameter estimates, but the degree of additional variability strongly depends on the type of parameter. Our open repository provides the data and code to generate informed prior distributions for any selection of studies included in the database so that researchers can construct informative prior distributions relevant to their own research questions. Naturally, paradigm/population-specific priors are only sensible when a sufficiently large number of parameter estimates are available in the database to fit the theoretical (mixture) distributions, or when researchers can augment the repository with estimates from additional studies.

Despite their usefulness, systematic quantitative parameter reviews are not without 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 toward 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 endeavor.

4.1. 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.

4.1.1. 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 et al., 2009b; van Maanen and Miletić, 2020). 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 influences 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 report 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 et al., 2009a) or the results of model-selection procedures (e.g., Heathcote et al., 2015; Strickland et al., 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 et al., 2002, 2014), WAIC (Watanabe, 2010), or Bayes factors (Kass and 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 et al., 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.

4.1.2. 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 and Voss, 2007), HDDM (Wiecki et al., 2013), DMC (Heathcote et al., 2019), DMAT (Vandekerckhove and Tuerlinckx, 2008), using a variety of estimation methods, such as maximum likelihood, Kolmogorov-Smirnov, chi-squared minimization (Voss and Voss, 2007), quantile maximum probability (Heathcote and Brown, 2004), or Bayesian Markov chain Monte Carlo (MCMC; e.g., Turner et al., 2013) techniques. We encourage researchers to report the software they used, and whenever possible, share their commented code to enable computational reproducibility (McDougal et al., 2016; Cohen-Boulakia et al., 2017). Knowledge about the estimation software can also provide valuable information about the parametrization and scaling issues described above.

4.1.3. Parameter Estimates and Uncertainty
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, this would mean reporting 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. Ideally, in the process of aggregation used to create prior distributions, estimates should be weighted by their relative uncertainty. The weighing should reflect the uncertainty of the individual estimates resulting from fitting the model to finite data and—if average parameters are used, as was the case here—also sampling error reflecting the sample size used in each study. Although we had access to the sample sizes, most studies reported parameter estimates averaged across participants without accounting for the uncertainty of the individual estimates. Moreover, the few studies that reported individual estimates provided only point estimates and failed to include measures of uncertainty. As a proxy to participant-level measures of uncertainty one may use the number of trials that provide information for the estimation of the various model parameters. However, this approach requires a level of detail about the experimental design and the corresponding model specification (including the number of excluded trials per participant) that was essentially never available in the surveyed studies.

Given these problems with reporting, we have decided to give equal weights to all (averaged) parameter estimates regardless of the sample size. The reason for this decision was that studies with large sample sizes typically used a small number of trials and likely resulted in relatively imprecise individual estimates, whereas studies with small sample sizes typically used a large number of trials and likely resulted in relatively precise individual estimates. We reasoned that as a result of this trade-off, the equal weighting may not be necessarily unreasonable. To remedy this problem in future parameter reviews, we urge researchers to either report properly weighted group average estimates or report individual estimates along with measures of uncertainty, let these be (analytic or bootstrapped) frequentist standard errors and confidence intervals (e.g., Visser and Poessé, 2017), or Bayesian credible intervals and full posterior distributions (Jeffreys, 1961; Lindley, 1965; Eberly and Casella, 2003).

4.1.4. Individual Parameters and Correlations
Ideally, researchers should report parameter estimates for each individual participant. 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. Access to individual parameters would allow estimation of these correlations, and thus enable priors to reflect this potentially important information. 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.

4.2. Limitations and Future Directions
The approach to parameter reviews taken here—obtaining values from texts, tables, and graphs from published papers and performing an aggregation across studies—has the advantage of sampling estimates that are representative of a wide variety of laboratories, paradigms, and estimation methods. Indeed, for the priors presented in **Figure 3** we included a few studies with much longer RTs than are typically fit with the DDM (e.g., Lerche and Voss, 2019). The larger parameter values from these studies had the effect of broadening the tails of the fitted distributions so they represent the full variety of estimates reported in the literature.

However, this approach 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. The most straightforward example concerns including parameters from studies with long RTs. The solution is equally straightforward: only including studies with RTs that fall in the range of interest specific to a particular application. A related but more subtle issue occurs in our DDM application where the meaning of the magnitude of the response bias (z_r) parameter is design specific, and so it is difficult to form useful aggregates over

different paradigms. To take a concrete example, a bias toward "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. Again, this problem can again be avoided by constructing priors based on a more specific (in this case task-specific) aggregation. 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. The priors for the lexical decision task reported in the **Supplementary Materials** provide an example that did not require us to mirror the bias estimates. We note, however, that we had to exclude a paper where it was unclear which response was mapped to which DDM boundary, so we would add a reporting guideline that this choice be spelled out. We also note that similar problems with aggregation over different designs 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 and 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 and 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[5] 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.

---

[6]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.

Problems related to this limitation 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. With Bayesian methods, 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 et al., 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 make this enterprise worthwhile.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://osf.io/9ycu5/and GitHub: https://github.com/nhtran93/DDM_priors.

AUTHOR CONTRIBUTIONS

N-HT designed the study, collected the data, performed the analyses, and wrote the manuscript. N-HT, LvM, and DM designed the study and the analyses. DM and AH provided critical feedback and helped shape the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.608287/full#supplementary-material

REFERENCES

Akaike, H. (1973). "Information theory and an extension of maximum likelihood principle," in *Proceedings of the Second International Symposium on Information Theory* (New York, NY), 267–281.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Control* 19, 716–723. doi: 10.1109/TAC.1974.1100705

Akaike, H. (1981). Likelihood of a model and information criteria. *J. Econometr.* 16, 3–14. doi: 10.1016/0304-4076(81)90071-3

Allaire, J., Cheng, J., Xie, Y., McPherson, J., Chang, W., Allen, J., et al. (2018). *rmarkdown: Dynamic Documents for R.* R package version 1.

American Psychological Association (2017). *Psycinfo.* Available online at: https://www.apa.org/pubs/databases/psycinfo/ (accessed December 26, 2017).

Arizona (2010). *GraphClick.*

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

Boehm, U., Annis, J., Frank, M. J., Hawkins, G. E., Heathcote, A., Kellen, D., et al. (2018). Estimating across-trial variability parameters of the diffusion decision model: expert advice and recommendations. *J. Math. Psychol.* 87, 46–75. doi: 10.1016/j.jmp.2018.09.004

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

Busemeyer, J. R., and Stout, J. C. (2002). A contribution of cognitive decision models to clinical assessment: decomposing performance on the Bechara gambling task. *Psychol. Assess.* 14, 253–262. doi: 10.1037/1040-3590.14.3.253

Cohen-Boulakia, S., Belhajjame, K., Collin, O., Chopard, J., Froidevaux, C., Gaignard, A., et al. (2017). Scientific workflows for computational reproducibility in the life sciences: status, challenges and opportunities. *Future Generat. Comput. Syst.* 75, 284–298. doi: 10.1016/j.future.2017.01.012

Damaso, K., Williams, P., and 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., and Heathcote, A. (2009a). Getting more from accuracy and response time data: Methods for fitting the linear ballistic accumulator. *Behav. Res. Methods* 41, 1095–1110. doi: 10.3758/BRM.41.4.1095

Donkin, C., and Brown, S. D. (2018). "Response times and decision-making," in *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, Volume 5: Methodology, 4th Edn.*, eds E. J. Wagenmakers and J. T. Wixted (John Wiley and Sons, Inc.), 349–382.

Donkin, C., Brown, S. D., and Heathcote, A. (2009b). The overconstraint of response time models: rethinking the scaling problem. *Psychon. Bull. Rev.* 16, 1129–1135. doi: 10.3758/PBR.16.6.1129

Dutilh, G., Annis, J., Brown, S. D., Cassey, P., Evans, N. J., Grasman, R. P., et al. (2019). The quality of response time data inference: a blinded, collaborative sssessment of the validity of cognitive models. *Psychon. Bull. Rev.* 26, 1051–1069. doi: 10.3758/s13423-017-1417-2

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

Eberly, L. E., and Casella, G. (2003). Estimating Bayesian credible intervals. *J. Stat. Plann. Inference* 112, 115–132. doi: 10.1016/S0378-3758(02)00327-0

Elsevier (2017). *Scopus.* Available online at: https://www.scopus.com/home.uri (accessed December 12, 2017).

Evans, N. J., and Annis, J. (2019). Thermodynamic integration via differential evolution: a method for estimating marginal likelihoods. *Behav. Res. Methods* 51, 930–947. doi: 10.3758/s13428-018-1172-y

Gershman, S. J. (2016). Empirical priors for reinforcement learning models. *J. Math. Psychol.* 71, 1–6. doi: 10.1016/j.jmp.2016.01.006

Gill, J. (2014). *Bayesian Methods:A Social and Behavioral Sciences Approach.* New York, NY: Chapman and Hall. doi: 10.1201/b17888

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

Gomez, P., Ratcliff, R., and Perea, M. (2007). A model of the Go/No-Go task. *J. Exp. Psychol. Gen.* 136, 389–413. doi: 10.1037/0096-3445.136.3.389

Grasman, R. P., Wagenmakers, E.-J., and van der Maas, H. L. (2009). On the mean and variance of response times under the diffusion model

with an application to parameter estimation. *J. Math. Psychol.* 53, 55–68. doi: 10.1016/j.jmp.2009.01.006

Gronau, Q. F., Heathcote, A., and Matzke, D. (2020). Computing bayes factors for evidence-accumulation models using Warp-III bridge sampling. *Behav. Res. Methods* 52, 918–937. doi: 10.3758/s13428-019-01290-6

Gunawan, D., Hawkins, G. E., Tran, M. N., Kohn, R., and Brown, S. D. (2020). New estimation approaches for the hierarchical Linear Ballistic Accumulator model. *J. Math. Psychol.* 96:102368. doi: 10.1016/j.jmp.2020.102368

Gutenkunst, R. N., Waterfall, J. J., Casey, F. P., Brown, K. S., Myers, C. R., and Sethna, J. P. (2007). Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput. Biol.* 3:e189. doi: 10.1371/journal.pcbi.0030189

Heathcote, A., and Brown, S. (2004). Reply to Speckman and Rouder: a theoretical basis for QML. *Psychon. Bull. Rev.* 11, 577–578. doi: 10.3758/BF03196614

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

Heathcote, A., Loft, S., and Remington, R. W. (2015). Slow down and remember to remember! A delay theory of prospective memory costs. *Psychol. Rev.* 122, 376–410. doi: 10.1037/a0038952

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

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. *Stat. Sci.* 14, 382–401. doi: 10.1214/ss/1009212519

Horn, S. S., Bayen, U. J., and Smith, R. E. (2011). What can the diffusion model tell Us about prospective memory? *Can. J. Exp. Psychol.* 65, 69–75. doi: 10.1037/a0022808

Jaynes, E. T. (1988). "The relation of Bayesian and maximum entropy methods," in *Maximum Entropy and Bayesian Methods in Science and Engineering*, eds G. J. Erickson and C. Smith (Dordrecht: Kluwer Academic Publishers), 25–29. doi: 10.1007/978-94-009-3049-0_2

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

Kass, R. E., and Raftery, A. E. (1995). Bayes factors. *J. Am. Stat. Assoc.* 90, 773–795. doi: 10.1080/01621459.1995.10476572

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

Lee, M. D. (2018). "Bayesian methods in cognitive modeling," in *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, Volume 5: Methodology, 4th Edn.*, eds E. J. Wagenmakers and J. T. Wixted (New York, NY: John Wiley and Sons, Inc.), 37–84.

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

Lee, M. D., and Wagenmakers, E.-J. (2014). *Bayesian Cognitive Modeling: A Practical Course.* Cambridge University Press.

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

Lerche, V., and Voss, A. (2019). Experimental validation of the diffusion model based on a slow response time paradigm. *Psychol. Res.* 83, 1194–1209. doi: 10.1007/s00426-017-0945-8

Lindley, D. V. (1965). *Introduction to Probability Theory and Statistics From a Bayesian Point of View.* Cambridge: Cambridge University Press.

Matzke, D., Hughes, M., Badcock, J. C., Michie, P., and Heathcote, A. (2017). Failures of cognitive control or attention? The case of stop-signal deficits in schizophrenia. *Attent. Percept. Psychophys.* 79, 1078–1086. doi: 10.3758/s13414-017-1287-8

Matzke, D., Logan, G. D., and Heathcote, A. (2020). A cautionary note on evidence-accumulation models of response inhibition in the stop-signal paradigm. *Comput. Brain Behav.* 3, 269–288. doi: 10.1007/s42113-020-00075-x

Matzke, D., and Wagenmakers, E.-J. (2009). Psychological interpretation of the ex-Gaussian and shifted Wald parameters: a diffusion model analysis. *Psychon. Bull. Rev.* 16, 798–817. doi: 10.3758/PBR.16.5.798

McDougal, R. A., Bulanova, A. S., and Lytton, W. W. (2016). Reproducibility in computational neuroscience models and simulations. *IEEE*

Trans. Biomed. Eng. 63, 2021–2035. doi: 10.1109/TBME.2016.25
39602

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., and Group, T. P. (2009).
Preferred reporting items for systematic reviews and meta-analyses: the prisma
statement. *PLoS Med.* 6:e1000097. doi: 10.1371/journal.pmed.1000097

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

Myung, I. J., and Pitt, M. A. (1997). Applying Occam's razor in modeling cognition:
a Bayesian approach. *Psychon. Bull. Rev.* 4, 79–95. doi: 10.3758/BF03210778

National Library of Medicine (2017). *Empirical Priors for DDM & LBA.*

Navarro, D. J. (2020). If mathematical psychology did not exist we might
need to invent it: a comment on theory building in psychology. *PsyArXiv.*
doi: 10.31234/osf.io/ygbjp

Oberauer, K., and Lewandowsky, S. (2019). Addressing the
theory crisis in psychology. *Psychon. Bull. Rev.* 26, 1596–1618.
doi: 10.3758/s13423-019-01645-2

O'Callaghan, C., Hall, J. M., Tomassini, A., Muller, A. J., Walpola, I. C., Moustafa,
A. A., et al. (2017). Visual hallucinations are characterized by impaired sensory
evidence accumulation: Insights from hierarchical drift diffusion modeling in
Parkinson's disease. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 2, 680–688.
doi: 10.1016/j.bpsc.2017.04.007

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

Pitt, M. A., and Myung, J. I. (2019). Robust modeling through design optimization.
Comput. Brain Behav. 2, 200–201. doi: 10.1007/s42113-019-00050-1

R Core Team (2020). *R: A Language and Environment for Statistical Computing.*
Vienna: R Core Team.

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

Ratcliff, R. (2008). The EZ diffusion method: Too EZ? *Psychon. Bull. Rev.* 15,
1218–1228. doi: 10.3758/PBR.15.6.1218

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

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

Ratcliff, R., Smith, P. L., Brown, S. D., and McKoon, G. (2016). Diffusion
decision model: current issues and history. *Trends Cogn. Sci.* 20, 260–281.
doi: 10.1016/j.tics.2016.01.007

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

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

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

Schmitz, F., and Voss, A. (2012). Decomposing task-switching costs with
the diffusion model. *J. Exp. Psychol. Hum. Percept. Perform.* 38, 222–250.
doi: 10.1037/a0026003

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.* 6, 461–464.
doi: 10.1214/aos/1176344136

Shankle, W. R., Hara, J., Mangrola, T., Hendrix, S., Alva, G., and Lee, M. D. (2013).
Hierarchical Bayesian cognitive processing models to analyze clinical trial data.
Alzheimers Dement. 9, 422–428. doi: 10.1016/j.jalz.2012.01.016

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian
measures of model complexity and fit. *J. R. Stat. Soc. Ser. B* 64, 583–616.
doi: 10.1111/1467-9868.00353

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2014). The
deviance information criterion: 12 years on. *J. R. Stat. Soc. Ser. B* 76, 485–493.
doi: 10.1111/rssb.12062

Stone, M. (1960). Models for choice-reaction time. *Psychometrika* 25, 251–260.
doi: 10.1007/BF02289729

Strickland, L., Loft, S., Remington, R. W., and Heathcote, A. (2018). Racing to
remember: a theory of decision control in event-based prospective memory.
Psychol. Rev. 125, 851–887. doi: 10.1037/rev0000113

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

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

Turner, B. M., Sederberg, P. B., Brown, S. D., and Steyvers, M. (2013). A method
for efficiently sampling from distributions with correlated dimensions. *Psychol.
Methods* 18, 368–384. doi: 10.1037/a0032222

van Maanen, L., and Miletić, S. (2020). The interpretation of behavior-
model correlations in unidentified cognitive models. *Psychon. Bull. Rev.*
doi: 10.3758/s13423-020-01783-y

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

van Ravenzwaaij, D., Donkin, C., and Vandekerckhove, J. (2017). The EZ diffusion
model provides a powerful test of simple empirical effects. *Psychon. Bull. Rev.*
24, 547–556. doi: 10.3758/s13423-016-1081-y

Vandekerckhove, J., and Tuerlinckx, F. (2008). Diffusion model analysis
with MATLAB: a DMAT primer. *Behav. Res. Methods* 40, 61–72.
doi: 10.3758/BRM.40.1.61

Vanpaemel, W. (2011). Constructing informative model priors using
hierarchical methods. *J. Math. Psychol.* 55, 106–117. doi: 10.1016/j.jmp.2010.
08.005

Vanpaemel, W., and Lee, M. D. (2012). Using priors to formalize theory: optimal
attention and the generalized context model. *Psychon. Bull. Rev.* 19, 1047–1056.
doi: 10.3758/s13423-012-0300-4

Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical Bayesian model evaluation
using leave-one-out cross-validation and WAIC. *Stat. Comput.* 27, 1413–1432.
doi: 10.1007/s11222-016-9696-4

Visser, I., and Poessé, R. (2017). Parameter recovery, bias and standard errors in
the linear ballistic accumulator model. *Brit. J. Math. Stat. Psychol.* 70, 280–296.
doi: 10.1111/bmsp.12100

Voss, A., Rothermund, K., and Brandtstädter, J. (2008). Interpreting ambiguous
stimuli: separating perceptual and judgmental biases. *J. Exp. Soc. Psychol.* 44,
1048–1056. doi: 10.1016/j.jesp.2007.10.009

Voss, A., and Voss, J. (2007). Fast-DM: a free program for efficient diffusion model
analysis. *Behav. Res. Methods* 39, 767–775. doi: 10.3758/BF03192967

Wabersich, D., and Vandekerckhove, J. (2014). The RWiener package: an R package
providing distribution functions for the Wiener diffusion model. *R J.* 6, 49–56.
doi: 10.32614/RJ-2014-005

Wagenmakers, E.-J., and Farrell, S. (2004). AIC model selection using
Akaike weights. *Psychon. Bull. Rev.* 11, 192–196. doi: 10.3758/BF032
06482

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

Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely
applicable information criterion in singular learning theory. *J. Mach. Learn. Res.*
11, 3571–3594. Available online at: https://dl.acm.org/doi/10.5555/1756006.
1953045. doi: 10.5555/1756006.1953045

White, C. N., and Poldrack, R. A. (2014). Decomposing bias in different
types of simple decisions. *J. Exp. Psychol. Learn. Mem. Cogn.* 40, 385–398.
doi: 10.1037/a0034851

Wiecki, T. V., Sofer, I., and Frank, M. J. (2013). HDDM: hierarchical Bayesian
estimation of the drift-diffusion model in Python. *Front. Neuroinform.* 7:14.
doi: 10.3389/fninf.2013.00014

Wo, S. (2017). *Web of Science.* Available online at: http://www.webofknowledge.
com/ (accessed December 26, 2017).

Conflict of Interest: 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.