

An EZ Bayesian hierarchical drift diffusion model for response time and accuracy

Adriana F. Chávez De la Peña^{a,b} and Joachim Vandekerckhove^{a,b,*}

This is the author final version of:

Chávez De la Peña, A. F., & Vandekerckhove, J. (in press). An EZ Bayesian hierarchical drift diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*.

The EZ-diffusion model is a simplification of the popular drift diffusion model of choice response times that allows researchers to calculate diffusion model parameters directly from data with no need for expensive computations. The EZ-diffusion model is based on a system of equations in which the diffusion model's drift rate, boundary separation, and nondecision time parameters are jointly used to predict three summary statistics (the accuracy rate and the mean and variance of the correct response times). These equations can then be inverted to obtain estimators for the three parameters from these summary statistics. Here, we describe a probabilistic formulation of the EZ-diffusion model that can serve as a hyper-efficient proxy model to the drift diffusion model. The new formulation is based on sampling distributions of summary statistics and consists only of normal and binomial distributions. It can easily be implemented in any probabilistic programming language. We demonstrate the validity of the proxy model through extensive simulation studies and provide multiple examples (via osf.io/bzkpn), including an implementation in JASP. We conclude that, although the recovery of some parameters with the proxy model is biased, the recovery of regression parameters is good, making the method useful for cognitive psychometrics (i.e., explanatory cognitive modeling). Casting the EZ-diffusion model in the broad family of Bayesian generative models allows us to benefit from mature implementations, practical workflows, and powerful extensions that are not possible without a probabilistic implementation and not feasible with the regular drift diffusion model. Code and example applications are provided via osf.io/bzkpn.

Hierarchical Bayesian | EZ diffusion | cognitive psychometrics | indirect inference

Cognitive psychometrics is a relatively new, specialized discipline in which advances in cognitive modeling are applied in psychological measurement (Batchelder, 2016). The advance of cognitive psychometrics reflects a broader trend towards the use of idealized models in the philosophy of measurement (Tal, 2020).

The relationship between cognitive science and cognitive psychometrics is by nature symbiotic: As cognitive scientists develop ever more sophisticated models of cognition and behavior, so are psychometricians empowered to use those models as measurement tools. The goal of cognitive psychometrics is to construct an algebra of data, so that complex and nonlinear data patterns can be expressed as simple composites of interpretable units. An example success story of this research program is the drift diffusion model (DDM) for two-choice response times (RTs). The model was developed, gradually, by Stone (1960) and Link (1975). It was then made popular through extensive work by Ratcliff (1978), who showed that diffusion model parameters behave like their namesakes in a wide variety of contexts (Wagenmakers, 2009; see also Voss, Rothermund, & Voss, 2004; but see Lerche & Voss, 2018, and Rafiei & Rahnev, 2021, for violations of selective influence). After that, the model was developed into a measurement tool (Vandekerckhove, Verheyen, & Tuerlinckx, 2010; Vandekerckhove, Tuerlinckx, & Lee, 2011), and many interesting applications have resulted (see, e.g., Ratcliff, Smith, Brown, & McKoon, 2016). Taken together, this is a cognitive psychometric research program, intended to quantify patterns of data that are indicative of underlying latent features that vary between individuals and other empirical

units. When quantified, we may then seek to *explain* the observed variability (De Boeck & Wilson, 2004; Vandekerckhove, 2014).

Theoretical progress in cognitive psychometrics is achieved through the development of cognitive models to describe the latent processes that generate data. Ideally, the models specify such processes with parameters that carry psychological meaning and that can be interpreted in the context of the data collection, while formalizing only those substantive assumptions about cognition to which the researcher is willing to commit.

Some of these models, however, are somewhat involved and require significant modeling expertise to be implemented. The DDM in particular became popular for its ability to account for data patterns that are known to be elicited by specific aspects of experimental task designs, but its implementation is computationally complex, erecting a barrier to applications. In response to these limitations, Wagenmakers, van der Maas, and Grasman (2007) developed an "EZ" version of the DDM that allows the user to estimate the model's key parameters from summary statistics of observed choice RT data (i.e., the mean and variance of the correct RTs and the accuracy rate). The EZ method (hereafter, EZ-diff) has since been widely applied in the study of, among others, perceptual decision-making (Bitzer, Park, Blankenburg, & Kiebel, 2014; Mulder, van Maanen, & Forstmann, 2014), of working memory and intelligence (Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007), of self-regulation (Enkavi et al., 2019).

The Bayesian implementation of cognitive process models is crucial to the cognitive psychometrician, as it facilitates their application as measurement models. Cognitive process models that are typically used to account for the underlying response processes at the individual level can be extended hierarchically to take into account between-subjects variability and capture individual differences (Lee, 2011; Vandekerckhove et al., 2011). Furthermore, models can be extended into cognitive latent variable models that

^aDepartment of Cognitive Sciences, University of California, Irvine; ^bDepartment of Statistics, University of California, Irvine
All authors contributed to the final draft.

*Correspondence should be addressed to Joachim Vandekerckhove (joachim@uci.edu).
JV was supported by National Science Foundation grants #1850849 and #2051186.

distinguish across different levels of variation, or incorporate meta-regression structures that capture predictors associated with any of the parameters (Vandekerckhove, 2014).

To deal with complex models with intractable likelihoods, statisticians will sometimes adopt so-called ‘indirect inference’ approaches, in which a target likelihood is approximated by an auxiliary or proxy model (Jiang & Turnbull, 2004; Price, Drovandi, Lee, & Nott, 2017). For example, Wood’s (2010) synthetic likelihood method models relevant summary statistics using a multivariate normal distribution.¹

In this paper, we propose a similar approach to working with the DDM. We construct a proxy model from the sampling distributions of the EZ-diff summary statistics. This simplified model can easily be implemented in specialized software as a Bayesian generative model, allowing for the implementation of hierarchical and explanatory meta-regression structures. We demonstrate the viability of this application through multiple examples developed in JASP and R (via osf.io/bzkpn). First, we will review the DDM and EZ-diff. Then we will introduce our proxy model, and then illustrate how it can be made into a proxy for a Bayesian hierarchical DDM.

The drift diffusion model

The drift diffusion model (DDM) is a cognitive process model that describes binary choice RT data as the result of a stochastic sampling process. The core assumption is that decision-makers accumulate information about stimuli presented, starting the moment they are asked to make a binary judgment and ending once a decision boundary is met. One implication of the principle of accumulation of information (Laming, 1968) is that people make two decisions: when they are ready to respond and what to respond. These aspects are captured by the RT and choice data, respectively.

The three key parameters of the DDM are illustrated in Figure 1. The drift rate parameter ν (nu) indicates the average amount of evidence sampled per unit of time; the boundary separation parameter α (alpha) corresponds to the distance between the two response boundaries; and the non-decision time parameter τ (tau) accounts for the time required to encode and process the information presented by the stimuli. The parameters of the DDM capture relevant aspects of the psychological process that underlies decision-making. These aspects can be manipulated through experimental design: The drift rate is affected by the quality of the information conveyed by the stimuli and by individual differences in processing efficiency. The boundary separation captures the speed-accuracy trade-off imposed by task instructions and serves as an indicator of individual caution. The nondecision time is an indicator of the perceptual complexity of the stimulus (encoding time) and of the response modality (motor response time), and seems to be affected by stimulus complexity (Nunez, Gosai, Vandekerckhove, & Srinivasan, 2019) and participant age (Ratcliff, Thapar, & McKoon, 2001).

The DDM is frequently implemented as a cognitive process model to account for the underlying mechanisms that generate the collected data. In contrast, cognitive psychometricians use the DDM as a statistically tractable measurement tool that can be extended to explore latent variable structures, quantify individual differences, and estimate regression coefficients for covariates of interest (Rouder, Province, Morey, Gomez, & Heathcote, 2014).

¹ Indirect inference has recently come to imply the use of simulated data sets to fit models and estimate parameters, but that is not what we will do here.

Parameter	Process role	Interpretation
Drift rate ν	Evidence accumulation rate	“Ability”
Boundary separation α	Speed-accuracy trade-off	“Caution”
Nondecision time τ	Perceptual encoding and motor execution time	

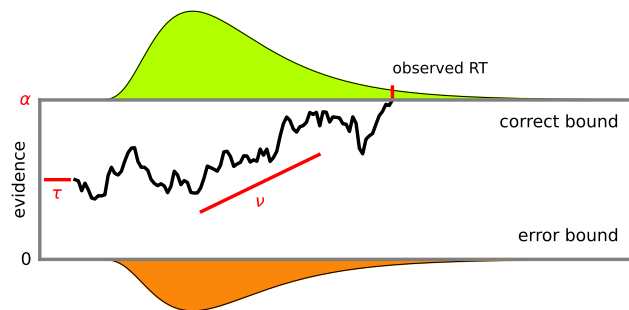


Fig. 1. Evidence accumulation in a two-choice decision task. The jagged line represents the noisy process of evidence accumulation over time until the observed reaction time, when the evidence hits the ‘correct’ decision bound. The distributions are predicted RT distributions for correct decisions (top) and errors (bottom). The drift rate ν captures the participant’s evidence accumulation rate: their ability at the task. The boundary separation α captures the participant’s speed-accuracy trade-off: their caution at the task. The nondecision time τ captures the time spent on processes other than decision making. Figure credit Vandekerckhove and Chávez De la Peña (2023).

In the present paper, we focus on the latter case: the implementation of the DDM as a measurement model to explain variability in the model’s parameters with exogenous predictors. The practical application of a process model as a measurement model will benefit from the use of a Bayesian hierarchical model (Dutilh et al., 2017; Rouder & Lu, 2005; Rouder et al., 2014; Rouder & Haaf, 2019; Schubert, Nunez, Hagemann, & Vandekerckhove, 2019; Vandekerckhove et al., 2011; Villarreal et al., 2024).

Hierarchical cognitive models

Cognitive models are traditionally implemented to describe individual-level performance (e.g., Lee, Newell, & Vandekerckhove, 2014), with group performance being described using aggregate measures that assume low between-subject variability (Estes, 1956). This traditional approach implies a multi-step analysis that starts by estimating individual parameter values and then using summary measures for inference regarding average group performance and individual differences (e.g., Klauer, Voss, Schmitz, & Teige-Mocigemba, 2007). While this two-stage approach is common, it is important to remember that the parameters estimated in the first stage have uncertainty associated with them (either posterior variability or standard error of measurement), and for the purposes of statistical inference this uncertainty needs to be propagated appropriately. Neglecting this uncertainty in the second stage leads to a risk of so-called *generated regressor bias* (Pagan, 1984; Boehm, Marsman, Matzke, & Wagenmakers, 2018).

Hierarchical cognitive models are extensions of cognitive models that model multiple sources of variability simultaneously (e.g., measurement error, variability between individuals, stimuli, experimental conditions, etc.; Coleman, 1964; Clark, 1973). In these models, parameter values across levels of variation are assumed to be sampled from parent distributions (Lee, 2011), such that variability in individual performance is taken into account when estimating population-level statistics (Lee & Webb, 2005; Rouder & Lu, 2005). In the context of nonlinear models of cognition, hierarchical modeling is especially critical because generated regressor bias

may be asymptotic (i.e., the bias does not go to zero even with very large samples; Rouder & Lu, 2005).

While it is possible to implement hierarchical cognitive models under any statistical philosophy, it is often most convenient to make use of flexible Bayesian modeling strategies (Lee & Wagenmakers, 2013), and this is the default practice.

Going forward, we will use the following notational conventions. The choice and RT y_{pi} of person p on trial i is modeled using a Wiener distribution with participant-specific drift rate ν_p , boundary separation α_p , and nondcision time τ_p (Eq. 1). The hierarchical model assumes that individual parameters ν_p , α_p , and τ_p are sampled from parent normal distributions with a mean and variance that describe the population (Eqs. 2, 3, and 4).

$$y_{pi} \sim \text{Wiener}(\alpha_p, \tau_p, \nu_p) \quad [1]$$

$$\nu_p \sim \text{Normal}(\mu_\nu + \beta x_p, \sigma_\nu^2) \quad [2]$$

$$\alpha_p \sim \text{Normal}(\mu_\alpha, \sigma_\alpha^2) \quad [3]$$

$$\tau_p \sim \text{Normal}(\mu_\tau, \sigma_\tau^2) \quad [4]$$

Equation 2 includes by way of example a *metaregression structure* on the drift rate. Here, person-specific drift rates are modeled as samples from normal distributions that are shifted from a shared population mean by an individual predictor x (indexed with p) multiplied by a regression coefficient β . Similar regression structures can be applied to the boundary separation and nondcision time parameters,² and nonlinear regressions may be implemented as well.

The EZ-diffusion model

The “full” DDM accounts for empirical data patterns commonly observed in binary choice tasks through the implementation of trial-by-trial variability in the drift rate, starting point, and nonddecision time (Ratcliff, 1978; Ratcliff & Rouder, 1998; Ratcliff & Tuerlinckx, 2002). However, this seven-parameter version of the DDM is significantly more complex and requires specific modeling expertise to be implemented (Vandekerckhove & Tuerlinckx, 2007). Furthermore, the data patterns specifically captured by the full DDM are elicited only by certain task designs. In many cases—including many of those where the DDM is used as a measurement model—using the full DDM is asking too much of the data (van Ravenzwaaij, Donkin, & Vandekerckhove, 2017).

In cases where it is reasonable to forgo these assumptions of between-trial variability in the parameters of the model, as well as the need for a response bias, parameter estimation is much easier through the ‘EZ’ implementation of the DDM (EZ-diff) introduced by Wagenmakers et al. (2007). EZ-diff was developed by first constructing a system of equations showing how the DDM parameters can be used to compute a predicted accuracy rate R^{pr} and mean M^{pr} and variance V^{pr} of RTs (the ‘forward’ system).³ The insight by Wagenmakers et al. is that the forward system is invertible, so that parameter estimates can be obtained directly from these three summary statistics.

Let $q = \exp(-\alpha\nu)$. The forward EZ equations are then:

$$R^{\text{pr}} = \frac{1}{q + 1} \quad [5]$$

² Although care should be taken with parameters that are naturally bounded to be positive. Sometimes it is preferable to apply a linear regression to $\log(\alpha)$ or $\log(\tau)$ instead.

³ In the original formulation of EZ-diff, a distinction is made between the distribution of ‘correct’ versus ‘incorrect’ RTs. However, under the drift diffusion model with only three parameters, the mean and variance of the decision time do not depend on the response (Stone, 1960), so we can let go of this distinction.

$$M^{\text{pr}} = \tau + \left(\frac{\alpha}{2\nu}\right) \left(\frac{1-q}{1+q}\right), \quad [6]$$

$$V^{\text{pr}} = \left(\frac{\alpha}{2\nu^3}\right) \left\{ \frac{1 - 2\alpha\nu q - q^2}{(q+1)^2} \right\} \quad [7]$$

and parameter estimation is made possible through the following inverse EZ equations:

$$L = \log\left(\frac{R^{\text{ob}}}{1 - R^{\text{ob}}}\right)$$

$$\nu^{\text{es}} = \text{sgn}\left(R^{\text{ob}} - \frac{1}{2}\right) \sqrt[4]{\frac{L(R^{\text{ob}^2}L - R^{\text{ob}}L + R^{\text{ob}} - \frac{1}{2})}{V^{\text{ob}}}} \quad [8]$$

$$\alpha^{\text{es}} = \frac{L}{\nu^{\text{es}}} \quad [9]$$

$$\tau^{\text{es}} = M^{\text{ob}} - \left(\frac{\alpha^{\text{es}}}{2\nu^{\text{es}}}\right) \left[\frac{1 - \exp(-\nu^{\text{es}}\alpha^{\text{es}})}{1 + \exp(-\nu^{\text{es}}\alpha^{\text{es}})}\right]. \quad [10]$$

Here, ν^{es} , α^{es} , and τ^{es} are parameter estimates, R^{ob} is the observed accuracy rate, M^{ob} is the observed mean of the RTs, and V^{ob} is their observed variance.

Sometimes EZ-diffusion is the better model

There are use cases in which we believe EZ-diff should be preferred not because it is easier and faster but because it is the better model.

EZ-diff gained popularity rapidly as a parameter estimation tool due to its simplicity and practicality. Despite initial criticism calling the EZ implementation “too EZ”—for model fit assessments and for obtaining meaningful parameter interpretations when its assumptions are not met, the trial size is limited, or there are outlier RTs (Ratcliff, 2008)—Wagenmakers, van der Maas, Dolan, and Grasman (2008) showed that EZ-diff can be extended to address most of these cases. Moreover, as van Ravenzwaaij et al. (2017) point out, most binary choice tasks do not elicit the data patterns specifically captured by the full DDM in the first place.

Importantly, van Ravenzwaaij and Oberauer (2009) compared the performance of the full and EZ-diff in a parameter recovery simulation study, and demonstrated that EZ-diff can capture individual differences in true data-generating parameters, while the full model failed to recover individual differences in parameters with across-trial variability. Further, van Ravenzwaaij et al. (2017) used the full DDM to generate data emulating experimental effects and then contrasted the ability of EZ-diff to detect those effects against that of the full model, and were able to conclude that EZ-diff provides a powerful test of simple empirical effects. This advantage was seen again in subsequent validation studies (Arnold, Bröder, & Bayen, 2015) and again in a recent many-modelers study (Dutilleul et al., 2019), in which a variation on EZ-diff was the better model in terms of its ability to locate group-level effects. A likely reason for this success is that the mean RT as a summary statistic is less sensitive to random noise in fast RTs, and this robustness is more noticeable in scenarios where the number of trials per participant is low.

Finally, because EZ-diff requires only summary statistics at the individual level, and indeed only requires those summary statistics that are conventionally reported in academic papers, the Bayesian hierarchical extension we will propose next can potentially be applied in a meta-analytic context.

A probabilistic proxy model for the drift diffusion model

Equations 8-10 provide deterministic estimators of the three diffusion model parameters. To implement EZ-diff in a probabilistic programming language, we require a probabilistic estimator – that is, we need a distribution over data that is conditional on the model parameters. Such a distribution can be derived from Equations 5-7 and application of sampling statistics (e.g., Rice, 2006).

If N observations are drawn from a diffusion model whose accuracy rate is R^{pr} , then the sampling distribution of the observed number of correct trials T^{ob} is:

$$T^{ob} \sim \text{Binomial}(R^{pr}, N). \quad [11]$$

Similarly, if N observations are drawn from a sample whose mean and variance of the RTs are M^{pr} and V^{pr} , then the sampling distribution of the observed mean RT M^{ob} is:⁴

$$M^{ob} \sim \text{Normal}\left(M^{pr}, \frac{V^{pr}}{N}\right). \quad [12]$$

Finally, the sampling distribution of the variance of the RTs follows this probability law:⁵

$$\begin{aligned} (N-1) \frac{V^{ob}}{V^{pr}} &\sim \text{Chi-squared}(N-1) \\ \Rightarrow (N-1) \frac{V^{ob}}{V^{pr}} &\sim \text{Gamma}\left(\frac{N-1}{2}, 2\right) \\ \Rightarrow V^{ob} &\sim \text{Gamma}\left(\frac{N-1}{2}, \frac{2V^{pr2}}{N-1}\right). \end{aligned}$$

As N becomes sufficiently large, this is well approximated by a normal distribution:⁶

$$V^{ob} \sim \text{Normal}\left(V^{pr}, \frac{2V^{pr2}}{N-1}\right). \quad [13]$$

Together with Equations 5, 6, and 7, Equations 11, 12, and 13 provide a predictive distribution of three summary statistics in terms of three DDM parameters – that is, a *likelihood*. We call this set of equations our “proxy model” for the DDM. It is easy to see from Equations 11-13 that the EZ-diff estimators in Equations 8-10 are the maximum likelihood estimators of the proxy model.

Equipped with a likelihood for EZ-diff, we can now implement it in a probabilistic programming language like JAGS (Plummer, 2003) or Stan (Carpenter et al., 2017) or PyMC (Abril-Pla et al., 2023). This will allow us, among other things, to use this proxy model as a component of a Bayesian hierarchical model.

The EZ Bayesian hierarchical drift diffusion model

A hierarchical Bayesian extension of EZ-diff allows for a practical and easy implementation of the hierarchical DDM as a measurement model. Equations 1-4 describe the initial hierarchical extension of the DDM (Vandekerckhove et al., 2011). To implement our proxy model, it suffices to substitute the likelihood equation (Eq. 1) with the three equations that constitute our proxy model (Eqs. 11-13). In practical applications, we will also have to adjust data preprocessing code so that mean accuracy and the mean and

⁴Here the normal distribution is expressed with a mean and variance parameter.

⁵Here the Gamma distribution is expressed with a shape and scale parameter.

⁶This approximation is not strictly speaking necessary since we could implement the Gamma distribution directly. However, the Normal distribution affords more flexibility for model-building, and the Gamma is not as numerically stable.

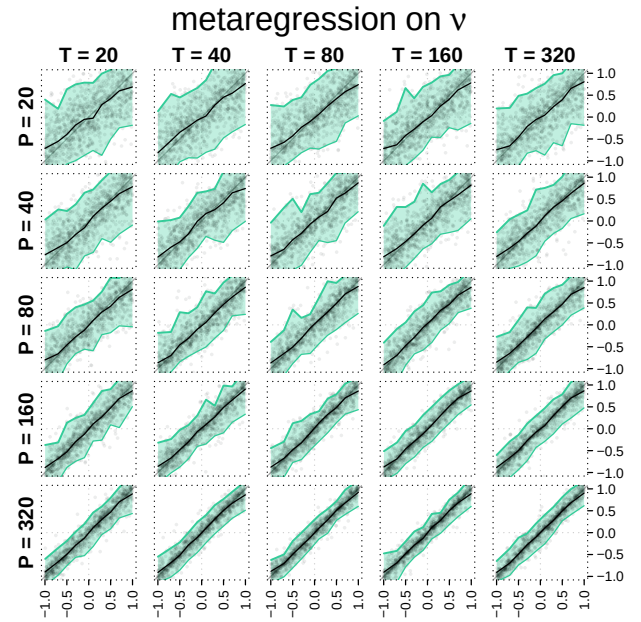


Fig. 2. Partial results from the simulation study. Here, we simulated a design in which participants' individual drift rates were drawn from a normal distribution whose mean was a linear function of some external predictor (as in Eq. 2). Depicted is the recovery of the regression weight β . We believe that testing a relationship between drift rate and an explanatory covariate would be the typical use of EZ-diff in a cognitive-psychometric context. The 25 panels indicate different combinations of number of participants P and number of trials T per participant. In each panel, the 'true' (generated) β value is on the horizontal axis and the mean a posteriori estimate is on the vertical axis. The dark line in each panel is the median of the estimated β s. The shaded area contains 95% of the estimates (with 2.5% of estimates falling above and below). Recovery is unbiased and the variability decreases rapidly with increasing P but only slowly with T . This illustrates the benefit of increasing the number of participants over the number of trials (note that the panels on the diagonal from bottom left to top right show scenarios with the exact same total number of observations, but the variability is clearly lower on the left).

variance of RTs in each design cell are available. An illustrative implementation of the EZ Bayesian hierarchical drift diffusion model in JAGS is given in Appendix A. The prior distributions in that example can be adjusted for specific applications – ours were inspired by Matzke and Wagenmakers (2009)'s review of the literature.

To test the ability of our proxy model to recover DDM parameters, we conducted a large simulation study. Reproducible code for the simulation study can be found via osf.io/bzkpn and more extensive detail is in Appendix B. Briefly, we simulated data sets from a hierarchical simple (3-parameter) diffusion model in which P participants provide T trials in exactly one condition (a between-subjects design). Exactly one of the three DDM parameters was selected as the 'criterion' and made a linear function of an exogenous covariate X (Eq. 2 shows the case where drift rate was the criterion). X was either binary (taking only values of 0 or 1; the “t test” design scenario) or took a value between 0 and 1 (inclusive, the “linear regression” design), yielding six distinct scenarios. Additionally, we varied the number of participants and trials per participant, $P \in \{20, 40, 80, 160, 320\}$ and $T \in \{20, 40, 80, 160, 320\}$, for a total of $6 \times 25 = 150$ conditions. In each condition, we simulated 1,000 data sets and used our proxy method (implemented in JAGS using R2jags; Su & Yajima, 2008) to recover the parameters.

Figure 2 shows partial results from this simulation study, and complete results and full details of the implementation are in Appendix B. Across conditions, we observe that our model estimates

the beta weights well. Recovery is unbiased even in the case with only 20 participants and improves with increasing P . The effect of T on recovery is much weaker. Note that the total number of trials in a simulated experiment (across all participants) is the same between each panel and its neighbor to the right and up (or left and down). This way, the figure also illustrates that greater estimation efficiency is often gained by increasing the sample size P than the trial count T , a known but underappreciated feature of hierarchical models (Cohen, Sanborn, & Shiffrin, 2008; DeKay, Rubinchik, Li, & De Boeck, 2022; Rubinchik, 2019; Vandekerckhove, 2024).

We additionally evaluate the recovery of the population means of the model parameters. The mean drift rate is generally recovered well with no notable bias. The mean boundary separation is recovered well, with a slight bias towards the mean of the prior in the case where we are recovering μ_α as the *intercept* of a regression with $P = 20$. The mean nondecision time shows the same pull towards the mean of the prior even with medium values of P . Moreover, μ_τ is frequently systematically underestimated even with large P and T . Note, however, that the fact that this bias is *systematic* means that the usefulness of the model is preserved – the effect of a covariate predictor is still estimated well.

In summary, the simulation study illustrates good recovery properties—especially of the beta weight parameter—but also shows that our proxy model inherits the estimation biases from EZ-diff. Consequently, we recommend the use of the EZ Bayesian hierarchical drift diffusion model specifically in the cognitive-psychometrical context – that is, scenarios where the interest is not in specific values of DDM parameters, but rather in *regression coefficients that link DDM parameters to external covariates* such as explanatory predictors or elements of an experimental design (i.e., differences between conditions).

Example applications and additional illustrations

To further illustrate the usefulness of our proxy model, we provide a number of worked-out applications. These applications can be found as appendices.

In Appendix C, we conduct a short simulation study with a within-subjects t-test design and illustrate an easy-to-use hypothesis testing procedure with a Bayes factor (Kass & Raftery, 1995). The Bayes factor has excellent inferential properties even for the case with only 20 participants with 20 observations.

In Appendix D, we revisit data from Ratcliff and Rouder (1998) and illustrate a more complex regression case. Despite the complexity, the code is simple enough that it can be implemented in the Bayesian analysis package JASP (Love et al., 2015).

Finally, in Appendix E, we revisit data from Vandekerckhove, Panis, and Wagemans (2007) and illustrate a practical hypothesis testing scenario in an incomplete ANOVA design.

Summary

The drift diffusion model (DDM) is a popular model of choice response time that comes at a significant computational cost. Moreover, many of its applications add computational complexity above and beyond that of calculating the expensive DDM likelihood. Such costly applications have included Bayesian hierarchical extensions (Vandekerckhove et al., 2011), latent variable structures (Vandekerckhove, 2014), applications to large population samples (Lucio et al., 2017; Salum et al., 2014a, 2014b), use of DDMs as a component in numerical experiments (Stafford, Pirrone, Croucher,

& Krystalli, 2020), real-time adaptive design optimization (Bahg et al., 2020), and model comparisons that involve high-dimensional numerical integration (Boehm et al., 2023; Gronau, Heathcote, & Matzke, 2019).

The EZ-diffusion model (EZ-diff; Wagenmakers et al., 2007) was a significant development in this regard, allowing for rapid approximations to DDM parameters with relative ease. However, enhancing EZ-diff with additional statistical structure—such as constraints on parameters over conditions, or introduction of explanatory covariates—was not yet possible. We have introduced the EZ Bayesian hierarchical DDM, a new formulation of EZ-diff as a Bayesian generative model that lends itself to implementation in generic probabilistic programming languages. The derivation was based in standard mathematical statistics – the key equations are Equations 11, 12, and 13. We conducted simulation studies showing good recovery of DDM parameters using our new method, including good recovery of hierarchical regression parameters, and we provide some examples in an online repository (osf.io/bzkpn).

With this new ‘EZ’ formulation, using the DDM is no more costly than using a normal distribution, opening the door to a wide array of modeling applications.

Open practices: Code availability and reproducibility

Our simulation studies can be reproduced using code available via GitHub, github.com/joachimvandekerckhove/ezbhdmm (Python) and github.com/Adrifelcha/EZ-project (R), and archived on the Open Science Framework (osf.io/bzkpn). The python codebase is in the `src/` directory and R/ contains R code. Python and R notebooks to repeat small-scale simulations and create figures are in the `notebooks/` directory. Finally, the `vm/` directory contains set-up information for a virtual machine with a suitable computational environment for reproducibility.

Code and example applications are additionally provided via osf.io/bzkpn.

References

- Abril-Pla, O., Andreani, V., Carroll, C., Dong, L., Fonnesbeck, C. J., Kochurov, M., ... Zinkov, R. (2023, September). Pymc: a modern, and comprehensive probabilistic programming framework in python. *PeerJ Computer Science*, 9, e1516. Retrieved from <http://dx.doi.org/10.7717/peerj-cs.1516> doi: 10.7717/peerj-cs.1516
- Arnold, N. R., Bröder, A., & Bayen, U. J. (2015). Empirical validation of the diffusion model for recognition memory and a comparison of parameter-estimation methods. *Psychological Research*, 79, 882–898.
- Bahg, G., Sederberg, P. B., Myung, J. I., Li, X., Pitt, M. A., Lu, Z.-L., & Turner, B. M. (2020). Real-time adaptive design optimization within functional mri experiments. *Computational Brain & Behavior*, 3(4), 400–429. doi: 10.1007/s42113-020-00079-7
- Batchelder, W. (2016). Cognitive psychometrics. In *Mathematical models of perception and cognition, volume I*. Psychology Press.
- Bitzer, S., Park, H., Blankenburg, F., & Kiebel, S. (2014). Perceptual decision making: drift-diffusion model is equivalent to a Bayesian model. *Frontiers in Human Neuroscience*, 8. doi: 10.3389/fnhum.2014.00102

- Boehm, U., Evans, N. J., Gronau, Q. F., Matzke, D., Wagenmakers, E.-J., & Heathcote, A. J. (2023). Inclusion Bayes factors for mixed hierarchical diffusion decision models. *Psychological Methods*. doi: 10.1037/met0000582
- Boehm, U., Marsman, M., Matzke, D., & Wagenmakers, E.-J. (2018, August). On the importance of avoiding shortcuts in applying cognitive models to hierarchical data. *Behavior Research Methods*, 50(4), 1614–1631. doi: 10.3758/s13428-018-1054-3
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, 76(1).
- Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12(4), 335–359. doi: 10.1016/S0022-5371(73)80014-3
- Cohen, A. L., Sanborn, A. N., & Shiffrin, R. M. (2008). Model evaluation using grouped or individual data. *Psychonomic Bulletin & Review*, 15, 692–712.
- Coleman, E. B. (1964). Generalizing to a Language Population. *Psychological Reports*, 14(1), 219–226. doi: 10.2466/pr0.1964.14.1.219
- De Boeck, P., & Wilson, M. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. New York: Springer.
- DeKay, M. L., Rubinchik, N., Li, Z., & De Boeck, P. (2022). Accelerating psychological science with metastudies: A demonstration using the risky-choice framing effect. *Perspectives on Psychological Science*, 17(6), 1704–1736. doi: 10.1177/17456916221079611
- Dutilh, G., Annis, J., Brown, S., Cassey, P., Evans, N. J., Grasman, R., ... Donkin, C. (2019). The quality of response time data inference: A blinded, collaborative approach to the validity of cognitive models. *Psychonomic Bulletin & Review*, 26, 1051–1069.
- Dutilh, G., Vandekerckhove, J., Ly, A., Matzke, D., Pedroni, A., Frey, R., ... Wagenmakers, E.-J. (2017). A test of the diffusion model explanation for the Worst Performance Rule using preregistration and blinding. *Attention, Perception, and Performance*, 79, 713–725.
- Enkavi, A. Z., Eisenberg, I. W., Bissett, P. G., Mazza, G. L., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019, March). Large-scale analysis of test–retest reliabilities of self-regulation measures. *Proceedings of the National Academy of Sciences*, 116(12), 5472–5477. doi: 10.1073/pnas.1818430116
- Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin*, 53, 134–140.
- Gronau, Q. F., Heathcote, A., & Matzke, D. (2019). Computing Bayes factors for evidence-accumulation models using Warp-III bridge sampling. *Behavior Research Methods*, 52(2), 918–937. doi: 10.3758/s13428-019-01290-6
- Jiang, W., & Turnbull, B. (2004). The indirect method: Inference based on intermediate statistics—a synthesis and examples. *Statistical Science*, 19(2). doi: 10.1214/088342304000000152
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90, 773–795.
- Klauer, K. C., Voss, A., Schmitz, F., & Teige-Mocigemba, S. (2007). Process components of the implicit association test: A diffusion–model analysis. *Journal of Personality and Social Psychology*, 93, 353–368.
- Laming, D. R. J. (1968). *Information theory of choice–reaction times*. London: Academic Press.
- Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology*, 55, 1–7.
- Lee, M. D., Newell, B., & Vandekerckhove, J. (2014). Modeling the adaptation of search termination in human decision making. *Decision*, 1, 223–251.
- Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.
- Lee, M. D., & Webb, M. R. (2005). Modeling individual differences in cognition. *Psychonomic Bulletin & Review*, 12, 605–621.
- Lerche, V., & Voss, A. (2018, March). Speed–accuracy manipulations and diffusion modeling: Lack of discriminant validity of the manipulation or of the parameter estimates? *Behavior Research Methods*, 50(6), 2568–2585. doi: 10.3758/s13428-018-1034-7
- Link, S. W. (1975). The relative judgement theory of two choice response time. *Journal of Mathematical Psychology*, 12, 114–135.
- Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A. J., ... Wagenmakers, E.-J. (2015). JASP [computer software]. <https://jasp-stats.org/>.
- Lucio, P., Salum, G., Rohde, L., Gadelha, A., Swardfager, W., Vandekerckhove, J., ... Cogo-Moreira, H. (2017). Poor stimulus discriminability as a common neuropsychological deficit between ADHD and reading ability in young children: a moderated mediation model. *Psychological Medicine*, 47, 255–266.
- Matzke, D., & Wagenmakers, E.-J. (2009). Psychological interpretation of ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review*, 16, 798–817.
- Mulder, M., van Maanen, L., & Forstmann, B. (2014). Perceptual decision neurosciences – a model-based review. *Neuroscience*, 277, 872–884. doi: 10.1016/j.neuroscience.2014.07.031
- Nunez, M. D., Gosai, A., Vandekerckhove, J., & Srinivasan, R. (2019). The latency of a visual evoked potential tracks the onset of decision making. *Neuroimage*, 197, 93–108.
- Pagan, A. (1984). Econometric issues in the analysis of regressions with generated regressors. *International Economic Review*, 25(1), 221. doi: 10.2307/2648877
- Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In K. Hornik, F. Leisch, & A. Zeileis (Eds.), *Proceedings of the 3rd international workshop on distributed statistical computing*. Vienna, Austria.
- Price, L. F., Drovandi, C. C., Lee, A., & Nott, D. J. (2017). Bayesian synthetic likelihood. *Journal of Computational and Graphical Statistics*, 27(1), 1–11. doi: 10.1080/10618600.2017.1302882
- Rafiei, F., & Rahnev, D. (2021, January). Qualitative speed–accuracy tradeoff effects that cannot be explained by the diffusion model under the selective influence assumption. *Scientific reports*, 11(1), 45. doi: 10.1038/s41598-020-79765-2
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59–108.
- Ratcliff, R. (2008). The EZ diffusion method: Too

- EZ? *Psychonomic Bulletin & Review*, 15(6), 1218–1228. doi: 10.3758/pbr.15.6.1218
- Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, 9, 347–356.
- 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: <https://doi.org/10.1016/j.tics.2016.01.007>
- Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology and aging*, 16(2), 323.
- Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, 9, 438–481.
- Rice, J. A. (2006). *Mathematical statistics and data analysis*. Cengage Learning.
- Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review*, 26(2), 452–467. doi: 10.3758/s13423-018-1558-y
- Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, 12, 573–604.
- Rouder, J. N., Province, J. M., Morey, R. D., Gomez, P., & Heathcote, A. (2014). The lognormal race: A cognitive-process model of choice and latency with desirable psychometric properties. *Psychometrika*, 80(2), 491–513. doi: 10.1007/s11336-013-9396-3
- Rubinchik, N. (2019). *A demonstration of the meta-studies methodology using the risky-choice framing effect*. Retrieved from <https://worldcat.org/title/1159175086> (Unpublished Master's thesis.)
- Salum, G., Sergeant, J., Sonuga-Barke, E., Vandekerckhove, J., Gadelha, A., Pan, P., ... Rohde, L. (2014a). Mechanisms underpinning inattention and hyperactivity: neurocognitive support for ADHD dimensionality. *Psychological Medicine*, 44, 3189–3201.
- Salum, G., Sergeant, J., Sonuga-Barke, E., Vandekerckhove, J., Gadelha, A., Pan, P., ... Rohde, L. (2014b). Specificity of basic information processing and inhibitory control in attention deficit/hyperactivity disorder. *Psychological Medicine*, 44, 617–631.
- Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H.-M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General*, 136, 414–429.
- Schubert, A.-L., Nunez, M. D., Hagemann, D., & Vandekerckhove, J. (2019). Individual differences in cortical processing speed predict cognitive abilities: A model-based cognitive neuroscience account. *Computational Brain & Behavior*, 2, 64–84.
- Stafford, T., Pirrone, A., Croucher, M., & Krystalli, A. (2020). Quantifying the benefits of using decision models with response time and accuracy data. *Behavior Research Methods*, 52(5), 2142–2155. doi: 10.3758/s13428-020-01372-w
- Stone, M. (1960). Models for choice–reaction time. *Psychometrika*, 25, 251–260.
- Su, Y.-S., & Yajima, M. (2008, October). *R2jags: Using r to run "jags"*. The R Foundation. Retrieved from <http://dx.doi.org/10.32614/CRAN.package.R2jags> doi: 10.32614/cran.package.r2jags
- Tal, E. (2020). Measurement in Science. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2020 ed.). Metaphysics Research Lab, Stanford University. Retrieved from <https://plato.stanford.edu/archives/fall2020/entries/measurement-science/>
- Tuerlinckx, F., Maris, E., Ratcliff, R., & De Boeck, P. (2001). A comparison of four methods for simulating the diffusion process. *Behavior Research Methods, Instruments, & Computers*, 33, 443–456.
- Vandekerckhove, J. (2014). A cognitive latent variable model for the simultaneous analysis of behavioral and personality data. *Journal of Mathematical Psychology*, 60, 58–71.
- Vandekerckhove, J. (2024). Commensurability engineering is first and foremost a theoretical exercise. *Behavioral and Brain Sciences*, 47, e63.
- Vandekerckhove, J., & Chávez De la Peña, A. F. (2023). *Simple ddm figure*. Code available via: <https://colab.research.google.com/drive/1onDh7Q3aPyLU8PK8V0Xh7oVbpdchxbs>.
- Vandekerckhove, J., Panis, S., & Wagemans, J. (2007). The concavity effect is a compound of local and global effects. *Perception & Psychophysics*, 69, 1253–1260.
- Vandekerckhove, J., & Tuerlinckx, F. (2007). Fitting the Ratcliff diffusion model to experimental data. *Psychonomic Bulletin & Review*, 14, 1011–1026.
- Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2011). Hierarchical diffusion models for two-choice response times. *Psychological Methods*, 16, 44–62.
- Vandekerckhove, J., Verheyen, S., & Tuerlinckx, F. (2010). A crossed random effects diffusion model for speeded semantic categorization data. *Acta Psychologica*, 133, 269–282.
- 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.
- van Ravenzwaaij, D., & Oberauer, K. (2009). How to use the diffusion model: Parameter recovery of three methods: EZ, fast-dm, and DMAT. *Journal of Mathematical Psychology*, 53(6), 463–473. doi: 10.1016/j.jmp.2009.09.004
- Villarreal, J. M., Chávez De la Peña, A. F., Mistry, P., Menon, V. E., Vandekerckhove, J., & Lee, M. D. (2024). Bayesian graphical modeling with the circular drift diffusion model. *Computational Brain & Behavior*, 7, 181–194.
- Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition*, 32, 1206–1220.
- Wagenmakers, E.-J. (2009). Methodological and empirical developments for the Ratcliff diffusion model of response times and accuracy. *European Journal of Cognitive Psychology*, 21, 641–671.
- Wagenmakers, E.-J., van der Maas, H. J. L., & Grasman, R. P. P. P. (2007). An EZ–diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, 14, 3–22.
- Wagenmakers, E.-J., van der Maas, H. J. L., Dolan, C. V., & Grasman, R. P. P. P. (2008). EZ does it! extensions of the EZ-diffusion model. *Psychonomic Bulletin & Review*, 15(6), 1229–1235. doi: 10.3758/pbr.15.6.1229
- Wood, S. N. (2010). Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, 466(7310), 1102–1104. doi: 10.1038/nature09319

A. Illustrative JAGS code

The following JAGS code implements the EZ Bayesian hierarchical drift diffusion model, with drift rate as the criterion of a regression with predictor X . The code consists of a block of priors, a block for the hierarchical level with the regression component, a block with the forward EZ equations, and a block with the sampling distributions for each summary statistic. Note that JAGS uses precision (inverse variance) as the second parameter of `dnorm()`.

The code is written to expect a data set with five variables, all vectors of length equal to the number of cells in the study design (here, there is one cell per participant because the design is fully between participants):

- `nTrials`: the total number of trials in this cell
- `correct`: the number of correct responses
- `meanRT`: the mean of the RTs (in seconds)
- `varRT`: the variance of the RTs (in seconds squared)
- `X`: the predictor

In the remaining appendices, we provide more complex examples. All of our code, including applications imported into JASP (Love et al., 2015), can be accessed via osf.io/jstgw.

```
## Illustrative JAGS code
model{
  bound_mean ~ dnorm(2.25,pow(1,-2))T(0.10,)
  nondt_mean ~ dnorm(0.55,pow(0.25,-2))T(0.05,)
  drift_mean ~ dnorm(0,pow(3,-2))
  bound_sdev ~ dunif(0.01,2)
  nondt_sdev ~ dunif(0.01,0.5)
  drift_sdev ~ dunif(0.01,2)
  betaweight ~ dnorm(0,pow(1,-2))

  for(p in 1:nParticipants){
    # Hierarchical distributions of diffusion
    # model parameters, with drift rate criterion.
    drift[p] ~ dnorm(drift_mean + betaweight*X[p],
                    pow(drift_sdev,-2))
    bound[p] ~ dnorm(bound_mean,
                    pow(bound_sdev,-2))T(0.10,)
    nondt[p] ~ dnorm(nondt_mean,
                    pow(nondt_sdev,-2))T(0.05,)

    # Forward equations from EZ-diffusion
    ey[p] = exp(-bound[p] * drift[p])
    Pc[p] = 1 / (1 + ey[p])
    PRT[p] = 2 * pow(drift[p], 3) / bound[p] *
             pow(ey[p] + 1, 2) / (2 * -bound[p] *
             drift[p] * ey[p] - ey[p]*ey[p] + 1)
    MDT[p] = (bound[p] / (2 * drift[p])) *
             (1 - ey[p]) / (1 + ey[p])
    MRT[p] = MDT[p] + nondt[p]

    # Sampling distributions
    correct[p] ~ dbin(Pc[p], nTrials)
    meanRT[p] ~ dnorm(MRT[p], PRT[p] * nTrials)
    varRT[p] ~ dnorm(1/PRT[p], 0.5 * (nTrials-1) *
                    PRT[p] * PRT[p])
  }
}
```

B. Parameter recovery simulation studies

We present the results of six simulation studies to test the ability of our proxy model to recover the parameters of the hierarchical DDM. In all studies, we simulated two-alternative forced-choice data from P participants $P \in \{20, 40, 80, 160, 320\}$, each providing T trials $T \in \{20, 40, 80, 160, 320\}$. In all cases reported below, we simulated a between-subjects design.

Sampling procedure. All simulated participants had a randomly chosen drift rate ν_p , boundary separation α_p , and nondecision time τ_p . The population-level distributions of all parameters were normal (see Eqs. 2-4). In all simulations, we chose one parameter $\theta \in \{\nu, \alpha, \tau\}$ as the ‘criterion’ and made its population mean a linear function of a single person-level predictor x_p :

$$\theta_p \sim N(\mu_\theta + \beta x_p, \sigma_\theta^2).$$

Here, the beta weight β is the parameter of greatest interest, as it describes the prediction of θ by x_p (i.e., it is an external covariate).

In the six simulation studies we describe below, θ was one of ν , α , or τ , and x_p was either binary (the “t test” design scenario) or took a value between 0 and 1 (inclusive; the “linear regression” design). Every study consisted of 1,000 iterations in which a data set was generated from these individual-level sampled parameter values.

Each iteration of each simulation study began by drawing a random set of hierarchical mean parameters:

$$\begin{aligned}\mu_\alpha &\sim U(0.50, 4.00) \\ \mu_\nu &\sim U(-5.50, 5.50) \\ \mu_\tau &\sim U(0.15, 0.50)\end{aligned}$$

A set of hierarchical standard deviation parameters σ_α , σ_ν and σ_τ was then defined to be proportionate to the hierarchical mean parameters sampled. This manipulation served the purpose of ensuring lower-value hierarchical means were not matched with higher-value standard deviations.

The beta weight parameter β to be used in any iteration was also sampled from a uniform distribution. For simulations with α or ν as the criterion, the range was $(-1, 1)$. When τ was the criterion, β was sampled from a $(0, 1)$ uniform interval.

We then drew P person-specific diffusion model parameters from the population-level distributions defined by these hierarchical parameters. Finally, for each person, we drew T realizations of the diffusion process using the rejection-based algorithm described in Tuerlinckx, Maris, Ratcliff, and De Boeck (2001). EZ-diff was developed from the simple DDM with no initial bias and no between-trial variability parameters. As such, the data simulated for every participant did not include any form of between-trial variability.

Parameter recovery. For parameter recovery, we drew samples from the posterior distribution of each of the hierarchical parameters using the JAGS code in Appendix A (with the regression structure applied to whichever parameter was the criterion in that simulation). For each simulation study, we drew three MCMC chains with 2,500 samples after a 500-sample burn-in period.

Speed. The posterior sampling, together with standard postprocessing in R, generally took between four and five seconds per iteration on an off-the-shelf desktop computer.

Evaluation. We evaluate the recovery by plotting true values on the horizontal axis and mean a posteriori estimates on the vertical axis in Figures 3 (for β), 4 (for μ_ν), 5 (for μ_α), and 6 (for μ_τ). Each figure contains six panels of 25 (5×5) scatter plots. The panels in the left column refer to the “t test” designs and those in the right column to the “linear regression” designs. The three rows of panels indicate the cases with the drift rate ν as criterion (top row), the boundary separation α as criterion (middle row), and the nondecision time τ as criterion (bottom row). Better recovery is signaled by points falling closer to the gray dashed diagonal. The shaded areas denote 95% of the estimates obtained across different regions of the true parameter space (2.5% falling on either side), with the thick black lines indicating the median estimates.

Note that Figures 4, 5, and 6 show the recovery of parameter-specific hierarchical means (μ_ν , μ_α , and μ_τ , respectively) and that the role of these parameters depends on the design. That is, in Figure 4, the panels on the top row (“t test on ν ” and “metaregression on ν ”) show the recovery of μ_ν , which in these cases is the intercept of the linear function $\mu_\nu + \beta x_p$. The same is true for the panels in the middle row in Figure 5 and the bottom row in Figure 6.

The 25 scatter plots differ in the number of simulated participants P and trials per participant T . As expected, precision of recovery improves as P and T increase. The values of P and T were chosen so that the total number of observations is equal between each scatter plot and its neighbor to the top right or bottom left. This helps illustrate that recovery of the beta weights improves faster with P than with T (Cohen et al., 2008). For small P and T we sometimes see a strong effect of the prior (i.e., extreme values are pulled towards the middle), and for mean nondecision time we see occasional large overestimates.

In summary, the simulation studies illustrate good recovery properties of the mean drift rate parameter and the beta weight parameter, but also shows that our proxy model inherits the estimation biases from EZ-diff.

C. A brief simulation exercise on hypothesis testing

We conducted a simulation study with a within-subjects t-test design and two experimental conditions. Data was generated according to the following hierarchical structure:

$$\begin{aligned}\nu_{p,k} &\sim \text{Normal}(\mu_\nu + \beta X_k, 0.25) \\ \alpha_p &\sim \text{Normal}(\mu_\alpha, \sigma_\alpha) \\ \tau_p &\sim \text{Normal}(\mu_\tau, \sigma_\tau) \\ \mathbf{y}_{p,k,i} &\sim \text{Wiener}(\alpha_i, \tau_i, \nu_{i,k}),\end{aligned}$$

so that choice and RT data $\mathbf{y}_{p,k,i}$ from trial i of participant p in condition $k \in \{1, 2\}$ was simulated from a simple diffusion process with participant-specific boundary separation α_p and nondecision time τ_p parameters, and participant-by-condition specific drift rates $\nu_{i,k}$. We used the indicator variable $X_k : \{X_1 = 0, X_2 = 1\}$ to enforce a t-test design on the drift parameter and a beta weight β parameter to capture the within-participant differences in performance across the two conditions.

We generated 1,000 data sets with 20 participants and 20 trials per condition across three fixed effects conditions: $\beta \in \{0.0, 0.2, 0.4\}$. Each iteration of the simulation study began by sampling hierarchical parameters from suited uniform distributions. From there, participant and participant-by-condition specific parameters were generated to simulate the data.

```
## JAGS code for hypothesis testing simulation
model{
  bound_mean ~ dnorm(2.25,pow(1,-2))T(0.10,)
  bound_sdev ~ dunif(0.01,1)
  nondt_mean ~ dnorm(0.55,pow(0.25,-2))T(0.05,)
  nondt_sdev ~ dunif(0.01,0.5)
  drift_mean ~ dnorm(0,pow(3,-2))
  drift_sdev ~ dunif(0.01,1)
  betaweight ~ dnorm(0,pow(1,-2))

  # Participant-level parameters
  for(p in 1:nParticipants) {
    bound[p] ~ dnorm(bound_mean,
                     pow(bound_sdev, -2))T(0.10,)
    nondt[p] ~ dnorm(nondt_mean,
                    pow(nondt_sdev, -2))T(0.05,)
    # Participant-by-condition parameters
    for(j in 1:2){
      drift[p,j] ~ dnorm(drift_mean + betaweight
                        * (j-1), pow(drift_sdev, -2))
    }
  }

  for (k in 1:length(meanRT)) {
    ey[k] = exp(-bound[P[k]] *
                drift[P[k],(X[k]+1)])
    Pc[k] = 1 / (1 + ey[k])
    PRT[k] = 2 * pow(drift[P[k],(X[k]+1)], 3) /
             bound[P[k]] * pow(ey[k] + 1, 2) / (2 * -
             bound[P[k]] * drift[P[k],(X[k]+1)] * ey[k]
             - ey[k]*ey[k] + 1)
    MDT[k] = (bound[P[k]] / (2 * drift[P[k],(X[k]
             +1)])) * (1 - ey[k]) / (1 + ey[k])
    MRT[k] = MDT[k] + nondt[P[k]]

    correct[k] ~ dbin(Pc[k], nTrialsPerCondition)
    meanRT[k] ~ dnorm(MRT[k], PRT[k] *
                     nTrialsPerCondition)
    varRT[k] ~ dnorm(1/PRT[k], 0.5*(
                     nTrialsPerCondition-1) * PRT[k] * PRT[k])
  }
}
```

Estimation. Once the trial-by-trial data had been generated, we computed the EZ-diff summary statistics for each participant-by-condition design cell. Estimation was conducted using the JAGS model below. We tested convergence of the MCMC algorithm by confirming that the potential scale reduction factor $\hat{R}$ was less than 1.05. If the algorithm failed to converge (approx. 0.01% of iterations), the iteration was discarded.

Speed. When fitting the JAGS model, we drew two MCMC chains with 1,500 samples after a 500-sample burn-in period. The posterior sampling and standard postprocessing in R took approximately five seconds per iteration on an off-the-shelf desktop computer.

Hypothesis testing procedure. The goal of this short simulation study is to demonstrate a simple Bayesian hypothesis testing procedure. We are interested in testing whether the simulated experimental condition had an effect on the population mean drift

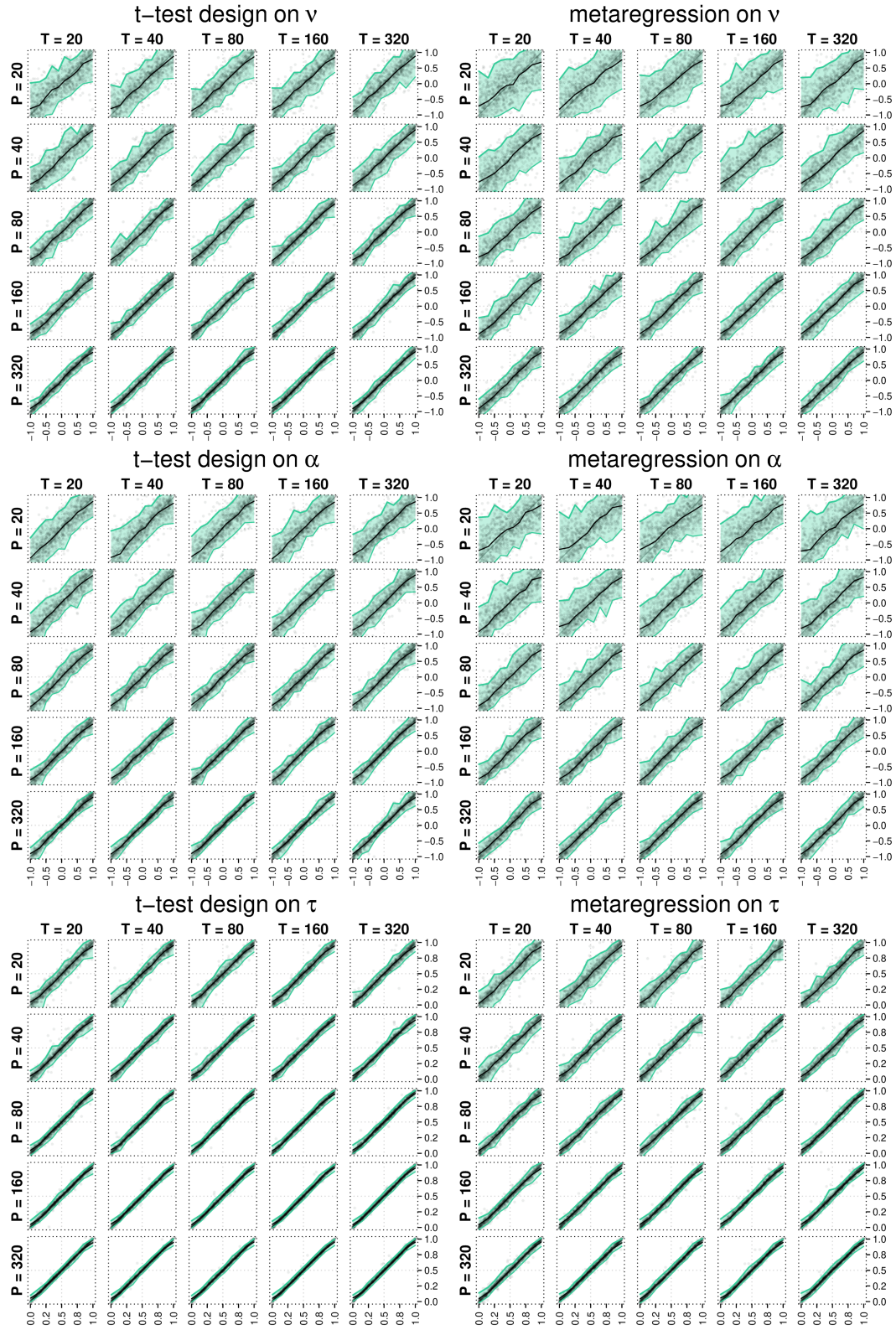


Fig. 3. Recovery of the β parameter across simulation studies (true values on the horizontal axis). The left three panels show studies with 't test' designs, and the right panels show 'linear regression' designs. Rows correspond to different criterion parameters: drift rate (ν , top row), boundary separation (α , middle row), and nondesicion time (τ , bottom). Shaded areas indicate where 95% of the estimated values fell, and the thick black lines represent median recovered values.

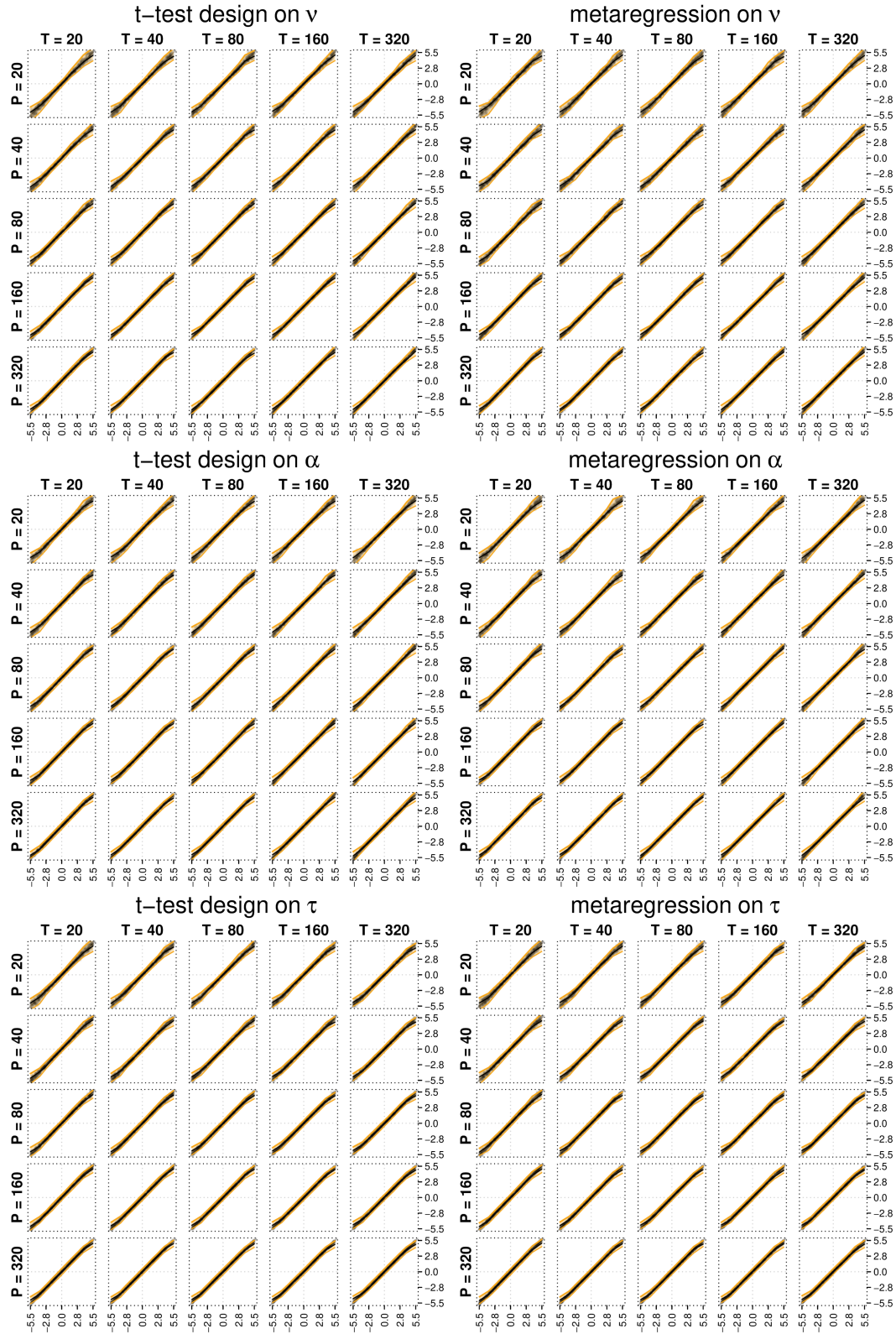


Fig. 4. Recovery of the population mean drift rate μ_{ν} parameter across simulation studies (true values on the horizontal axis). The left three panels show studies with 't test' designs, and the right panels show 'linear regression' designs. Rows correspond to different criterion parameters: drift rate (ν , top row), boundary separation (α , middle row), and nondesicision time (τ , bottom). Shaded areas indicate where 95% of the estimated values fell, and the thick black lines represent median recovered values.

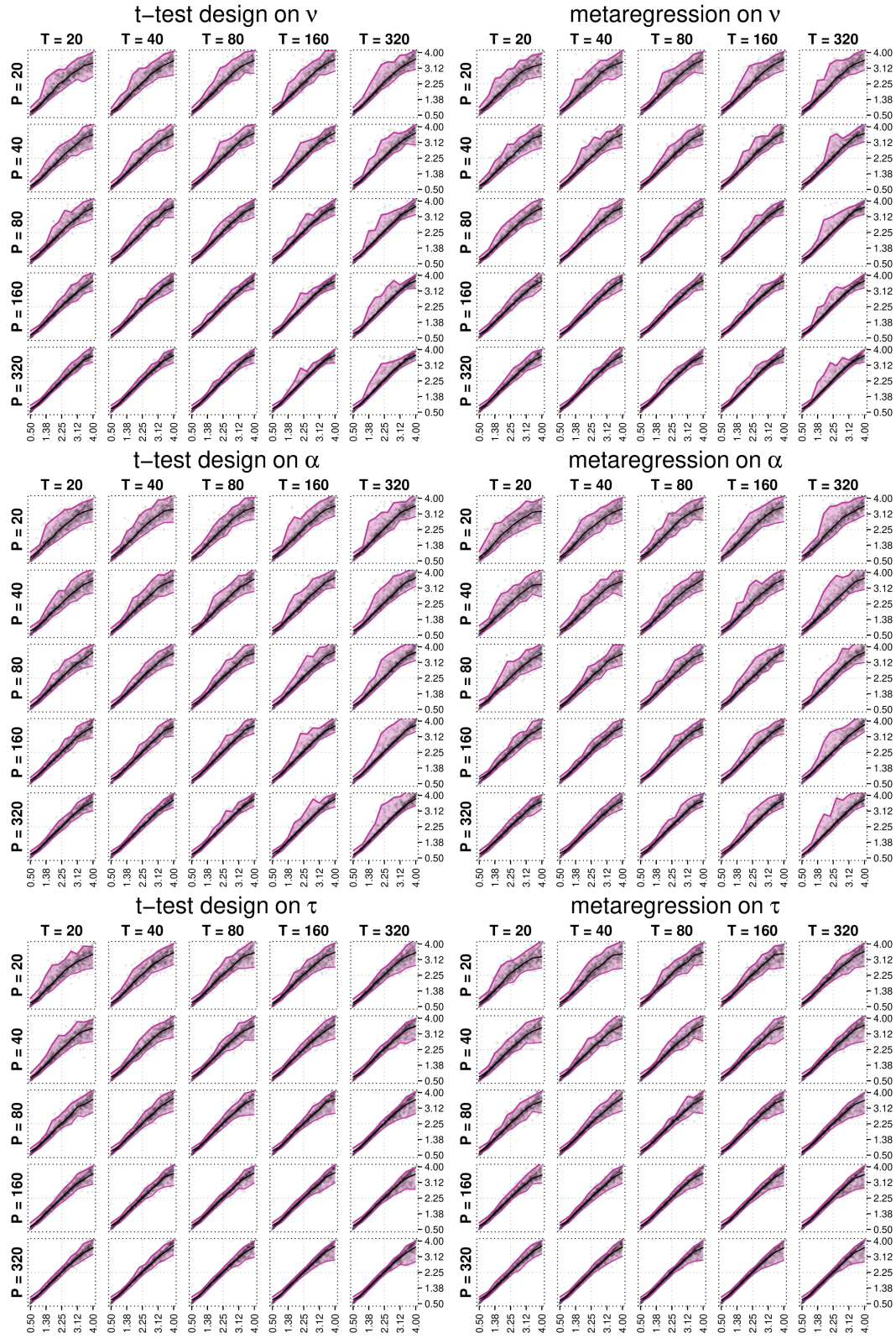


Fig. 5. Recovery of the population mean boundary separation μ_{α} parameter across simulation studies (true values on the horizontal axis). The left three panels show studies with 't test' designs, and the right panels show 'linear regression' designs. Rows correspond to different criterion parameters: drift rate (ν , top row), boundary separation (α , middle row), and nondcision time (τ , bottom). Shaded areas indicate where 95% of the estimated values fell, and the thick black lines represent median recovered values.

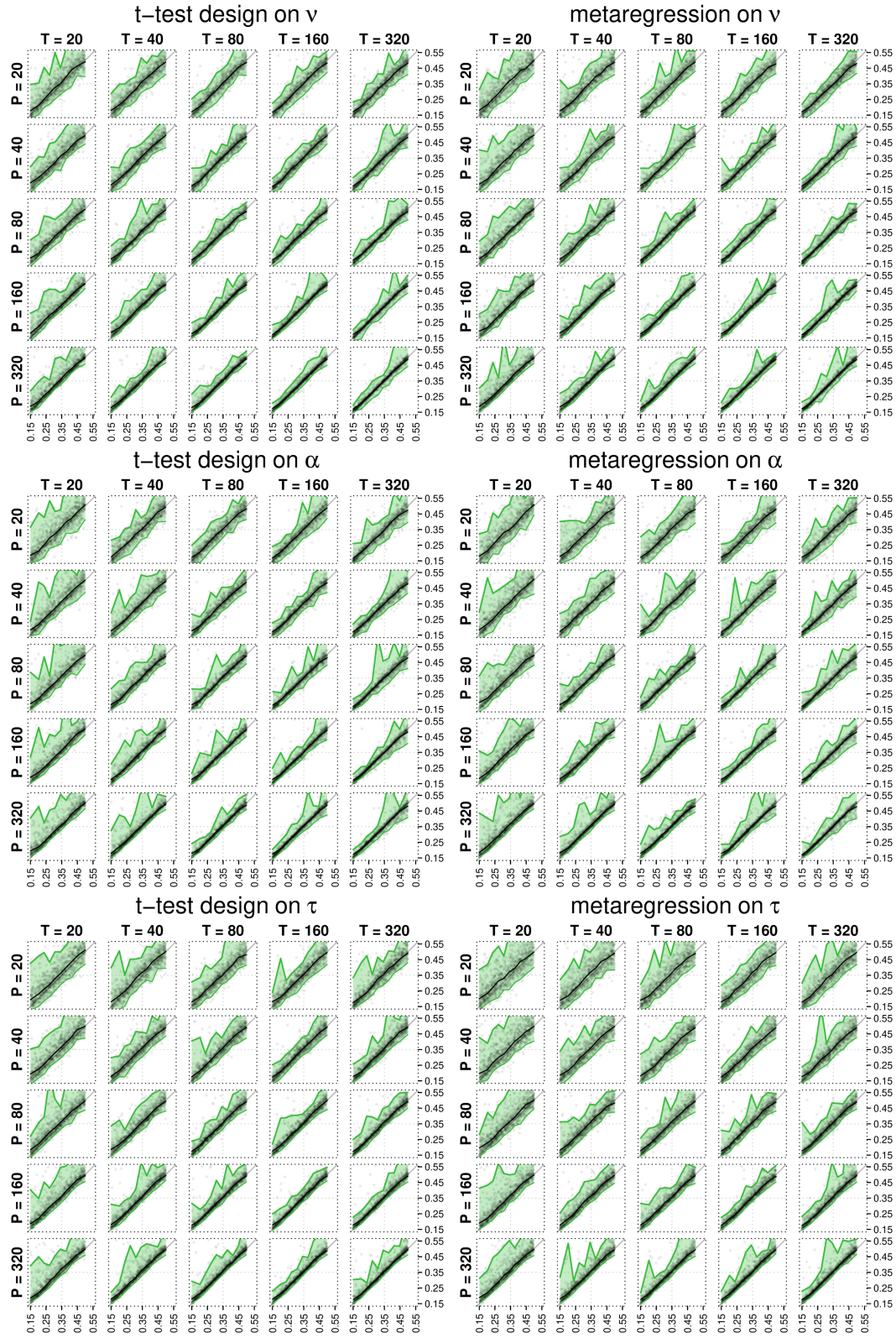


Fig. 6. Recovery of the population mean nondesign time μ_τ parameter across simulation studies (true values on the horizontal axis). The left three panels show studies with 't test' designs, and the right panels show 'linear regression' designs. Rows correspond to different criterion parameters: drift rate (ν , top row), boundary separation (α , middle row), and nondesign time (τ , bottom). Shaded areas indicate where 95% of the estimated values fell, and the thick black lines represent median recovered values.

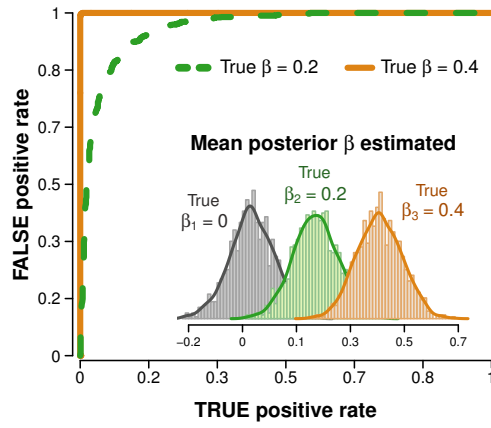


Fig. 7. Main panel: Receiver operating characteristic curves illustrating the trade-off between true positive rate (vertical axis) and false positive rate (horizontal axis) for different Bayes factor thresholds (solid line: true $\beta = 0.4$; dashed line: true $\beta = 0.2$). **Inset:** Distributions of estimated β parameters for each condition.

rate. For each iteration in our simulation study, we compute a Bayes factor (Kass & Raftery, 1995) to evaluate the constraint that $\beta \approx 0$. The Bayes factor favoring the alternative is simply the ratio between the prior mass near 0 and the posterior mass near zero:

$$B = \frac{P_{\text{prior}}(\beta \in (-\varepsilon, \varepsilon))}{P_{\text{posterior}}(\beta \in (-\varepsilon, \varepsilon))},$$

for some small ε (we used 0.1). We then decide that β is nonzero if B exceeds some threshold that is determined by the strength of evidence we need. In Figure 7, we show the resulting receiver operating characteristic (ROC) curve. Even with this small data size, the ROC curve demonstrates excellent sensitivity and selectivity for β .

D. Applied example: Metaregression

In this example, we showcase the advantages of implementing our proxy model with a meta-regression extension that involves a non-linear function. We re-analyze data from Ratcliff and Rouder (1998), who conducted a numerosity study in which participants had to identify the brightness of pixel arrays shown on screen as “high” or “low.” The pixel arrays varied across 33 configuration levels that differed in the proportion of white pixels, with the first 16 levels indicating a majority of black pixels. The task incorporated two instruction conditions that primed participants to prioritize the speed or the accuracy of their responses. We analyze the data from a single participant, excluding the stimulus configuration with an equal number of black and white pixels (for which there is no correct response).

The model. The model has two components. The main component is the proxy model built from the sampling distributions for the summary statistics computed from the data. The second component is a meta-regression structure with which we incorporate an effect β of instruction X_i on the boundary separation α ,

$$\alpha_i \sim \text{Normal}(\mu_\alpha + \beta X_i, \sigma_\alpha).$$

and a nonlinear regression on the drift rate ν using instruction X_i and stimulus configuration Z_s as predictors:

$$\begin{aligned} S_{i,s} &= \Phi(\beta_1 + \beta_2 |Z_s| + \beta_3 X_i |Z_s|) \\ \nu_{i,s}^{\text{pred}} &= \mu_\nu + \beta_0 S_{i,s} + \beta_4 X_i \\ \nu_{i,s} &\sim \text{Normal}(\nu_{i,s}^{\text{pred}}, \sigma_\delta). \end{aligned}$$

Here, Z_s is the difference between the number of black and white pixels (standardized). We use the absolute value $|Z_s|$ since we expect task difficulty to change as the proportion of white pixels departs from 50%, regardless of the direction of the difference.

```
## JAGS code for metaregression application
model {
  betaweight ~ dnorm(0.00, 1.00)
  beta0 ~ dnorm(0.00, 1.00)
  beta1 ~ dnorm(0.00, 1.00)
  beta2 ~ dnorm(0.00, 1.00)
  beta3 ~ dnorm(0.00, 1.00)
  beta4 ~ dnorm(0.00, 1.00)
  bound_mean ~ dnorm(1.50, (0.20^-2))T( 0.10, 3.00)
  drift_mean ~ dnorm(0.50, (0.50^-2))
  nondt_mean ~ dnorm(0.30, (0.06^-2))T( 0, )
  bound_sdev ~ dunif(0.01, 1.00)
  drift_sdev ~ dunif(0.01, 3.00)
  nondt_sdev ~ dunif(0.01, 0.50)

  # Hierarchical distributions of diffusion model
  # parameters, with drift rate criterion.
  for (p in 1:length(meanRT)) {
    drift_pred[p] = drift_mean + beta0 * phi(beta1
      + beta2*abs(Xs[p]) + beta3*Xi[p]*abs(Xs[p]))
      + beta4 * Xi[p]
    drift[p] ~ dnorm(drift_pred[p], (drift_sdev^-2))
    bound_pred[p] = bound_mean + betaweight * Xi[p]
    bound[p] ~ dnorm(bound_pred[p], (bound_sdev^-2))T
      ( 0.10, 3.00)
    nondt[p] ~ dnorm(nondt_mean, (nondt_sdev^-2))T
      ( 0.05, )

    # Forward equations from EZ DDM
    ey[p] = exp(-bound[p] * drift[p])
    Pc[p] = 1 / (1 + ey[p])
    PRT[p] = 2 * pow(drift[p], 3) / bound[p] * pow(
      ey[p] + 1, 2) / (2 * -bound[p] * drift[p] *
      ey[p] - ey[p] * ey[p] + 1)
    MDT[p] = (bound[p] / (2*drift[p])) * (1 - ey[p])
      / (1 + ey[p])
    MRT[p] = MDT[p] + nondt[p]

    # Sampling distributions for summary statistics
    correct[p] ~ dbin(Pc[p], nTrials[p])
    varRT[p] ~ dnorm(1/PRT[p], 0.5*(nTrials[p]-1) *
      PRT[p] * PRT[p])
    meanRT[p] ~ dnorm(MRT[p], PRT[p] * nTrials[p])

    # Noiseless predictions for figures
    ey_pred[p] = exp(-bound_pred[p] * drift_pred[p])
    Pc_pred[p] = 1 / (1 + ey_pred[p])
    PRT_pred[p] = 2 * pow(drift_pred[p], 3) /
      bound_pred[p] * pow(ey_pred[p]+1, 2) / (2 *
      -bound_pred[p] * drift_pred[p] * ey_pred[p]
      - ey_pred[p] * ey_pred[p] + 1)
    MDT_pred[p] = (bound_pred[p] / (2 * drift_pred[p]
      )) * (1 - ey_pred[p]) / (1 + ey_pred[p])
    MRT_pred[p] = MDT_pred[p] + nondt_mean
  }
}
```

Results. We use the proposed metaregression structure to explore the effect of instruction on the drift rate. Figure 8 presents the posterior distributions for the regression coefficients of interest: the effect on slope β_3 and the main effect β_4 of instruction on the drift rate parameter. Both panels show very low posterior density near zero, suggesting that the instruction condition had an effect on the drift rate parameter.

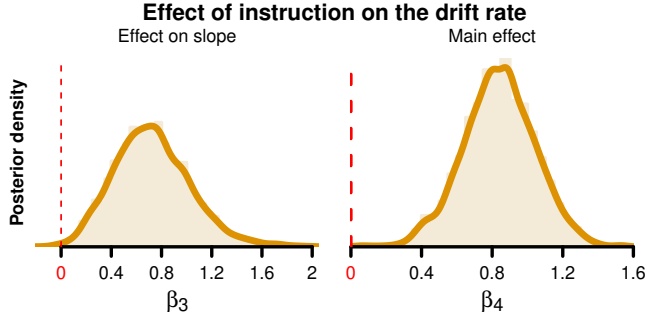


Fig. 8. Posterior distributions for the effect on slope (β_3 , left panel) and the main effect (β_4 , right panel) of the instruction condition on the drift rate. The red vertical lines indicate the 'no-effect' values $\beta_3 = 0$ and $\beta_4 = 0$.

Figure 9 shows the posterior mean predicted drift rates $\nu_{i,s}^{\text{pred}}$ (thick, orange line) and the posterior mean estimated drift rates $\nu_{i,s}$ (brown markers) with their respective 95% CI (brown whiskers) for each stimulus configuration level s and instruction condition i . As expected, drift rates decrease in both instruction conditions as the stimulus configuration approaches the 50/50 condition, where the task is more difficult. Moreover, as suggested by Figure 8, drift rates seem to be shifted upwards in the 'Speed' instruction condition, indicating an effect of instruction condition on drift rate.

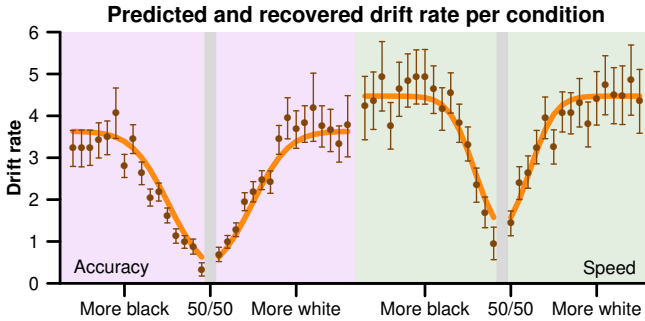


Fig. 9. Predicted $\nu_{i,s}^{\text{pred}}$ and estimated $\nu_{i,s}$ drift rates across conditions. The thick orange lines show the posterior mean predicted drift rates $\nu_{i,s}^{\text{pred}}$ for every stimulus configuration level across instruction conditions. The brown dots correspond to the posterior mean drift rates estimated $\nu_{i,s}$, with the brown whiskers indicating their corresponding 95% CI. The 'Accuracy' and 'Speed' instruction conditions are shown on the left and right panels, respectively. The vertical gray bars indicate the 50/50 black-and-white condition omitted from our analysis for lack of a 'correct' response.

Finally, to demonstrate the adequacy of our meta-regression model, in Figure 10 we present the posterior predictive checks obtained with respect to all three EZ summary statistics. In all panels, the thick line and shaded region indicate the mean posterior prediction and 95% CI, respectively, with the black dots corresponding to the summary statistics computed from the data for each stimulus configuration and instruction condition. Overall, our model provides an adequate account of the data analyzed.

Speed. This analysis with 7,802 observations after data cleaning took 8.6 seconds on an off-the-shelf desktop computer.

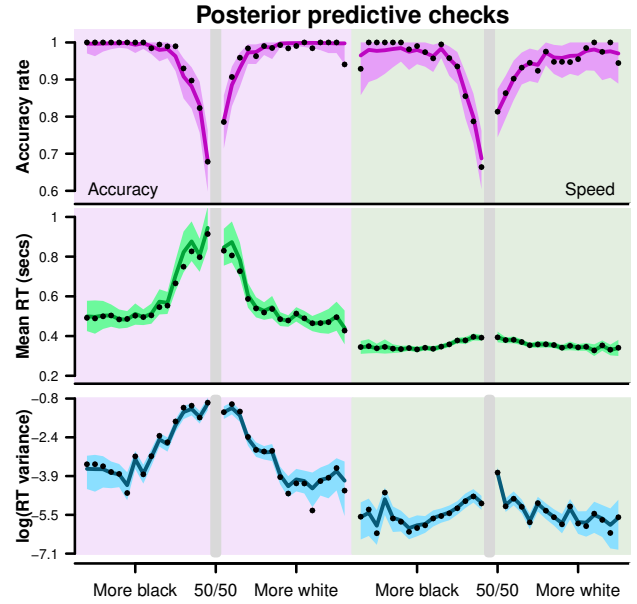


Fig. 10. Posterior predictive checks for all EZ summary statistics: the accuracy rate (top panel), the mean RT in seconds (middle panel), and the RT variance expressed in a log scale (bottom panel). All panels are split by instruction condition, with the 'Accuracy' condition on the left and the 'Speed' condition on the right. Thick colored lines show the mean posterior predictions for each stimulus configuration level and instruction condition, with the shaded area marking the 95% CI. The black markers indicate the actual summary statistics computed from the data analyzed.

E. Applied example: Hypothesis testing

This example features an implementation of our proxy model with a meta-regression extension that uses a multiple linear regression function for hypothesis testing. We re-analyze the data from nine participants in a shape perception study (Vandekerckhove et al., 2007), who were asked to compare pairs of irregular shapes in a "Same/Different" task. The experimental design included three factors: (1) whether there was a change between the shapes, and if so (2) what was the change type (did it affect a *convexity* or *concavity*); and (3) the change quality (did it introduce a *qualitatively* new vertex or did it *quantitatively* change an existing one). Due to the incompletely crossed design, this leads to five experimental conditions (see Tab. 1).

Table 1. Experimental conditions in Vandekerckhove et al. (2007).

Condition	Change (A)	Change quality (B)	Change type (C)
$k = 1$	Yes ($A = 1$)	Qualitative ($B = 0$)	Convexity ($C = 0$)
$k = 2$	Yes ($A = 1$)	Quantitative ($B = 1$)	Convexity ($C = 0$)
$k = 3$	Yes ($A = 1$)	Qualitative ($B = 0$)	Concavity ($C = 1$)
$k = 4$	Yes ($A = 1$)	Quantitative ($B = 1$)	Concavity ($C = 1$)
$k = 5$	No ($A = 0$)	n/a	n/a

The model. We explore the variability in the drift rate parameter ν_k across conditions k with the following multiple linear regression model:

$$\nu_k^{\text{pred}} = \mu + A_k(\gamma_1 B_k + \gamma_2 C_k + \gamma_3 B_k C_k) + (1 - A_k)\gamma_4$$

$$\nu_k \sim \text{Normal}(\nu_k^{\text{pred}}, \sigma_\nu).$$

The predicted drift rate ν_k^{pred} for each condition k is determined by its unique configuration of the dummy variables A , B , and C (see Tab. 1). Table 2 summarizes the drift rates predicted per condition and the implied interpretation for each model parameter. We will focus on the regression effects γ_1 , γ_2 , and γ_3 .

Table 2. Regression coefficients for the predicted drift rates.

K	Drift predicted ($\nu_{pred,k}$)	Interpretation of parameter added
1	μ_ν	μ_ν is the baseline drift rate
2	$\mu_\nu + \gamma_1$	γ_1 is the effect of a quantitative change
3	$\mu_\nu + \gamma_2$	γ_2 is the effect of a change in concavity
4	$\mu_\nu + \gamma_1 + \gamma_2 + \gamma_3$	γ_3 is an interaction term
5	$\mu_\nu + \gamma_4$	γ_4 is the effect of not having any change

```
## JAGS code for the hypothesis testing application
model {
  ##### Priors
  drift_mu ~ dnorm(0,1) # Baseline
  drift_lambda ~ dgamma(2,1)
  drift_sigma = pow(drift_lambda, -0.5)
  for(i in 1:4){
    gamma[i] ~ dnorm(0,1)
  }

  for(j in 1:5){
    drift_pred[j] = drift_mu + X[j]*(gamma[1]*Y[j]+
      gamma[2]*Z[j]+gamma[3]*Y[j]*Z[j]) + (1-X[j])
      *gamma[4]
  }

  ##### Sampling model
  for (k in 1:length(nTrials)) {
    # Person-by-condition parameters for DM
    parameters
    bound[k] ~ dgamma(2,1)
    nondt[k] ~ dexp(1)
    drift[k] ~ dnorm(drift_pred[cond[k]],
      drift_lambda)

    # Forward equations from EZ Diffusion
    ey[k] = exp(-bound[k] * drift[k])
    Pc[k] = 1 / (1 + ey[k])
    PRT[k] = 2 * pow(drift[k], 3) / bound[k] * pow(
      ey[k] + 1, 2) / (2 * -bound[k] * drift[k] *
      ey[k] - ey[k] * ey[k] + 1)
    MDT[k] = (bound[k] / (2*drift[k])) * (1 - ey[k])
      / (1 + ey[k])
    MRT[k] = MDT[k] + nondt[k]

    # Sampling distributions for summary statistics
    correct[k] ~ dbin(Pc[k], nTrials[k])
    varRT[k] ~ dnorm(1/PRT[k], 0.5*(nTrials[k]-1) *
      PRT[k] * PRT[k])
    meanRT[k] ~ dnorm(MRT[k], PRT[k] * nTrials[k])
  }
}
```

Results. We use the multiple linear regression model shown above to explore the effect of change quality and change type on the drift rate parameter. The top two panels in Figure 11 show the posterior distributions for the main effects of change quality γ_1 and change type γ_2 . These posterior distributions assign a low—but not nonzero—posterior density around zero (marked with a red dashed vertical line). To evaluate the strength of evidence these estimates provide, we can compute Bayes factors against a restricted model in which the weight is close to 0 (as in Appendix C, we use a tolerance of 0.1, meaning that we test the notion that the weight is less than 0.1 in absolute value). For γ_1 , we obtain a Bayes factor of 0.93, which is close enough to 1 as to be nearly perfectly equivocal. For γ_2 , we obtain a Bayes factor of 26, which is strong evidence for a difference between the concavity and convexity conditions.

The bottom left panel in Figure 11 shows the posterior distribution for the interaction effect of change type and change quality γ_3 , which has some posterior density around zero despite being centered around -0.45 . This suggests that there is at best weak evidence for an interaction effect between these two factors. Indeed, when we compute the Bayes factor against $\gamma_3 \in (-0.1, 0.1)$, we get a value of 0.67, again close to 1.

In the bottom right panel of Figure 11, we show the posterior distributions of the predicted drift rates across each possible combination of change quality and change type. We see a clear effect of change type (concavity vs. convexity), but the other effects are more murky.

Speed. Fitting the JAGS model for this example on the 5,722 observations after data cleaning took 3.8 seconds in an off-the-shelf desktop computer.

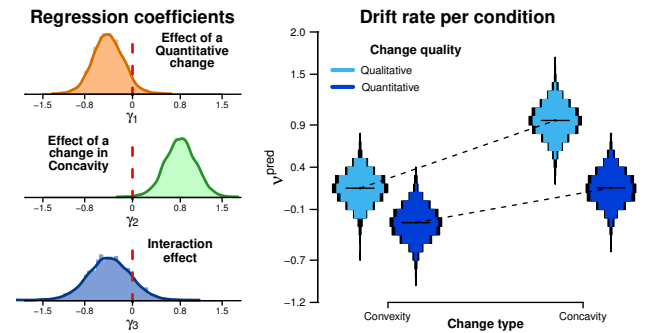


Fig. 11. Posterior distributions of the regression coefficients γ_1 , γ_2 and γ_3 , and posterior distributions of the predicted drift rates across all possible combinations of change type $\times$ change quality. The top left panel shows the posterior distribution of the main effect of the change quality γ_1 . The top right panel presents the posterior distribution of the main effect of change type γ_2 . The bottom left panel displays the posterior distribution of the interaction effect between change type and change quality. All these distributions mark the zero-value position with a vertical dashed red line. The bottom right panel presents the posterior distributions for the predicted drift rates across all change type $\times$ change quality combinations.