
Article:
Bose, T., Pirrone, A., Reina, A. et al. (1 more author) (2020) Comparison of magnitude-sensitive sequential sampling models in a simulation-based study. Journal of Mathematical Psychology, 94. 102298. ISSN 0022-2496

https://doi.org/10.1016/j.jmp.2019.102298

Comparison of magnitude-sensitive sequential sampling models in a simulation-based study

Thomas Bose[a,*], Angelo Pirrone[b,c], Andreagiovanni Reina[a], James A. R. Marshall[a]

[a] *Department of Computer Science, University of Sheffield, Sheffield (UK)*
[b] *Department of Psychosocial Science, University of Bergen, Bergen (Norway)*
[c] *Department of Psychology, University of Sheffield, Sheffield (UK)*

Abstract

Modelling plays a key role in explaining data in psychology and neuroscience and helps elucidate neural computations. Recent observations of magnitude-sensitivity (i.e. sensitivity to overall magnitudes and magnitude differences) in both humans engaged in perceptual decision making and monkeys engaged in value-based decisions have shown that new assumptions (such as the inclusion of noise that is proportional to magnitudes of external stimuli) in routinely-used sequential sampling models need to be considered to fit this type of magnitude-sensitive data. In this paper, we studied different variants of diffusion-type models and a leaky-competing accumulator model, and compared their behaviour in response to varying input magnitudes as well as their ability to resemble each other. We evaluated the extent to which these models can give good fits to simulated reaction time distributions for choices between unequal and equal alternatives. As a result, in some cases we obtained good fits of model and data, even when the underlying model used for data generation was different compared to the model used to fit these data. Our results underpin the importance of both overall magnitude and magnitude difference effects in models describing the sequential integration of evidence, and contribute to the debate over possible model candidate explanations. We discuss how magnitude-dependent input noise and lateral inhibition may be used to regulate different magnitude-sensitive effects and the implications for quantitative analyses of experimental data.

Keywords: decision making, magnitude sensitivity, sequential sampling models, model mimicry

1. Introduction

Decision-making models describing the sequential accumulation of evidence have proven to be important quantitative tools to describe decision making behaviour in a variety of cognitive tasks (Bogacz et al., 2006;

☆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.
☆☆Computer code for data generation and model fitting is available under: https://github.com/DiODeProject/magnitude-sensitive-sequential-sampling-models.
*Corresponding author
Email address: t.bose@sheffield.ac.uk (Thomas Bose)

Mulder et al., 2014; Forstmann et al., 2016; Ratcliff et al., 2016; O'Connell et al., 2018). Among these models, the drift-diffusion model (DDM) (Ratcliff, 1978; Ratcliff et al., 2016) has been particularly influential, and different variants of the standard DDM have been shown to provide good and psychologically-plausible fits to a wide variety of behavioural and neural data, such as obtained from motion discrimination tasks (Shadlen & Newsome, 1996, 2001; Gold & Shadlen, 2007), value-based decision making experiments (Krajbich et al., 2010; Basten et al., 2010; Krajbich & Rangel, 2011) and the study of social decisions (Krajbich et al., 2015).

Recently, magnitude-sensitive reaction times in decision making have been observed in a brightness discrimination task (Teodorescu et al., 2016). In this study, Teodorescu et al. (2016) showed that subjects have faster reaction times in conditions which maintain the same ratio or difference in evidence between two stimuli when the overall magnitude is increased. A further demonstration of magnitude-sensitivity in decision making was provided by Pirrone et al. (2018a), who showed magnitude-sensitive reaction times in equal alternative decision cases, both in humans performing a perceptual decision making task and in monkeys performing a reward-based task. In both experiments, when the overall magnitude of the alternatives increased, reaction times decreased. The finding of faster decisions when overall intensities are increased (Teodorescu et al., 2016; Pirrone et al., 2018a) is also in agreement with other observations in perceptual decision making (Pins & Bonnet, 1996; Stafford & Gurney, 2004; Palmer et al., 2005; Teodorescu et al., 2016; Pirrone et al., 2018a; Polanía et al., 2014; Ratcliff et al., 2018; van Maanen et al., 2012; Simen et al., 2016; Bose et al., 2019a), even in single trial dynamics Pirrone et al. (2018b), as well as in economic choices (Hunt et al., 2012; Polanía et al., 2014). Furthermore, it has been shown that magnitudes also affect the attention-choice link (Cavanagh et al., 2014; Smith & Krajbich, 2019). In line with an evolutionary perspective on naturalistic decisions (Pirrone et al., 2014), magnitude-sensitive responses to stimuli have also been found in other areas, such as collective behaviour of social insects (Pais et al., 2013; Reina et al., 2017, 2018; Bose et al., 2017) and in an ongoing decision making task related to dietary choice (Bose et al., 2019b).

Regarding possible model candidates, the new aspect previously discussed by Pirrone et al. (2014) and advanced by Teodorescu et al. (2016) is that the DDM in its canonical form fails to explain the data they obtained. Furthermore, Teodorescu et al. (2016) could show that both a DDM with multiplicative noise (mDDM) and a variant of the leaky-competing accumulator model (LCA) (Usher & McClelland, 2001; Bogacz et al., 2006) are able to explain magnitude-sensitive behavioural data. Both models contain components sensitive to relative as well as absolute evidence, which appears to be necessary to explain experimental data (Teodorescu et al., 2016; Ratcliff et al., 2018). In the mDDM the relative part is expressed by the drift rate and sensitivity to the overall magnitude is provided by input magnitude-dependent noise or (Teodorescu et al., 2016). Ratcliff et al. (2018) also studied an alternative DDM-variant where the authors assume magnitude-dependent variability of the drift-rate. This model performed equally well compared with the mDDM. The LCA is intrinsically sensitive to absolute magnitudes and relative evidence is mediated in the

LCA through lateral inhibition that couples otherwise independent evidence-integrating units (Teodorescu et al., 2016).

Regarding magnitude-sensitive data, the results reported by Teodorescu et al. (2016) highlight that mDDM and LCA may explain magnitude-sensitive behavioural data equally well, whereas Ratcliff et al. (2018) found that magnitude-sensitive DDM variants outperform the LCA in their analysis. This raises the question in the context of magnitude-sensitivity, as to what extent a magnitude-sensitive model is able to resemble performance predicted by another model sensitive to both overall magnitudes and magnitude differences? Detached from a magnitude-sensitive context, for example, Ratcliff & Smith (2004) investigated mimicry between the DDM and stable Ornstein-Uhlenbeck (OU) processes (see also Busemeyer & Townsend, 1992, 1993, for introduction of OU model and its application to value-based decisions), and showed that stable OU models with small-to-moderate values of the decay parameter are difficult to discriminate from the DDM. In another study, Teodorescu & Usher (2013) demonstrated that an independent race model and a feed-forward inhibition DDM fail to mimic classical DDM or LCA.

Building on previous studies, we performed a model comparison analysis including different sequential sampling models. In particular, we focused on four magnitude-sensitive models with multiplicative noise (mDDM, multiplicative stable OU (mSOU), multiplicative unstable OU (mUOU)), and LCA, i.e. models responsive to both overall magnitudes and magnitude differences, and compared those models with each other as well as with a pure DDM (pDDM) model that is only sensitive to relative evidence, i.e. magnitude differences which determine the drift rate. Based on the simulation of a brightness discrimination task (Teodorescu et al., 2016; Pirrone et al., 2018a), we generated artificial decision time data with each model for four different conditions. Subsequently we fitted each model to each joint data set including all four conditions. This procedure was repeated for five different parameters sets for each model. Model parameters were not allowed to vary across conditions. To constrain the sensory input term (e.g. drift rate in the diffusion-type models) we took into account a nonlinear transfer function between physical and internal stimulus in form of a power law (cf. Teodorescu et al., 2016; Ratcliff et al., 2018). This provides a strong coupling between properties of the external stimulus and internal model dynamics. As a result, we found that diffusion-type models sensitive to absolute and relative evidence (mDDM, mSOU and mUOU) are able to mimic each other with similar sublinear-to-linear shapes of the psychophysical transfer function. We also observed an asymmetry between LCA and diffusion-type models in the sense that the LCA fits better diffusion-type models with and without multiplicative noise than the diffusion-type models under study fit the LCA.

2. Models and Methods

2.1. Magnitude-sensitive models

In our study, we consider five different models, four diffusion-type models and a linear LCA model, under varying conditions to simulate and analyse two-alternative choice tasks. Regarding the diffusion-type models, we assume that the temporal evolution of the corresponding decision variable, $x(t)$, which describes sequential accumulation of evidence, may be subject to input-depend multiplicative noise (Teodorescu et al., 2016; Brunton et al., 2013). The models applied in the present paper may be summarised as

$$dx = (I_1(t) - I_2(t) + B\,x)\,dt + \Gamma(I_1, I_2)\,dW(t)\,, \tag{1}$$

where $I_{1,2}(t)$ are time-dependent internal representations of the applied stimuli further described below in Section 2.2, B represents the growth parameter ($B > 0$) or decay parameter ($B < 0$), respectively, and dW is the increment of a Wiener process, which is normally distributed, i.e. $dW \sim \mathcal{N}(\text{mean} = 0, \text{SD} = 1)$. The term $I_1(t) - I_2(t)$ is usually interpreted as drift, i.e. the evidence to decide in favour of one of the options available. We assume that the internal representation of the drift term underlies a trial-to-trial variability. This means that we add a small Gaussian random number sampled from $\mathcal{N}(\text{mean} = 0, \text{SD} = \sigma_{\text{drift}})$ to the drift at the beginning of each trial. We also take into account that the initial condition is not perfectly symmetric by assuming starting point variability (SPV) across trials, and sample the starting value $x(t = 0)$ from a uniform distribution $\mathcal{U}(-\text{SPV}, \text{SPV})$. The inclusion of across-trial variability in drift rate and starting point values in diffusion models has been shown to better explain behavioural data (Ratcliff & Rouder, 1998; Ratcliff & Tuerlinckx, 2002). $\Gamma(I_1, I_2)$ is an input-dependent coefficient of the noise term and has the form (Teodorescu et al., 2016)

$$\Gamma(I_1, I_2) = \sqrt{\sigma^2 + \Phi\,(I_1^2 + I_2^2)}\,, \tag{2}$$

where σ characterises a constant processing noise in the decision variable $x(t)$ and Φ quantifies the strength of the multiplicative noise originating from the transformed input signals. Through inputs I_1 and I_2, Γ depends on the magnitudes of the stimuli (see Eq. (4) below).

Eq. (1) describes noisy accumulation of evidence over time. Using the notation of the present paper, we obtain the mSOU model if $B < 0$ and the mUOU model if $B > 0$. We also note that under specific model assumptions these processes might have other properties (Diederich & Oswald, 2014, 2016). Instead, assuming $B = 0$ yields the mDDM. Further to this, if we set $B = 0$ and $\Phi = 0$ a DDM is recovered, which we call pure DDM (pDDM) throughout this study. In the pDDM, $\Gamma = \sigma = \text{const.}$ is insensitive to absolute magnitude-values.

The fifth model in our study is a linear LCA model (Usher & McClelland, 2001; Bogacz et al., 2006)

which has the form

$$\begin{aligned} dy_1(t) &= (-k\, y_1(t) - \beta\, y_2(t) + I_1(t))\, dt + \sigma\, dW_1(t)\,, \\ dy_2(t) &= (-k\, y_2(t) - \beta\, y_1(t) + I_2(t))\, dt + \sigma\, dW_2(t)\,, \end{aligned} \tag{3}$$

where y_1 and y_2 describe the activity levels of evidence-integrating units, in response to internal stimulus representations $I_{1,2}$. To avoid negative activity levels we applied $\max(0, y_j)$, $j = 1, 2$, at every simulation step. The activity level of each accumulator is independently affected by fluctuations modelled by Wiener processes with increments dW_1 and dW_2, where we again have $dW_j \sim \mathcal{N}(\text{mean} = 0, \text{SD} = 1)$, $j = 1,\ 2$. Information loss in the accumulators is characterised by the leak rate k. Cross-inhibition is included by the terms $\propto \beta$, where β denotes the inhibition strength. In the LCA model we also take into account starting point variability across trials and sample initial conditions from uniform distributions, i.e. $y_j(t = 0) \sim \mathcal{U}(0,\ SPV)$, $j = 1,\ 2$.

2.2. Data generation

We used all five models (pDDM, mDDM, mSOU, mUOU, LCA) to generate decision time data. To do so, we followed the experimental design by Teodorescu et al. (2016) and Pirrone et al. (2018a), who studied brightness discrimination of two visual stimuli. The internal representation of the physical stimulus is given as

$$I_j(t) = (m_j + \xi(t))^{\gamma}\ \ , j = 1,\ 2\,, \tag{4}$$

where the m_j represent the stimulus magnitudes (that can be controlled externally in the experiment), and γ is an exponent characterising the nonlinear relationship between the physical stimulus and its internal representation. In brightness discrimination tasks typical values of the exponent are given as $\gamma \sim 0.5$ (Geisler, 1989). However, in recently obtained empirical data relevant for our study, Teodorescu et al. (2016) obtained participant-specific γ-values between 0.5 and 0.85 based on model fits to the data using a model similar to the mDDM used in our study. In the same study the authors obtained a value of $\gamma \approx 0.3$ for the LCA model. Similar values for the LCA model have been obtained from LCA model fits in a similar empirical study by Ratcliff et al. (2018). However, with regard to diffusion-type models with multiplicative noise Ratcliff et al. (2018) find γ-values that are closer to 0.5 in their model fits but those fitted parameter values also show a standard deviations of approximately 0.3. Taking into account these empirical findings (Teodorescu et al., 2016; Ratcliff et al., 2018), here we have chosen $\gamma \sim \mathcal{U}(0.3, 0.7)$ for all data generating models (pDDM, mDDM, mSOU, mUOU, LCA). We argue that these choices are suitable values for our model comparison study, as they are motivated by relevant empirical findings and reflect the quantitative difference of γ-values observed in experiments (Teodorescu et al., 2016; Ratcliff et al., 2018). $\xi(t)$ is a Gaussian random number, i.e. $\xi \sim \mathcal{N}(\text{mean} = 0,\ \text{STD} = 0.1)$, which is sampled at the beginning of each trial and then again every

20 ms during the trial. This corresponds to refreshing the physical stimulus at a rate of 50 Hz. In accordance with the experimental implementation (Teodorescu et al., 2016; Pirrone et al., 2018a) we introduced lower and upper cut-off values for the stimulus magnitudes. If the sum $m_j + \xi(t)$, $j = 1, 2$, was below 0.1 we reset it to this value, and if this sum was larger than 1 we reset it to this upper limit, i.e. $0.1 \leq m_j + \xi(t) \leq 1$ in all simulations.

Each model was simulated for $N_{\text{cond}} = 4$ different conditions given by the following four combinations of stimulus magnitudes m_1 and m_2 (Teodorescu et al., 2016; Pirrone et al., 2018a): *baseline*: ($m_1 = 0.4$, $m_2 = 0.3$), *additive*: ($m_1 = 0.6$, $m_2 = 0.5$), *multiplicative*: ($m_1 = 0.6$, $m_2 = 0.45$) and *equal*: ($m_1 = 0.45$, $m_2 = 0.45$). In the additive condition, the magnitudes corresponding to the baseline condition, m_1 and m_2, are increased by an equal amount to maintain the difference between them, whereas in the multiplicative condition both baseline condition magnitudes are increased by different amounts to maintain the ratio between them. More precisely, using magnitudes m_1 and m_2 let us define the:

- *magnitude difference* as $\Delta = m_1 - m_2$,
- *magnitude ratio* as $\rho = m_1/m_2$,
- *overall magnitude* as $\mu = m_1 + m_2$.

Comparing baseline and additive conditions we see that the magnitude difference $\Delta = 0.1$ is maintained and the magnitude ratio ρ decreases from 4/3 to 6/5 with increasing magnitudes. In contrast, a comparison between baseline and multiplicative conditions shows that the magnitude difference Δ increases from 0.1 to 0.15 with increasing magnitudes whilst the magnitude ratio $\rho = 4/3$ remains the same. The magnitude chosen for the equal alternatives case is the mean value obtained from the largest and lowest magnitudes used for $m_{1,2}$ in our study, i.e. $(0.6 + 0.3)/2$. Hence, the corresponding overall magnitude for the equal condition ($\mu = 0.9$) lies in between that of the baseline condition ($\mu = 0.7$) and that of the additive condition ($\mu = 1.1$). We point out that the nonlinear transfer function in Eq. 4 causes the pDDM to be sensitive to absolute magnitudes even if the noise term in Eq. (2) with $\Phi = 0$ is not, i.e. the drift term $I_1 - I_2$ is not the same in baseline and additive conditions although the magnitude differences are equal ($\Delta = 0.1$). This also applies to mDDM, mSOU and mUOU. In case $\gamma = 1$ the pDDM becomes completely insensitive to absolute magnitudes.

We used the same model-specific parameter set for each condition and model, i.e. for every condition only m_1 and m_2 were varied. Each of the four conditions was simulated for $N = 2 \times 10^4$ trials using an Euler method with step size $\Delta t = 0.002\,(\equiv 2\,\text{ms})$. This gave a total number of trials of $N_{\text{conds}} \cdot N = 8 \cdot 10^4$. To allow the decision making process to be concluded, we introduced the decision threshold z as another model parameter. As soon as the decision variables $x(t)$ (in the diffusion-type models) or $y_{1,2}(t)$ (in the LCA) crossed threshold z, i.e. $|x| \geq z$, or $y_{1,2} \geq z$, respectively, the decision process came to an end

and the response was recorded (free response paradigm). The procedure was repeated for five different, randomly generated parameter sets for each model. The range from which model parameters were sampled are summarised in table 1. If the decision variable did not meet the decision criterion within $t \leq T_{cut}$, where $T_{cut} = 6\,\mathrm{s}$ is the cut-off time, we excluded the result. We achieved an exclusion rate far below 1% for most of the data sets. Only occasionally the exclusion rate was slightly higher but never exceeded 2.5%. Additive noise characterised by σ was fixed and kept constant.

Table 1: Overview of the generation of model parameter values sampled from uniform distributions that were used to create the simulated data. The additive noise characterised by σ was fixed.

model	z	B	γ	σ	Φ	σ_{drift}	SPV	k	β
pDDM	$\sim\mathcal{U}(0.1, 0.4)$	0	$\sim\mathcal{U}(0.3, 0.7)$	0.1	0	$\sim\mathcal{U}(0.04, 0.08)$	$\sim\mathcal{U}(0.05, 0.1)$	—	—
mDDM	$\sim\mathcal{U}(0.1, 0.4)$	0	$\sim\mathcal{U}(0.3, 0.7)$	0.1	$\sim\mathcal{U}(0.05, 0.2)$	$\sim\mathcal{U}(0.04, 0.08)$	$\sim\mathcal{U}(0.05, 0.1)$	—	—
mSOU	$\sim\mathcal{U}(0.1, 0.4)$	$\sim -\mathcal{U}(1, 4)$	$\sim\mathcal{U}(0.3, 0.7)$	0.1	$\sim\mathcal{U}(0.05, 0.2)$	$\sim\mathcal{U}(0.04, 0.08)$	$\sim\mathcal{U}(0.05, 0.1)$	—	—
mUOU	$\sim\mathcal{U}(0.1, 0.4)$	$\sim\mathcal{U}(1, 4)$	$\sim\mathcal{U}(0.3, 0.7)$	0.1	$\sim\mathcal{U}(0.05, 0.2)$	$\sim\mathcal{U}(0.04, 0.08)$	$\sim\mathcal{U}(0.05, 0.1)$	—	—
LCA	$\sim\mathcal{U}(0.3, 0.6)$	—	$\sim\mathcal{U}(0.3, 0.7)$	$0.1/\sqrt{2}$	—	—	$\sim\mathcal{U}(0.05, 0.3)$	$\sim\mathcal{U}(0.2, 2)$	$\sim\mathcal{U}(0.2, 2)$

2.3. Model fitting

We applied a quantile maximum likelihood estimation (QMLE) method (Heathcote et al., 2002; Teodorescu et al., 2016), to test if the four different diffusion models and the LCA can be made equivalent in the sense that different models yield statistically similar behavioural data. For both responses in favour of option 1 and option 2 we divided the simulated data into six bins each for decision times ranging in the interval $[0,\, T_{\mathrm{cut}}]$. This gave $N_{\mathrm{bins}} = 12$ bins per condition. Bin widths were derived from the 0.1, 0.3, 0.5, 0.7, and 0.9 decision time quantiles for choosing option 1 and option 2, respectively. This procedure gave a total number of $N_{\mathrm{cond}} \cdot N_{\mathrm{bins}} = 48$ bins included in each model fit. Every model was fitted simultaneously to all conditions (magnitude combinations of m_1 and m_2).

The QMLE method applied is based on the minimisation of the Bayesian Information Criterion (BIC) given as (Teodorescu et al., 2016)

$$\mathrm{BIC} = F + K_{\mathrm{par}}\,\ln(N_{\mathrm{tot}})\,, \qquad F = -2\sum_{\alpha=1}^{N_{\mathrm{cond}}}\sum_{j=1}^{N_{\mathrm{bins}}} n_j^{\alpha}\,\ln\left(p_j^{\alpha}\right)\,, \tag{5}$$

where p_j^{α} represents the probability of observing a decision time in a particular bin j under condition α predicted by the model ($\sum_j p_j^{\alpha} = 1$ for each condition α). The number of observations of decision times in bin j for condition α that results from the simulation of empirical data is denoted n_j^{α}, and $N_{\mathrm{tot}} = \sum_{\alpha}\sum_j n_j^{\alpha}$ is the total number of observations, i.e. the total sample size. The expression for F in Eq. (5) describes twice the negative maximum likelihood that data n_j^{α} was produced by the model yielding p_j^{α}. K_{par} is the total number of model parameters included in the fitting.

To obtain good starting values we applied a simple genetic algorithm that uses the principles of mutation and selection but not crossover. More precisely, starting from an initial parameter set 45 randomly modified parameter sets were generated. Then the 15 parameter sets with the lowest value of the objective function were kept. Each of these parameter sets had three offspring, again generated by random mutation. This procedure was repeated for 40 iterations. At each iteration the permitted parameter range used for the mutation step was shrunk by multiplying it with a factor of 0.95 to facilitate convergence. The simulation for each parameter set was done using 10^3 trials.

We applied the BIC defined in Eq. (5) to quantify the goodness of fit. To move systematically in parameter space to find the minimal BIC-value, we made use of the Nelder-Mead SIMPLEX algorithm (Nelder & Mead, 1965) implemented in the Python library SciPy. We fitted each model to every other model. This also included the fits of each model to data generated from the same model. This way we checked whether our fitting method was able to reproduce the original parameter set. Sometimes models can yield good fits with parameter sets different from the original one, e.g. see Miletić et al. (2017) for a recent study relating to the LCA, which considers the case of abstract drift rates that are not connected to the stimulus values (or magnitudes). That is, it seems to be intrinsically difficult to recover model parameters of the LCA model even in case the model is known (Miletić et al., 2017).

When models to produce and fit the data were identical, we therefore included the parameter set used to generate the data in the initial parameter grid of the fitted model to confirm that our method is sensitive enough to select this from a range of different parameter sets that yield very similar BIC scores. At the same time we also obtained a reference value of the BIC in case of (almost) perfect recovery of the original parameters.

During the search for the initial parameter set we kept track of the three best parameter combinations giving the three lowest BIC scores. Using those three parameter sets, after the 40 iterations we simulated the model again for $2 \cdot 10^4$ trials per condition and re-calculated the BIC scores for each parameter set. The starting parameter set was then identified as the one giving the lowest BIC score with the increased number of trials. This improved the sensitivity of our fitting method, as occasionally the order of BIC scores changed after increasing the number of trials. This is a consequence of the noise inherent in the decision-making process, and including the comparison of several potential parameter sets at the end of a grid search helped to reduce the randomness in the fitting (see Eqs. (6) and (8) below).

After we obtained the starting parameter values, we used them as input for the fitting routine. We then repeated every model fit six times with the same set of starting values. From these computations we obtained statistics on the goodness of fit. The results are summarised in tables 4-8 in the Appendix.

We used absolute tolerances for the BIC as quantified in table 3 in the Appendix. Every calculation of the BIC within the SIMPLEX algorithm was based on $4 \cdot 10^4$ trials. The optimisation terminated successfully

when two successive calculations of the BIC value were below the absolute tolerance and the absolute variation in the estimated parameters was ≤ 0.005. Then we performed a final simulation with twice the number of trials, i.e. $8 \cdot 10^4$ trials for all conditions in total.

When choosing the tolerance value of the BIC that determined the criterion for a successful fit we took into account that there is a limitation on the accuracy of the BIC which is due to the stochastic nature of the accumulator models in Eqs. (1) and (3). We estimated this inherent uncertainty of the BIC, denoted δ_{BIC}, using

$$\delta_{\mathrm{BIC}} = \sum_{\beta=1}^{N_{\mathrm{cond}}} \sum_{k=1}^{N_{\mathrm{bins}}} \left(\left| \frac{\partial\,\mathrm{BIC}}{\partial\, n_k^\beta} \right| \delta n_k^\beta + \left| \frac{\partial\,\mathrm{BIC}}{\partial\, p_k^\beta} \right| \delta p_k^\beta \right) . \tag{6}$$

As can be seen from Eq. (6), δ_{BIC} has two sources of inherent variability – there is uncertainty stemming from the generation of the data sets $\propto \delta n_k^\beta$ and uncertainty originated in the fitting method $\propto \delta p_k^\beta$. There are two implicit contributions to δp_k^β – one arising from the discrepancy between model and data and another one due to randomness even if the the model can explain the data well. However, this means that not only the BIC score increases with increasing discrepancy between model and data but also fluctuations in the BIC (an insight we applied to the fitting routine, as discussed below). Using Eq. (5) we can calculate the partial derivatives in Eq. (6) as

$$\begin{aligned} \frac{\partial\,\mathrm{BIC}}{\partial\, n_k^\beta} &= -2 \sum_{\alpha=1}^{N_{\mathrm{cond}}} \sum_{j=1}^{N_{\mathrm{bins}}} \ln\left(p_j^\alpha\right)\, \delta_{\alpha\beta}\delta_{jk} = -2\, \ln\left(p_k^\beta\right) , \\ \frac{\partial\,\mathrm{BIC}}{\partial\, p_k^\beta} &= -2 \sum_{\alpha=1}^{N_{\mathrm{cond}}} \sum_{j=1}^{N_{\mathrm{bins}}} \frac{n_j^\alpha}{p_j^\alpha}\, \delta_{\alpha\beta}\delta_{jk} = -2\, \frac{n_k^\beta}{p_k^\beta} , \end{aligned} \tag{7}$$

where $\delta_{\alpha\beta}$ and δ_{jk} are Kronecker deltas. To obtain a rough estimate of δ_{BIC} we make the following approximations. We assume that δn_k^β and δp_k^β are independent of condition and bin, i.e. we set $\delta n_k^\beta = \delta n$ and $\delta p_k^\beta = \delta p$, $\forall \beta \in \{1, .., N_{\mathrm{cond}}\}$, $\forall\, k \in \{1, .., N_{\mathrm{bins}}\}$. Furthermore, we use the equal alternatives condition to approximate all other conditions, that is $n_k^\beta = n_k$ and $p_k^\beta = p_k$, $\forall \beta \in \{1, .., N_{\mathrm{cond}}\}$. Assuming 10^4 trials per condition, then, considering an idealised simulation of the equal alternatives condition, the first and last bin of every distribution contain $n_k = 500$ values each and every other bin contains $n_k = 1000$ values (i.e. $4 \cdot 500 + 8 \cdot 1000 = 10^4$). Regarding the probabilities, we have $p_k = 1/20$ if k refers to the first or last bin of a distribution and $p_k = 1/10$ if k refers to any other bin (i.e. $4 \cdot 1/20 + 8 \cdot 1/10 = 1$, as required). As we approximate all other conditions with the result for the equal alternatives estimation, the sum over β in Eq. (6) becomes the prefactor 4. Calculating the sum over k in Eq. (6) by using the approximations above

and the expressions in Eq. (7) we find

$$\delta_{\mathrm{BIC}} \simeq 32 \left(7.6\,\delta n + 3 \cdot 10^4\,\delta p\right) \tag{8}$$

as an estimate for δ_{BIC}. To get an upper bound on accuracy we assume that $\delta n = 0$, i.e. the experimental data represents the true data set which is, of course, an idealisation. If we further assume in the fitting that per condition only two decision times (this is an arbitrary choice reflecting small errors) fall into wrong bins, i.e. 2 out of 10^4 divided by the number of bins this gives $\delta p = 1/6 \cdot 10^{-4}$. Inserting these values for δn and δp in Eq. (8) we obtain $\delta_{\mathrm{BIC}}^{(\mathrm{low})} \simeq 16$. Hence, even for very small uncertainties, randomness in the system does not allow for arbitrary accuracy levels in the fitting. Moreover, if we take into account a nonzero δn and more realistic (bigger) values of δp then δ_{BIC} can be much larger. In table 2 we show uncertainties of δ_{BIC} for some combinations of δn and δp.

Table 2: Overview of δ_{BIC}-values for different combinations of δn and δp. Those may be compared with the lower bound as estimated in the main text, i.e. $\delta_{\mathrm{BIC}}^{(\mathrm{low})} \simeq 16$.

δn	0	0	0	1	1	1	10	10	10
δp	10^{-4}	10^{-3}	10^{-2}	10^{-4}	10^{-3}	10^{-2}	10^{-4}	10^{-3}	10^{-2}
δ_{BIC}	96	960	9600	340	1200	9840	2530	3390	12030

We also point out that due to the number of trials included to generate the data in our study, typical BIC-reference values are in the order of magnitude of $\mathrm{BIC}_{\mathrm{ref}} \sim 4 \cdot 10^5$, and therefore the relative uncertainty, approximated by $\delta_{\mathrm{BIC}}/\mathrm{BIC}_{\mathrm{ref}}$, is about $3 \cdot 10^{-3}$ using $\delta n = 1$ and $\delta p = 10^{-3}$, for example (see table 2). We took this inherent bound on accuracy into account in the fitting method. Comparing our estimates for the δ_{BIC} values in table 2 with the absolute tolerances used in our study (cf. table 3 in the Appendix), we see that fits, which successfully terminated, were achieved using reasonable tolerances.

Note that we did not include a non-decision time in our study, as the simulated models directly yield the pure decision time without any pre-processing, such as stimulus encoding, or any post-processing, such as executing a motor response. The non-decision time is usually taken into account as an additional free parameter allowed to vary in models fitted to experimental data. By not including this parameter we increase the constraints on the different models and hence reduce the potential for model mimicry. The additive noise characterised by σ was not varied during the model fitting.

3. Results

3.1. Overview of magnitude-sensitive models

Here we show the condition-dependent behaviour of the different models used to generate behavioural data. Fig. 1 depicts the mean decision times of choosing options 1 and 2, denoted $\langle DT_1 \rangle$ and $\langle DT_2 \rangle$, and the

response proportion of choosing option 1, denoted $Pr(opt1)$. Results are summarised for five different model parameter sets for each model that were used to generate the data. We can see that mSOU, mUOU, mDDM and LCA qualitatively exhibit similar behaviour. This means that $\langle DT_1 \rangle$ and $\langle DT_2 \rangle$ decrease when changing the condition from baseline to additive for these four models, in which case the magnitude difference $\Delta = 0.1$ remains constant, the magnitude ratio ρ decreases from 4/3 to 6/5, and the overall magnitude μ increases from 0.7 to 1.1. Therefore, the reduction of the decision time is clearly caused by the increase of the overall magnitude. At the same time $Pr(opt1)$ goes down, as illustrated in Figs. 1C, F, I, and O. This is an effect of keeping Δ constant whilst increasing μ (i.e. a decrease of ρ), which makes the decision problem more difficult. Furthermore, a comparison of mean reaction times between the multiplicative and the baseline condition demonstrates that $\langle DT_1 \rangle$ and $\langle DT_2 \rangle$ also decrease when μ increases whilst ρ is kept constant (i.e. Δ increases). However, if we compare mean decision times of additive and multiplicative conditions, we see that the multiplicative condition yields slightly larger mean decision times than the additive condition. We attribute this result to the slightly larger overall magnitude in the additive condition ($\mu = 1.1$) compared with that in the multiplicative condition ($\mu = 1.05$), which demonstrates sensitivity to absolute values of the input stimuli. In addition, we have a larger magnitude difference in the multiplicative condition ($\Delta = 0.15$) than in the additive condition ($\Delta = 0.1$). Thus, the decision problem is easier under the multiplicative condition compared with the additive one. For example, this can be seen when comparing the normalised magnitude difference Δ/μ which gives 0.14 for the multiplicative and 0.09 for the additive condition. In accordance with the decision problem being easier in the multiplicative condition we find an increase of $Pr(opt1)$ in this condition compared with the additive one. We also see that the value of $Pr(opt1)$ in the multiplicative condition approaches that of the baseline condition again.

The qualitative model behaviours observed for mSOU, mUOU, mDDM and LCA are largely in agreement with data obtained from magnitude-sensitive experiments (Teodorescu et al., 2016). That is, mean decision times decrease in both additive and multiplicative conditions compared with the baseline condition. However, comparing mean decision times of additive and multiplicative conditions, Teodorescu et al. (2016) find that the mean decision time of an average observer in the additive condition is slightly larger compared with that of the multiplicative condition, which is the opposite of our observation where the additive condition yields slightly shorter decision times, as shown in Fig. 2. In contrast, Pirrone et al. (2018a) did not obtain similar differences between baseline, additive and multiplicative conditions in their implementation of the experiment. However, the deviating observations regarding mean decision times made by Teodorescu et al. (2016) and Pirrone et al. (2018a) are probably due to not linearising the display screen with respect to brightness in the study by Pirrone et al. (2018a). Additionally, the deviations may also indicate participant-specific behaviours when comparing baseline, additive and multiplicative conditions which may not always follow a regular pattern. The patterns we observe in our simulations are most likely due to the specific set of

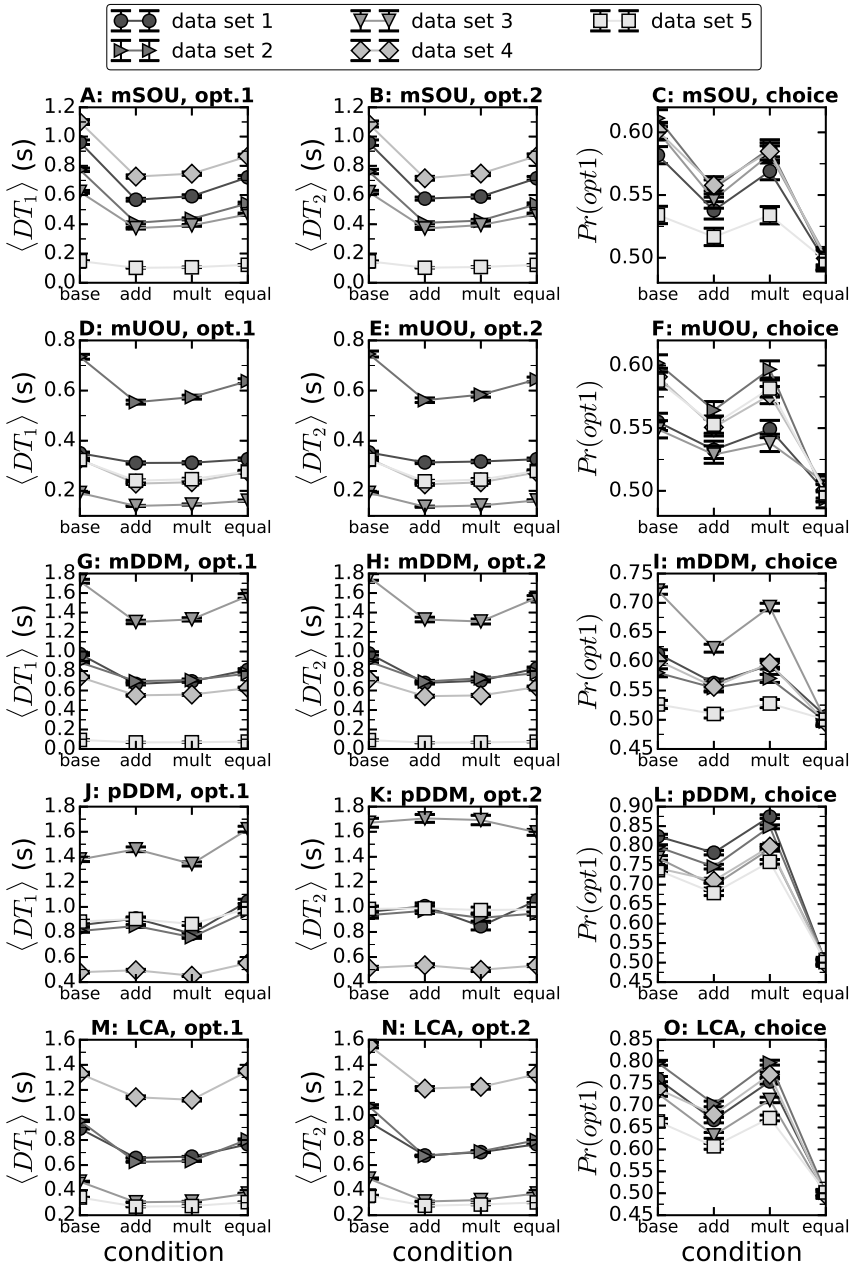

Figure 1: Comparison of simulated data sets for four different conditions. Data were generated for mSOU (A-C), mUOU (D-F), mDDM (G-I), pDDM (J-L) and LCA (M-O) in baseline, additive, multiplicative and equal alternatives conditions. Shown are mean decision times for choosing option 1 ($\langle DT_1 \rangle$), mean decision times for choosing option 2 ($\langle DT_2 \rangle$), and response proportions in favour of option 1 ($Pr(opt1)$). Qualitatively similar behaviour is observed for mSOU, mUOU, mDDM and LCA, whereas the behaviour of the pDDM differs compared with the other models. Error bars denote 95% confidence intervals. All model parameters for data generation are given in table 1.

parameters chosen in our study and suggest that observing faster decisions in the additive compared with the multiplicative condition, or vice versa, depends on the joint effect of varying overall magnitude, magnitude

difference and magnitude ratio in conjunction with decision-maker-specific characteristics expressed by the other model parameters.

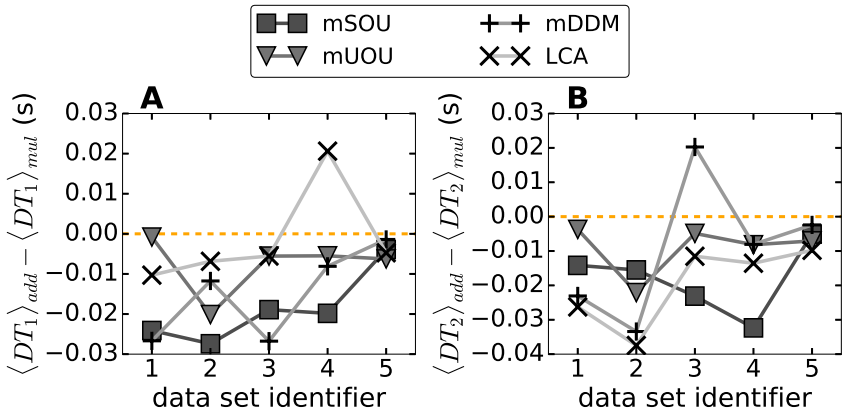

Figure 2: Comparison of mean decision times in additive and multiplicative conditions separately displayed for responses in favour of option 1 (A) and option 2 (B). All model parameters for data generation are given in table 1.

Our simulation results show qualitatively the same behaviour of the mean decision time for all magnitude-sensitive models with multiplicative noise (mSOU, mUOU, mDDM, LCA), when comparing additive and multiplicative conditions (cf. Fig. 2). We also note that the behaviour of $Pr(opt1)$ when comparing baseline, additive and multiplicative conditions is similar to that observed by Teodorescu et al. (2016) and Pirrone et al. (2018a).

Looking at the behaviour obtained for the condition with equal magnitudes ($\rho = 1$, $\Delta = 0$) in Fig. 1, we find that $\langle DT_1 \rangle$ and $\langle DT_2 \rangle$ are below the mean decision times corresponding to the baseline condition and above those corresponding to additive and multiplicative conditions. This may again be interpreted a consequence of the overall magnitude in the equal condition ($\mu = 0.9$) being larger than that of the baseline condition ($\mu = 0.7$) but smaller than those of additive ($\mu = 1.1$) and multiplicative ($\mu = 1.05$) conditions. This further supports the finding that mean decision times are strongly dependent on absolute magnitude values. Furthermore, the value of $Pr(opt1) = 0.5$ shows that without a difference in magnitudes ($\Delta = 0$) option 1 is chosen in half of the total number of trials, as expected.

The simulation results of the pDDM are depicted in Figs. 1J, K and L. This model is not able to reproduce the same magnitude-sensitive patterns compared with the other models discussed above. In particular, in the additive condition the simulation of the pDDM yields an increase of $\langle DT_1 \rangle$ and $\langle DT_2 \rangle$, compared with the baseline condition. Obtaining different mean decision times in baseline and additive conditions for the pDDM is a result of the nonlinear transfer function in Eq. (4) with $\gamma \neq 1$. Moreover, we observe the lowest mean decision times for the multiplicative condition and the highest mean decision times for the condition with equal magnitudes of both stimuli. The behaviour of $Pr(opt1)$ in Fig. 1L is also different compared with

mSOU (Fig. 1C), mUOU (Fig. 1F), mDDM (Fig. 1I) and LCA (Fig. 1O), and shows a significant increase of $Pr(opt1)$ in the multiplicative condition compared with both additive and baseline condition.

3.2. Graphical overview of results obtained from model fitting

Model fitting results are summarised in Fig. 3 which gives a graphical overview highlighting which models were able to resemble each other. We particularly emphasise in Fig. 3 conditions between important model parameters that seem to indicate mimicry between different models, as suggested by our analysis. These important model parameters are the multiplicative noise strengths (Φ) in mDDM, mSOU and mUOU; the leak parameter (k) and the inhibition strength (β) in the LCA; and the growth (decay) parameter $B > 0$ ($B < 0$) in mUOU (mSOU). A more detailed description of our model fitting results can be found in sections 3.3-3.7.

3.3. Model fits to data generated by the pure drift-diffusion model

In this and in the subsequent sections we present a more detailed model comparison based on the analysis of the decision time quantiles for correct and error choices. We are particularly interested in the question of how well different models can explain data simulated using another model. In Fig. 4 we show how the different model fits performed when the data set was generated by the pDDM. We emphasise again that, by model design, the pDDM is sensitive to magnitude differences but not to absolute magnitudes (apart from the psychophysical transformation of input stimuli, see Eq. (4)). It is therefore not surprising that using the pDDM as model to fit pDDM-data shows the best agreement (cf. BIC-values in table 4 in the Appendix). As the multiplicative noise strength, Φ, was allowed to vary during the fitting process, we also see that the mDDM with $\Phi \simeq 0$, as obtained from the fitting (table 4 in the Appendix), gives an excellent agreement with the data. Using the mSOU model to fit the pDDM data we also found that the best agreement was achieved with almost zero noise strength, i.e. $\Phi \simeq 0$ (table 4). We note that the resemblance of the pDDM and stable OU processes for small-to-medium values of the decay parameter in absence of multiplicative noise is well-known Ratcliff & Smith (2004). Therefore, our results confirm previous findings.

As a new aspect, we found that the mUOU model in which $\Phi \simeq 0$ gave reasonable fits to pDDM data, too. However, the LCA did not perform as well to fit the pDDM data. A close resemblance between these two models requires specific assumptions on the model parameters, i.e. $\beta = k$ and $\beta + k$ sufficiently large in the LCA (Bogacz et al., 2006). Therefore we imposed that $\beta = k$ during the fitting of the LCA model to pDDM data to improve the model fits. Reasons that the LCA was outperformed by multiplicative diffusion-type models might be related to the drift variability which was not taken into account in the LCA and the type of experiment simulated in our study. That is, the stimulus was flickering and not constant over the course of a decision and that the stimulus was transduced into an internal representation via a nonlinear transfer function. Taken together this may have led to the differences observed.

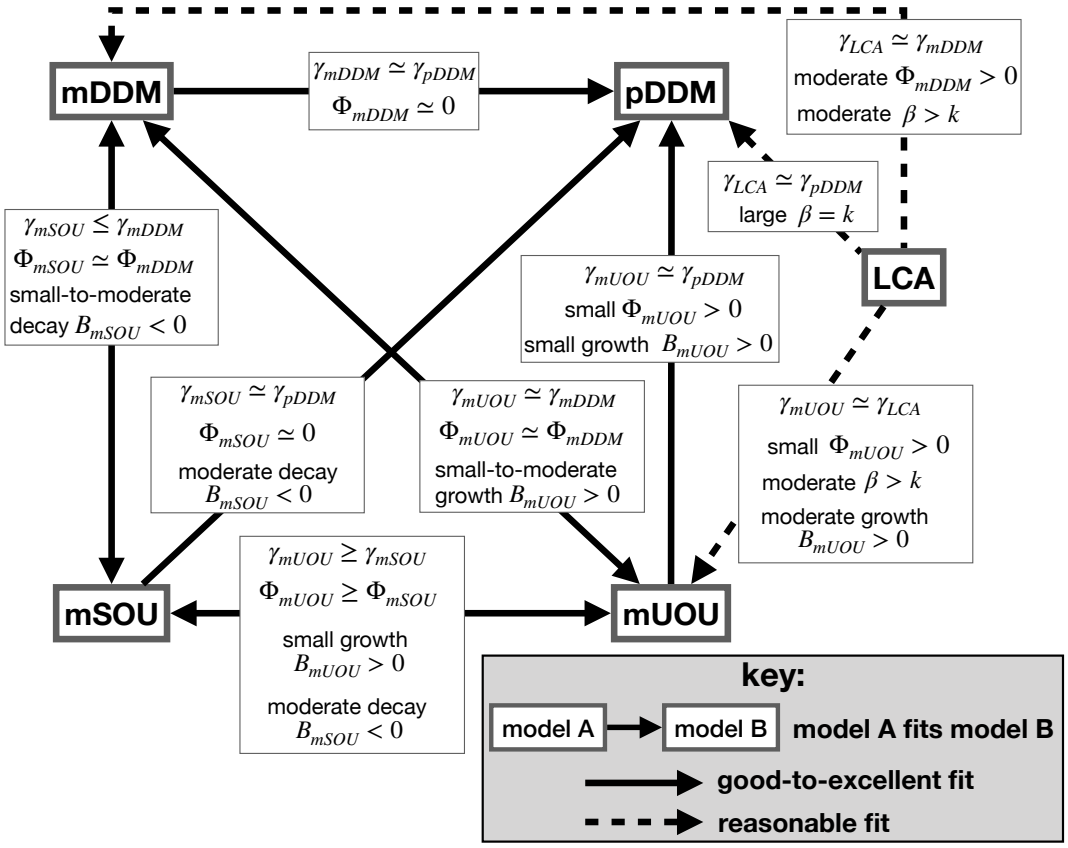

Figure 3: Directed graph summarising relations between different models that may lead to model mimicry. The following models are compared: mDDM: multiplicative drift-diffusion model, pDDM: pure drift-diffusion model, mSOU: multiplicative stable Ornstein-Uhlenbeck model, mUOU: multiplicative unstable Ornstein-Uhlenbeck model, and LCA: leaky-competing accumulator. Important model parameters are: the multiplicative noise strengths (Φ) in mDDM, mSOU and mUOU; the leak parameter (k) and the inhibition strength (β) in the LCA; the growth (decay) parameter $B > 0$ ($B < 0$) in mUOU (mSOU). Arrows indicate which model could be fitted to data generated by another model. Bi-directional arrows indicate that model fits worked well either way but in some cases fits were not reciprocal (unidirectional arrows) or fits were poor regardless (no connecting arrows). To display the goodness of model fits we used *dashed lines* (indicating reasonable model fits) and *solid lines* (indicating good-to-excellent model fits). BIC values are given in tables 4-8 for comparison. Fitted parameter values are available in the Supplementary Material.

3.4. Model fits to data generated by the multiplicative drift-diffusion model

Fig. 5 shows that the mDDM-model provides the best model fit, however, both mSOU and mUOU are able to fit the mDDM-data well, too (see also table 5). More specifically, we find good agreement for all quantiles and conditions between model fits (mDDM, mSOU, mUOU) and mDDM-data. In contrast, the pDDM cannot explain the mDDM-data well, underpinning the different behavioural outcome between presence and absence of multiplicative noise. Regarding the LCA-fits to mDDM-data, we expected a good agreement between LCA and mDDM, as it has been observed previously that both models are able to explain the same set of magnitude-sensitive data (Teodorescu et al., 2016). The LCA-fits in Fig. 5 show that for data sets 1-3 we observe a reasonable agreement between mDDM data and LCA model fit, whereas for mDDM

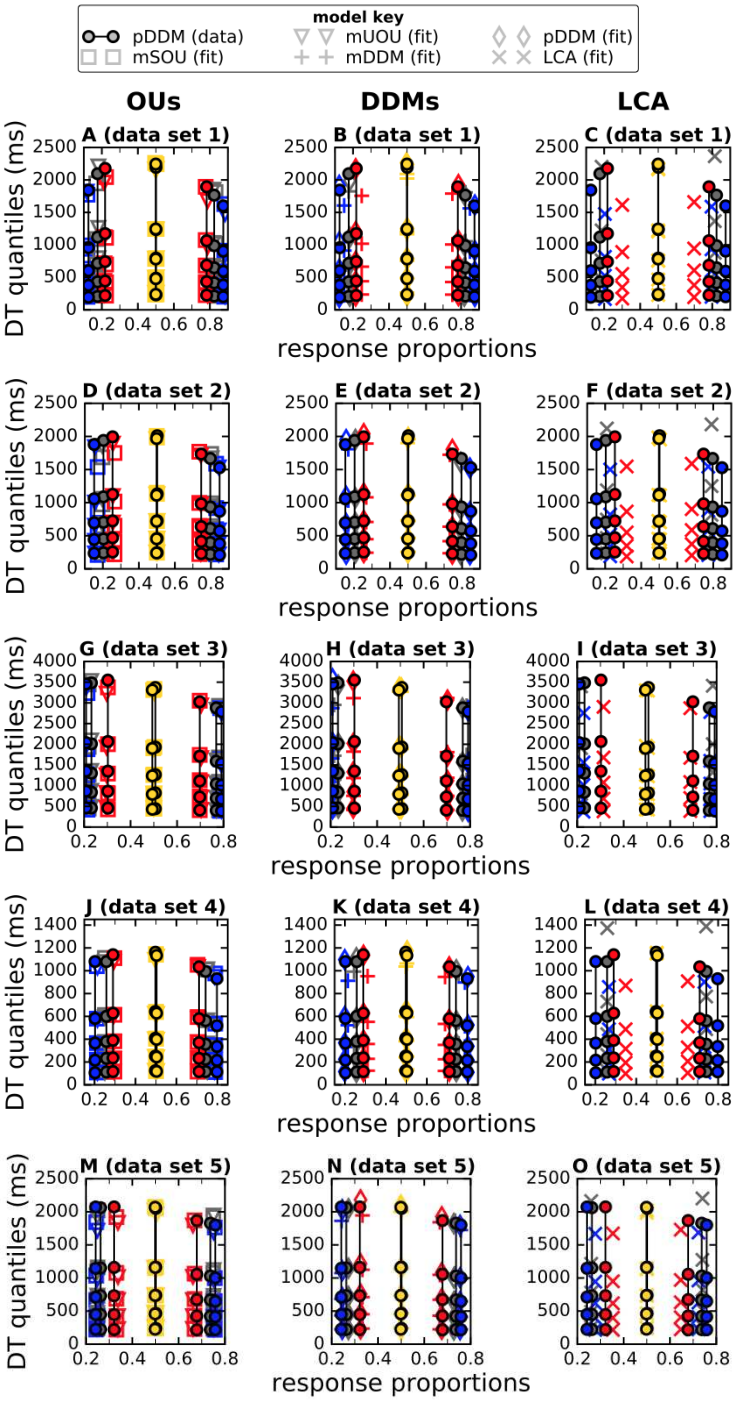

Figure 4: Comparison of different model fits to pDDM data. Four different conditions were studied: baseline (gray), additive (red), multiplicative (blue) and equal (yellow). For each condition, response proportions are plotted along the horizontal axis and decision time quantiles (0.1, 0.3, 0.5, 0.7, 0.9) for both responses in favour of option 1 (greater than 0.5) and option 2 (smaller than 0.5) are plotted vertically. BIC values are given in table 4. Fitted parameter values are available in the Supplementary Material.

data sets 4 and 5 visual inspection shows discrepancies between mDDM data and LCA model fits. Here we observe something interesting: the exponent in the psychophysical transfer function, γ, gets quite large, i.e. $\gamma \approx 7.3$ (fit to data set 4) and $\gamma \approx 5.3$ (fit to data set 5). A large superlinear γ value has the effect that the baseline condition can hardly be distinguished from the equal alternatives condition. This becomes obvious in Figs. 5L and O, where the baseline condition yields response proportions close to 0.5 for both responses. However, decision times are reproduced quite accurately. In particular, this leads to a good BIC score of the LCA fit in Fig. 5O, as the decision time quantiles shown are between $0 - 200$ ms (table 3), which is comparatively small (when compared to decision times in the other data sets). In another recent study, Ratcliff et al. (2018) observe differences between LCA and different variants of the mDDM which supports our finding that mDDM and LCA may not be able to mimic each other in general. In addition, the authors also find that the exponent in the psychophysical power-law transformation function yields values > 1 in some of the brightness-discrimination tasks studied there (Ratcliff et al., 2018), though there values are not as large.

Although it is known from previous work that stable OU processes and pDDM processes can mimic each other for small-to-moderate decay parameters (Ratcliff & Smith, 2004), which is also supported by our findings in Section 3.3 above, the question if one of the models is able to resemble the other in case of the inclusion of multiplicative noise has not been studied previously. Yet again we find evidence that the same conclusion may also be true for mDDM and mSOU accumulator models. In addition, our results also show that the mUOU model can explain data produced by the mDDM quite well (see also corresponding BIC-values in table 5 in the Appendix for quantitative comparison of goodness-of-fit).

3.5. Model fits to data generated by the multiplicative stable Ornstein-Uhlenbeck process

Fits to the data set obtained from simulating the mSOU process demonstrate that, besides the mSOU-model itself, the mDDM and mUOU are also able to explain the data accurately (see Fig. 6 and compare BIC scores in table 6 in the Appendix). Overall, decision time quantiles and response proportions are reproduced quite well. This finding further underlines the potential of mimicry in the group of diffusion models with multiplicative noise studied in the present paper (mDDM, mSOU, mUOU). Comparing mSOU data and LCA fit, we see that the LCA provides a good match in the equal alternatives conditions but also mimics the additive and multiplicative conditions in the fits to data sets 3 and 4 quite well. However, fitted to other data sets the LCA cannot mimic the values for response proportions in the baseline, additive and multiplicative conditions. In particular, when fitted to mSOU data set 5 the LCA fit yields an almost zero exponent in the psychophysical transfer function ($\gamma \approx 0.002$). This makes baseline, additive and multiplicative conditions indistinguishable from the equal alternatives condition, see Fig. 6O. This apparently unrealistic γ value was obtained when decision time quantiles were between $0 - 300$ ms. Although decision times are rarely less than 300ms, this could suggest that data produced by mSOU yielding small decision times are harder to fit using

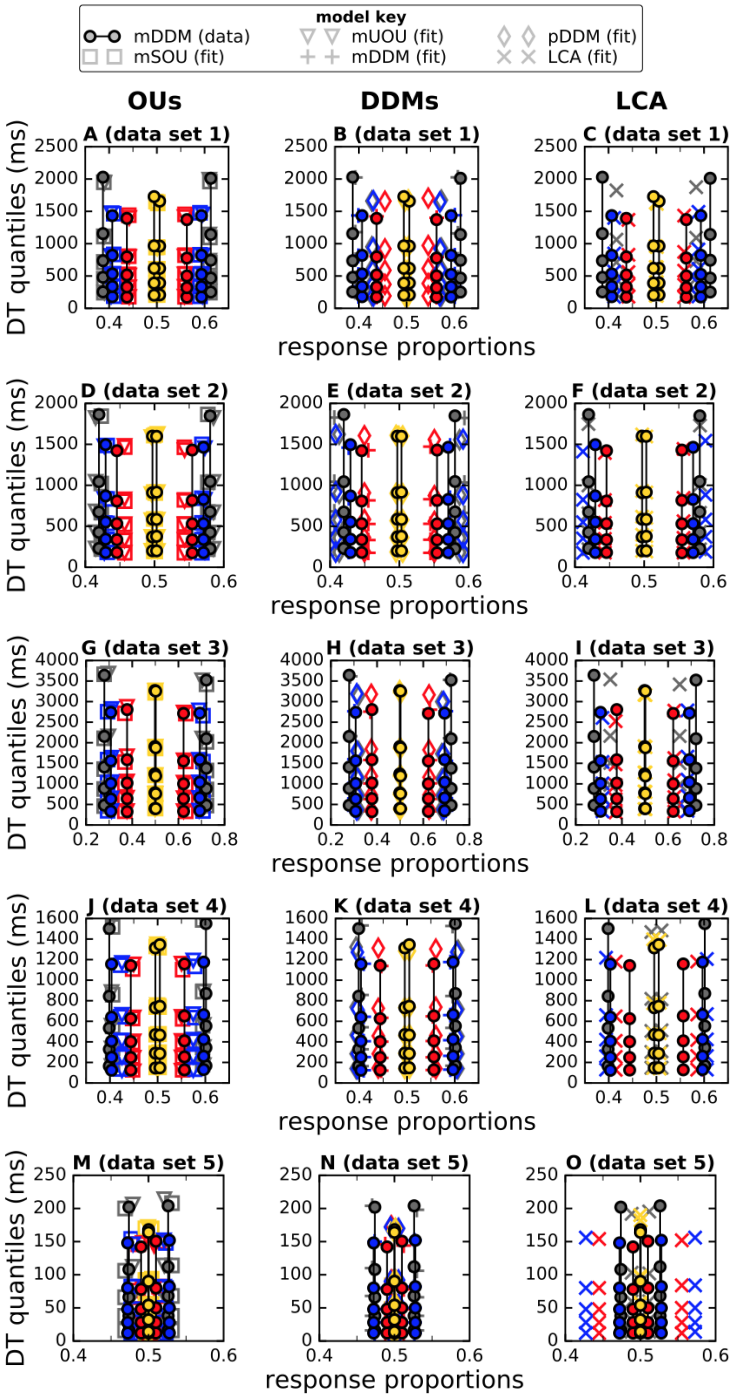

Figure 5: Comparison of different model fits to mDDM data. Four different conditions were studied: baseline (gray), additive (red), multiplicative (blue) and equal (yellow). Other plotting conventions are the same as in Fig. 4. BIC values are given in table 5. Fitted parameter values are available in the Supplementary Material.

the LCA than mSOU data giving larger decision times. In addition, the fit of the pDDM cannot account for the mSOU-data, neither qualitatively nor quantitatively. This demonstrates once more the incompatibility of the pDDM and magnitude-sensitive data.

3.6. Model fits to data generated by the multiplicative unstable Ornstein-Uhlenbeck process

Fig. 7 shows how the different models performed when the data set was generated by the mUOU. Here we found that, besides the mUOU itself, the mSOU model yielded the best fits to mUOU data, in general. One exception is data set 1, where we observe that the LCA fit gives the lowest BIC value (cf. table 6). Comparing the performance of mDDM and mSOU we see that the mDDM did not perform as well. Furthermore, our results seem to indicate that the exponent γ in the mSOU model is smaller, or equal, to that of the mUOU. Regarding the multiplicative noise strength we find that good fits are obtained when $\Phi_{mUOU} \geq \Phi_{mSOU}$.

3.7. Model fits to data generated by the leaky-competing accumulator model

If we generate data using the LCA, our results depicted in Fig. 8 show that, besides the LCA model itself, all model fits are generally poor. However, within the group of diffusion-type models the mUOU model provides the best fit to the simulated LCA data. In particular, the mUOU fit to the LCA data set 4 yielded a good agreement between data and model. The mUOU model mainly achieves reasonable decision time quantiles for all conditions but often does not yield the accurate response proportions. All other diffusion-type models (pDDM, mDDM, mSOU) did not give good fits. Our results seem to indicate that the LCA is the more flexible model, as we achieved better scores when fitting LCA model to data generated by the diffusion type models included in this study than vice versa. At the same time, our results also seem to suggest that model fitting given the data obtained from the simulation of a brightness discrimination task which was previously studied experimentally (Teodorescu et al., 2016; Pirrone et al., 2018a; Ratcliff et al., 2018) is inherently difficult. We discuss this further in the following section.

4. Discussion

In a model-based analysis, primarily inspired by brightness discrimination tasks (Teodorescu et al., 2016; Pirrone et al., 2018a), we examined to what extent sequential sampling models are able to resemble each other when noise in the evidence accumulation process increases with stimulus strength. Previously, Teodorescu et al. (2016) have shown that both a version of the mDDM and the LCA are able to explain the magnitude-sensitive data they obtained experimentally. However, our results show that mimicry between mDDM and LCA is not always the case, which is also in agreement with the results reported by Ratcliff et al. (2018) who found that two magnitude-sensitive versions of the DDM (one which is similar to the mDDM studied in this paper and another DDM-variant where across-trial variability in drift is proportional to the magnitude of the external stimulus) could explain their data obtained from a similar brightness discrimination experiment

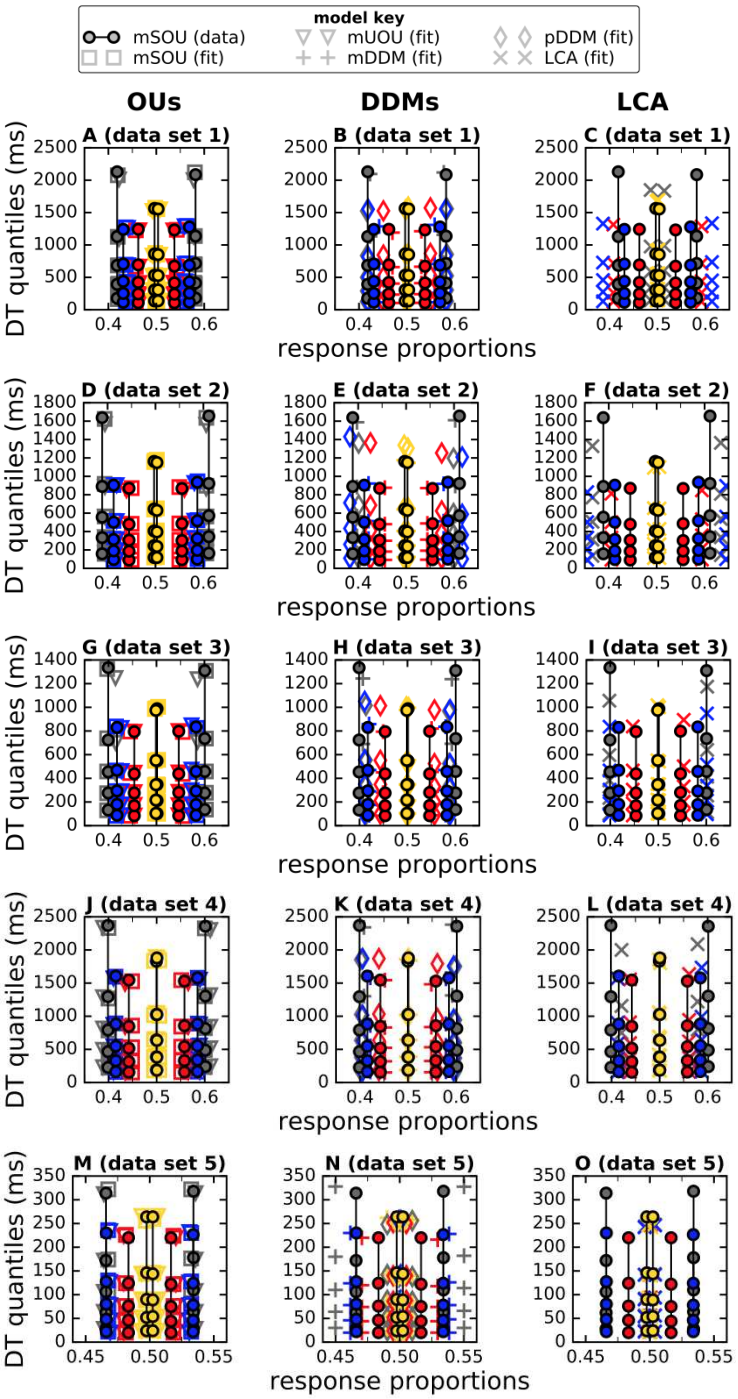

Figure 6: Comparison of different model fits to mSOU data. Four different conditions were studied: baseline (gray), additive (red), multiplicative (blue) and equal (yellow). Other plotting conventions are the same as in Fig. 4. BIC values are given in table 6. Fitted parameter values are available in the Supplementary Material.

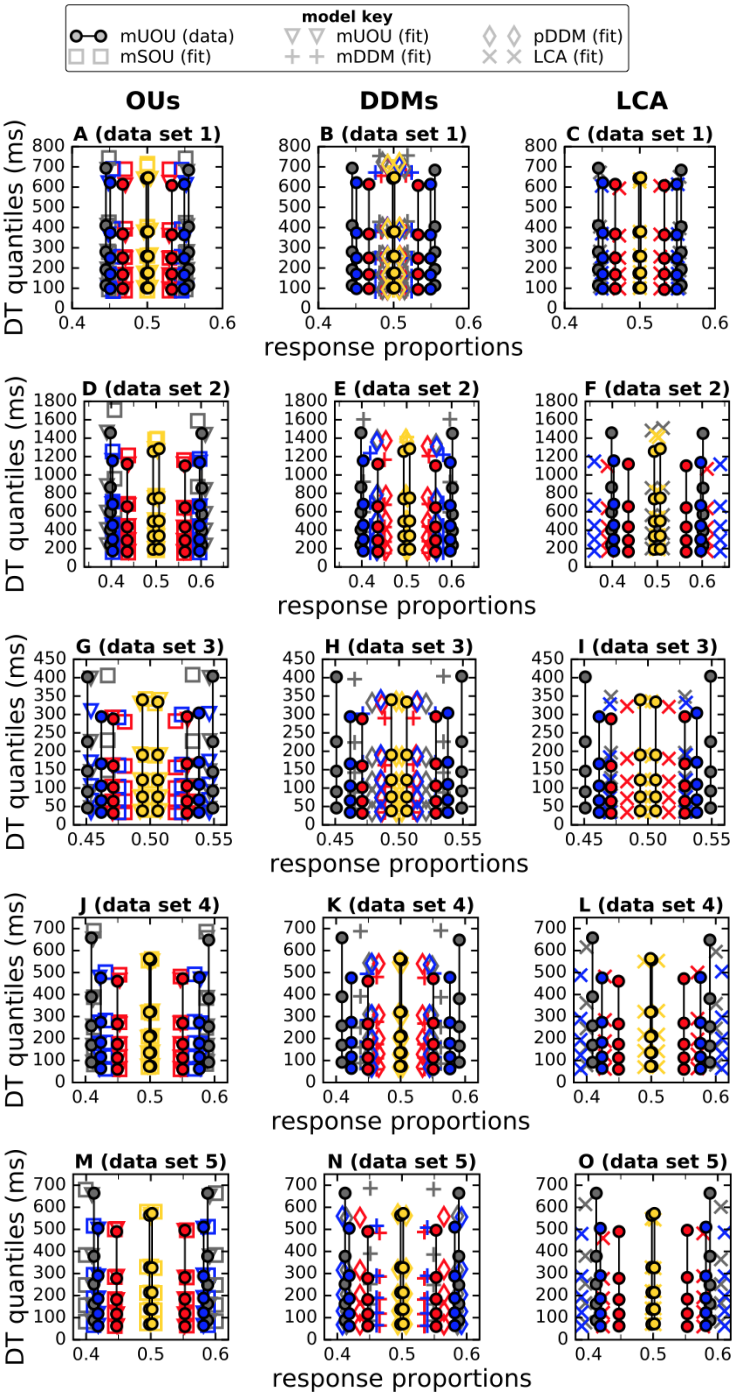

Figure 7: Comparison of different model fits to mUOU data. Four different conditions were studied: baseline (gray), additive (red), multiplicative (blue) and equal (yellow). Other plotting conventions are the same as in Fig. 4. BIC values are given in table 7. Fitted parameter values are available in the Supplementary Material.

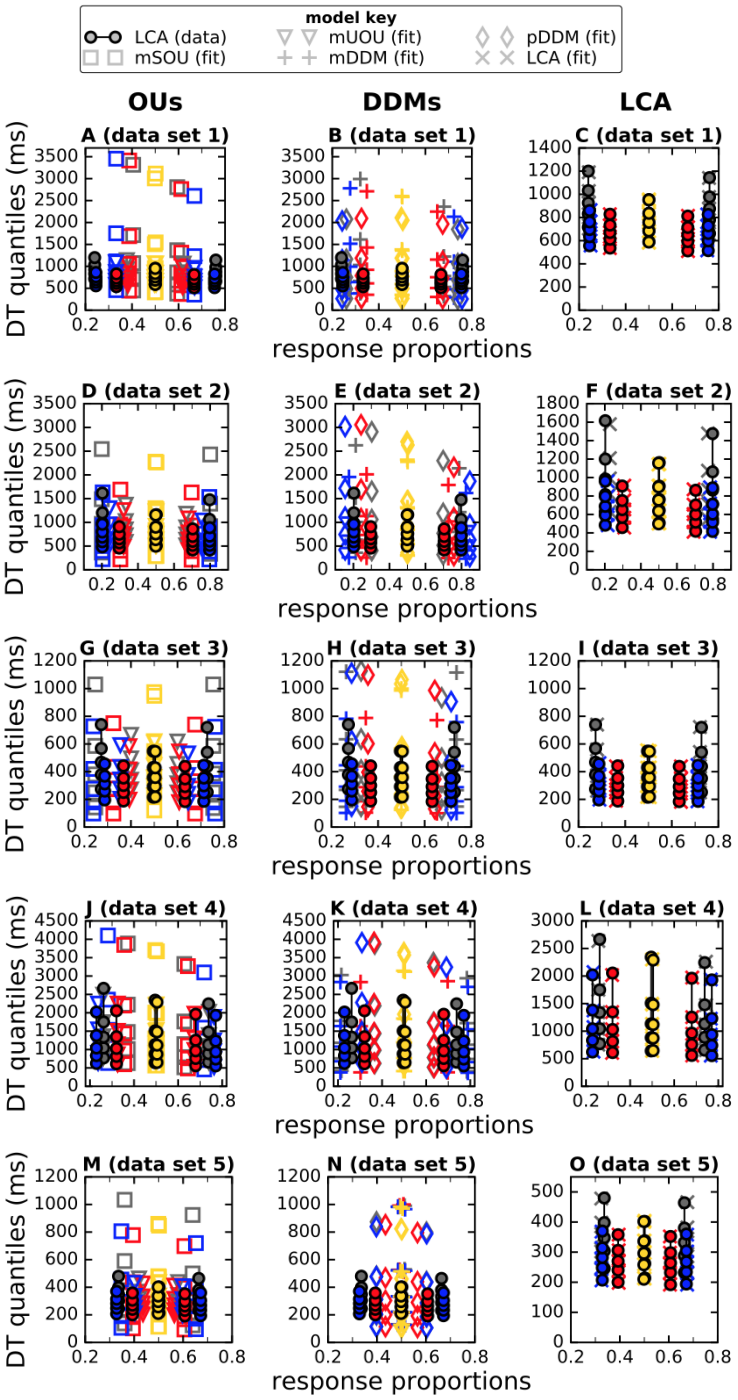

Figure 8: Comparison of different model fits to LCA data. Four different conditions were studied: baseline (gray), additive (red), multiplicative (blue) and equal (yellow). Other plotting conventions are the same as in Fig. 4. BIC values are given in table 8. Fitted parameter values are available in the Supplementary Material.

better than the LCA. Both magnitude-sensitive DDM variants performed equally well in the study by Ratcliff et al. (2018) and therefore we expected both models to behave largely similarly, in general. Hence, we decided to use only one of both models in our study. We chose the mDDM-variant as it provided good fits in related empirical studies Teodorescu et al. (2016); Ratcliff et al. (2018) and, furthermore, has also been shown to enable the determination of the origin of noise in other decision making tasks Brunton et al. (2013), which indicates that there are further use cases. We note, however, that it has also been argued that a DDM with magnitude-dependent across-trial variability in drift rate might be preferred for theoretical reasons (Ratcliff et al., 2018).

All models under consideration have contributions stemming from changes of the overall magnitude. This also includes the pDDM in case of $\gamma \neq 1$ in the psychophysical transfer function in Eq. (4). More precisely, the magnitudes m_1 and m_2 which are not under control of a decision maker as well as the exponent γ in Eq. (4), which is specific to a decision maker, regulate the effects of overall magnitudes. In the pDDM, mDDM, mSOU and mUOU the drift term $\propto I_1 - I_2$ will change when altering m_1, m_2 or γ. Even when m_1 and m_2 are increased or decreased by the same amount the internal representation of the evidence (i.e. the drift term) will change quantitatively if $\gamma \neq 1$. This is a direct consequence of the nonlinearity in the psychophysical transfer function (see Eq. (4)). Hence, Eq. (4) entangles effects of overall magnitudes and magnitude differences in the diffusion-type models. However, in contrast to the pDDM, diffusion models with magnitude-dependent noise (mDDM, mSOU, mUOU) have an additional contribution that is characterised by noise strength Φ and solely represents an effect of the overall magnitude. Therefore, increasing (decreasing) m_1, m_2 or Φ will enhance (reduce) the effect of multiplicative noise. In the LCA, a change of magnitudes $m_{1,2}$ and γ will affect absolute evidence integrated by the two decision units $y_{1,2}$. Lateral inhibition $\propto \beta$ then mediates the relative evidence between y_1 and y_2 which becomes more effective for larger activity levels $y_{1,2}$, suggesting that magnitude-sensitivity is more likely to affect the late and not so much the early stages of evidence accumulation in the LCA (Pirrone, 2018).

In our study, effects of overall magnitude may be contrasted with magnitude difference effects by detecting model-specific changes when making the transitions from baseline-to-additive condition, baseline-to-multiplicative condition, and baseline-to-equal alternatives condition. Comparing mDDM (cf. Figs. 1G-I) and pDDM (cf. Figs. 1J-L), for example, we see that the mDDM (with $\Phi > 0$) produces mean decision time and choice probability patterns that distinguishes this model from the pDDM. This underlines that the multiplicative noise strength Φ is an important model parameter which may be used to regulate responses of the decision-maker. Given the observations made by Teodorescu et al. (2016), Pirrone et al. (2018a) and the results presented in our study we conjecture that the competition between effects of overall magnitude and magnitude difference is nontrivial and depends on the experimental design and subtle model assumptions. The decisive parameters identified in our study are the externally controllable magnitudes of the physical

stimulus (m_1 and m_2 in all models), the exponent γ in the psychophysical transfer function (in all models), the decision threshold z (in all models), the multiplicative noise strength Φ in mDDM, mSOU and mUOU, and the cross-inhibition strength β in the LCA.

In the present paper we found that when the data were generated by the LCA model the mDDM did not yield a good fit (Fig. 5 and table 5), whereas when the data were generated by the mDDM the LCA model fit was much better (Fig. 8 and table 8), compared with the opposite case. This indicates that fitting mDDM to LCA data and fitting LCA model to mDDM data does not seem to be reciprocal. The LCA model seems more flexible to fit arbitrary parameter configurations in the mDDM, than vice versa. However, the best fit (although not perfect) of a diffusion-type model to LCA data was obtained for the mUOU model (Fig. 8 and table 8). Comparing these two models, we also found that the LCA model could fit several conditions and occasionally the full data set reasonably well when the data was generated by the mUOU model (Fig. 7 and table 7).

Within the set of diffusion-type models studied here, we observed resemblance between mDDM, mSOU and mUOU (Fig. 5 and table 5). Although model mimicry between classical DDM and stable OU process in case of small-to-moderate decay parameter is well-known Ratcliff & Smith (2004), observing the same type of mimicry including magnitude-dependent noise has not been reported previously, yet here we demonstrated that resemblance also occurs in the presence of multiplicative noise. Furthermore, model mimicry should become more likely with increasing multiplicative noise strengths, because mDDM, mSOU and mUOU should become more and more similar when the input-dependent noise increases.

From a more general point of view, Jones & Dzhafarov (2014) report that the predictive content of diffusion-type models (DDM and OU models without multiplicative noise) mainly depends on the assumptions made on the model parameters that are assumed to be sampled from distributions (e.g. drift rates are sampled from normal distributions and starting points from uniform distributions), rather than on the structural assumption of the model. Removing the distributional constraints, the authors come to the conclusion that these models are able to match arbitrary patterns of reaction time probabilities and distributions (Jones & Dzhafarov, 2014). However, this conclusion can also be considered as an argument which supports current practice of using constrained distributional assumptions in diffusion models, as those types of models have been applied successfully to fit response time data over decades (Heathcote et al., 2014). Additionally, the conclusions by Jones & Dzhafarov (2014) have also been criticised as a trivialisation of diffusion-type models, as it was argued that Jones & Dzhafarov (2014) included them in a non-representative general class of models (Smith et al., 2014). This discussion shows that falsifiability of models, or vice versa model mimicry, strongly depends on the constraints made on the model parameters. Considering the inclusion of magnitude-dependent noise, we found that the different models under study were very sensitive to this parameter. Hence, the modification of the amount of multiplicative noise might be a way to improve falsifiability within the

class of diffusion-type models (e.g. pDDM versus mDDM). However, if magnitude-dependent noise becomes too large then different diffusion-type models with multiplicative noise might become indistinguishable.

Another route that has obtained interest recently is the combination of computational accumulator models and neural data to enable data-driven refinement of model assumptions and to reduce model mimicry (Churchland et al., 2011; Churchland & Kiani, 2016; Turner et al., 2017; Purcell & Palmeri, 2017; O'Connell et al., 2018; Turner et al., 2018; Busemeyer et al., 2019). Although beyond the scope of the present paper, it would be interesting to see how models that are responsive to both magnitude differences and overall magnitudes relate to neural signals, which may help identify a correspondence between magnitude-sensitivity and neural dynamics[1].

The observation of magnitude-sensitive reaction times is consistent with evolutionary and ecological arguments (Pais et al., 2013; Pirrone et al., 2014; Teodorescu et al., 2016; Bose et al., 2017); decision-makers in naturalistic scenarios are typically rewarded by the value of the option they select, rather than simply according to whether or not they selected the best option, thus their decision mechanisms should have been shaped by natural selection to optimise the trade-off between decision time and expected decision reward. In house-hunting honeybees, for instance, mutual inhibition (similar to the β-term in the LCA) facilitates the decision making process so that the swarm is able to break costly decision-deadlocks and choose one of the available nest sites (Seeley et al., 2012; Pais et al., 2013; Reina et al., 2017). In particular, an inverse relationship between the minimum strength of the cross-inhibition and the quality values of the nest-sites necessary to break decision-deadlock in the two-alternative case with equal options has been found (Pais et al., 2013). Another nonlinear decision making circuit involving interneuronal units embedded in a nutrition-based framework has also been shown to account for the kind of magnitude-sensitivity which is in qualitative agreement with experimental observations in perceptual decision making (Bose et al., 2019b). Notably, there is a close mathematical link between the study by Pais et al. (2013) and the present paper, as the dynamics of the model by Pais et al. (2013) corresponds to a change from a stable OU to an unstable OU process when a decision maker moves from decision-deadlock to deadlock-breaking. Similar results have been observed in models of competing neural accumulators describing simple forced choice experiments (Brown & Holmes, 2001). The link between OU processes and value-based decision making has also been established previously Busemeyer & Townsend (1992, 1993). However, magnitude-dependent noise in the OU model has not been studied systematically before, although we point out that Brunton et al. (2013) included a similar decay/growth term in their model too. Here we included magnitude-dependent noise in the stable and unstable OU processes and propose that this could be a useful addition to the conventional form of OU processes in value-based decisions.

[1]One of the co-authors (A.P.) is currently conducting a neuroimaging investigation into magnitude sensitivity the results of which will be published elsewhere.

While starting out as a primarily descriptive model, the classical DDM (Ratcliff, 1978; Ratcliff et al., 2016) has transformed into an optimality model through its link with statistically-optimal decision making (Bogacz et al., 2006); similarly mechanistic models such as the LCA can be made normative by considering under what parameterisations they approximate the DDM (Bogacz et al., 2006; Feng et al., 2009; Holmes & Cohen, 2014). As discussed above, for decisions in which the subject is rewarded by the value of the option chosen, rather than whether or not it was the best, it has been argued that the simple DDM is not normative (Pirrone et al., 2014). Recently, an optimality analysis of such value-based decisions was conducted by Tajima et al. (2016). The authors concluded that the optimal policy is implemented by a DDM with boundaries that collapse over time; however in the two-dimensional choice-value space that represents binary decisions, these boundaries are parallel and have slope one (Tajima et al., 2016). This means that the optimal policy does not predict magnitude-sensitive reaction times of the kind observed in Pirrone et al. (2018a,b); all equal-alternative choices are represented by points lying on the line with slope one that passes through the origin of the choice-value space, therefore as decision boundaries with slope one collapse over time, the hitting time of these boundaries will not vary with the magnitude of the equal choice values. Magnitude-sensitive reaction times therefore seem to be, for the model of Tajima et al. (2016), the kind of falsifying observation discussed above (Pirrone et al., 2018b). However, this model can account for such magnitude-sensitive reaction times by assuming non-linear utility functions, in which case the decision boundaries cease to have slope one (Tajima et al., 2016), and different equal-alternatives pairs can have different reaction times.

To conclude, as decision making theory is considered to link neural events and behaviour, and to provide a deeper understanding of cognitive functions (e.g. see Schall, 2004; Busemeyer & Diederich, 2010; Shadlen & Kiani, 2013; Hanks & Summerfield, 2017), our qualitative and quantitative comparison of magnitude-sensitive models may help in modelling decision making processes to build a bridge between the presentation of physical stimuli, evidence accumulation in neuronal units, and eventually behaviour caused by the decision process. Although our model comparison study was based on one-dimensional and two-dimensional neuron-like accumulators, it has previously been demonstrated that the activity of a large ensemble of redundant neurons map onto a single model accumulator (Zandbelt et al., 2014). Further to this, low dimensional magnitude-responsive models have been shown to be in agreement with recent experimental observations (Teodorescu et al., 2016; Pirrone et al., 2018a; Ratcliff et al., 2018). Low-dimensional accumulator models with absolute and relative magnitude-sensitive components may thus further enhance our understanding of neural design principles represented by biologically more realistic network models, such as the one introduced by Wang (2002) where inhibition is provided by interneurons in a large network of neuronal units. In particular, appropriate experimental realisations could help in model selection as magnitude-dependent noise may change model dynamics for different models in different ways, and it could increase our understanding

of similarities and differences between perceptual and value-based decisions, such as the effect of overall magnitude or value versus magnitude or value difference, and the origin of fluctuations affecting the decision process. In discriminating between competing hypotheses for magnitude-sensitive decision behaviour, whether input noise processing (Brunton et al., 2013; Teodorescu et al., 2016; Bose et al., 2019a), non-linear utility functions (Bogacz et al., 2007; Tajima et al., 2016), non-linear decision dynamics (Pais et al., 2013; Bose et al., 2019b; Roxin & Ledberg, 2008; Wang, 2002; Wong & Wang, 2006), or any other candidate explanation (for example see Kacelnik et al., 2011), further empirical and theoretical work will be required.

Supplementary Material

We made the simulation code as well as the data analysis code (with documentation) available via a public GitHub repository: https://github.com/DiODeProject/magnitude-sensitive-sequential-sampling-models. In this GitHub repository, we also provide an interactive Jupyter notebook that can be used to display fitting parameters (including uncertainties) for each of the five different parameter sets for each of the models (i.e. 25 tables with five parameter sets each). All tables are also available as csv-files in the repository.

Acknowledgments

We are grateful for helpful advice from Jerome Busemeyer (Indiana University Bloomington (US)), Andrei Teodorescu (University of Haifa (Israel)) and two anonymous reviewers of a previous (significantly different and unpublished) version of this manuscript, which motivated us to improve our modelling approach, to simulate a more challenging data set and to employ a more sensitive fitting method in the present paper.

Funding

The authors acknowledge funding by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement number 647704).

Appendix: Overview of BIC-tolerance values obtained from the fitting

Here, we summarise the BIC-values we used as tolerances values to achieve a successful termination of the model fit in table 3. For some model fits we had to vary this tolerance value to make sure that the fitting routine was set up in accordance with the possible degree of agreement between model and data. This means that the better the model fit the lower we could choose the BIC-tolerance value to improve our results. This was done iteratively, as mentioned in table 3.

In tables 4-8 we give an overview of the final value of the objective function, $\mathrm{BIC}_{\mathrm{fit}}$, of each computation. The numerical values shown represent mean value and corresponding standard deviation (values in brackets) of BIC-scores, $\mathrm{BIC}_{\mathrm{fit}}$, obtained from 6 model fit repetitions.

Table 3: Overview of BIC-tolerances used for fitting decision models to the simulated data. To find the model-specific tolerance values we decreased the BIC-tolerance in the following sequence: 2000, 1000, 500, 100 and chose the most suitable tolerance value which yielded a successful termination of the model fit.

	model used for fitting				
data generated by	pDDM	mDDM	mSOU	mUOU	LCA
pDDM	100	500	500	500	500
mDDM	500	100	500	500	500
mSOU	500	500	100	500	500
mUOU	500	500	500	100	500
LCA	$1000 - 2000$	$1000 - 2000$	$1000 - 2000$	$1000 - 2000$	100

Table 4: Overview of BIC-scores for models fitted to data generated by pDDM (corresponds to Fig. 4).

fits to pDDM data – data set 1					
fitted model	pDDM	mDDM	mSOU	mUOU	LCA
BIC_{fit}	361749 (18)	363153 (68)	361824 (18)	362395 (19)	366222 (216)
fits to pDDM data – data set 2					
fitted model	pDDM	mDDM	mSOU	mUOU	LCA
BIC_{fit}	367225 (12)	367364 (24)	367593 (29)	367282 (15)	371411 (122)
fits to pDDM data – data set 3					
fitted model	pDDM	mDDM	mSOU	mUOU	LCA
BIC_{fit}	367654 (12)	368514 (19)	367790 (16)	367816 (22)	370087 (41)
fits to pDDM data – data set 4					
fitted model	pDDM	mDDM	mSOU	mUOU	LCA
BIC_{fit}	374542 (14)	375709 (43)	374614 (5)	374603 (11)	382261 (104)
fits to pDDM data – data set 5					
fitted model	pDDM	mDDM	mSOU	mUOU	LCA
BIC_{fit}	377393 (11)	377555 (33)	377574 (18)	378358 (51)	380720 (57)

Table 5: Overview of BIC-score for models fitted to data generated by mDDM (corresponds to Fig. 5).

fits to mDDM data – data set 1					
fitted model	pDDM	mDDM	mSOU	mUOU	LCA
BIC_{fit}	391313 (116)	388505 (7)	388577 (15)	388577 (13)	389491 (43)
fits to mDDM data – data set 2					
fitted model	pDDM	mDDM	mSOU	mUOU	LCA
BIC_{fit}	390717 (54)	389468 (7)	389499 (16)	389482 (9)	389786 (39)
fits to mDDM data – data set 3					
fitted model	pDDM	mDDM	mSOU	mUOU	LCA
BIC_{fit}	379680 (87)	377569 (15)	377624 (11)	377647 (22)	378616 (61)
fits to mDDM data – data set 4					
fitted model	pDDM	mDDM	mSOU	mUOU	LCA
BIC_{fit}	390243 (51)	388795 (9)	388972 (14)	388846 (9)	389703 (33)
fits to mDDM data – data set 5					
fitted model	pDDM	mDDM	mSOU	mUOU	LCA
BIC_{fit}	391039 (55)	389655 (14)	389652 (11)	389671 (15)	390309 (32)

Table 6: Overview of BIC-score for models fitted to data generated by mSOU (corresponds to Fig. 6).

fits to mSOU data – data set 1					
fitted model	pDDM	mDDM	mSOU	mUOU	LCA
BIC_{fit}	393432 (132)	389515 (9)	389498 (6)	389535 (27)	390975 (82)
fits to mSOU data – data set 2					
fitted model	pDDM	mDDM	mSOU	mUOU	LCA
BIC_{fit}	394685 (168)	388785 (22)	388708 (6)	388787 (14)	391004 (59)
fits to mSOU data – data set 3					
fitted model	pDDM	mDDM	mSOU	mUOU	LCA
BIC_{fit}	393735 (81)	389050 (14)	389002 (12)	389071 (14)	390235 (50)
fits to mSOU data – data set 4					
fitted model	pDDM	mDDM	mSOU	mUOU	LCA
BIC_{fit}	391497 (88)	388716 (19)	388672 (14)	388737 (16)	389981 (63)
fits to mSOU data – data set 5					
fitted model	pDDM	mDDM	mSOU	mUOU	LCA
BIC_{fit}	391445 (53)	389610 (9)	389557 (5)	389566 (7)	392268 (61)

Table 7: Overview of BIC-score for models fitted to data generated by mUOU (corresponds to Fig. 7).

fits to mUOU data – data set 1					
fitted model	pDDM	mDDM	mSOU	mUOU	LCA
$\mathrm{BIC}_{\mathrm{fit}}$	391621 (63)	391765 (53)	391678 (57)	389939 (7)	390058 (17)
fits to mUOU data – data set 2					
fitted model	pDDM	mDDM	mSOU	mUOU	LCA
$\mathrm{BIC}_{\mathrm{fit}}$	391199 (75)	389828 (27)	389497 (52)	388654 (9)	390008 (65)
fits to mUOU data – data set 3					
fitted model	pDDM	mDDM	mSOU	mUOU	LCA
$\mathrm{BIC}_{\mathrm{fit}}$	391688 (48)	389835 (13)	389915 (11)	389794 (5)	390949 (83)
fits to mUOU data – data set 4					
fitted model	pDDM	mDDM	mSOU	mUOU	LCA
$\mathrm{BIC}_{\mathrm{fit}}$	392088 (106)	389278 (15)	389222 (11)	389058 (9)	390052 (65)
fits to mUOU data – data set 5					
fitted model	pDDM	mDDM	mSOU	mUOU	LCA
$\mathrm{BIC}_{\mathrm{fit}}$	390811 (49)	389383 (23)	389115 (17)	389012 (11)	389770 (28)

Table 8: Overview of BIC-score for models fitted to data generated by LCA (corresponds to Fig. 8).

fits to LCA data – data set 1					
fitted model	pDDM	mDDM	mSOU	mUOU	LCA
$\mathrm{BIC}_{\mathrm{fit}}$	495374 (656)	493745 (507)	491444 (372)	393443 (174)	377335 (13)
fits to LCA data – data set 2					
fitted model	pDDM	mDDM	mSOU	mUOU	LCA
$\mathrm{BIC}_{\mathrm{fit}}$	433773 (279)	431619 (113)	432621 (266)	382440 (103)	372313 (16)
fits to LCA data – data set 3					
fitted model	pDDM	mDDM	mSOU	mUOU	LCA
$\mathrm{BIC}_{\mathrm{fit}}$	431791 (297)	430826 (114)	430822 (388)	390186 (83)	381242 (9)
fits to LCA data – data set 4					
fitted model	pDDM	mDDM	mSOU	mUOU	LCA
$\mathrm{BIC}_{\mathrm{fit}}$	395535 (179)	400044 (141)	395589 (173)	378072 (18)	376895 (10)
fits to LCA data – data set 5					
fitted model	pDDM	mDDM	mSOU	mUOU	LCA
$\mathrm{BIC}_{\mathrm{fit}}$	474000 (566)	477170 (1013)	471662 (233)	392022 (90)	385112 (9)

References

Basten, U., Biele, G., Heekeren, H. R., & Fiebach, C. J. (2010). How the brain integrates costs and benefits during decision making. *Proceedings of the National Academy of Sciences*, *107*, 21767–21772. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.0908104107. doi:10.1073/pnas.0908104107.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol. Rev.*, *113*, 700–765. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-295X.113.4.700. doi:10.1037/0033-295X.113.4.700.

Bogacz, R., Usher, M., Zhang, J., & McClelland, J. L. (2007). Extending a biologically inspired model of choice: multi-alternatives, nonlinearity and value-based multidimensional choice. *Phil. Trans. R. Soc. B*, *362*, 1655–1670. doi:10.1098/rstb.2007.2059.

Bose, T., Bottom, F., Reina, A., & Marshall, J. A. R. (2019a). Frequency-sensitivity and magnitude-sensitivity in decision-making: Predictions of a theoretical model-based study. *Computational Brain & Behavior*, . URL: https://doi.org/10.1007/s42113-019-00031-4. doi:10.1007/s42113-019-00031-4.

Bose, T., Reina, A., & Marshall, J. A. (2017). Collective decision-making. *Curr. Opin. Behav. Sci.*, *16*, 30–34. doi:10.1016/j.cobeha.2017.03.004.

Bose, T., Reina, A., & Marshall, J. A. R. (2019b). Inhibition and excitation shape activity selection: Effect of oscillations in a decision-making circuit. *Neural Comput.*, *31*, 870–896. doi:10.1162/neco_a_01185. PMID: 30883280.

Brown, E., & Holmes, P. (2001). Modeling a simple choice task: Stochastic dynamics of mutually inhibitory neural groups. *Stochastics and Dynamics*, *1*, 159–191. doi:10.1142/S0219493701000102.

Brunton, B. W., Botvinick, M. M., & Brody, C. D. (2013). Rats and Humans Can Optimally Accumulate Evidence for Decision-Making. *Science*, *340*, 95–98. doi:10.1126/science.1233912.

Busemeyer, J., Gluth, S., Rieskamp, J., & Turner, B. M. (2019). Cognitive and neural bases of multi-attribute, multi-alternative, value-based decisions. *Tr. Cogn. Sci.*, *23*, 251–263. URL: https://www.cell.com/trends/cognitive-sciences/fulltext/S1364-6613(18)30284-5. doi:10.1016/j.tics.2018.12.003.

Busemeyer, J. R., & Diederich, A. (2010). *Cognitive modeling*. Sage. URL: https://books.google.co.uk/books?hl=en{&}lr={&}id=R7KDF35g5LQC{&}oi=fnd{&}pg=PP2{&}ots=YT-7M4syLY{&}sig=CUWv60pWjBDEfhhAPtTtGLNSUJg{&}redir{_}esc=y{#}v=onepage{&}q{&}f=false.

Busemeyer, J. R., & Townsend, J. T. (1992). Fundamental derivations from decision field theory. *Math. Soc. Sci.*, *23*, 255–282. URL: https://www.sciencedirect.com/science/article/pii/0165489692900435. doi:10.1016/0165-4896(92)90043-5.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychol. Rev.*, *100*, 432–459. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-295X.100.3.432. doi:10.1037/0033-295X.100.3.432.

Cavanagh, J. F., Wiecki, T. V., Kochar, A., & Frank, M. J. (2014). Eye tracking and pupillometry are indicators of dissociable latent decision processes. *J. Experiment. Psychol.: Gen.*, *143*, 1476–1488. doi:10.1037/a0035813.

Churchland, A. K., & Kiani, R. (2016). Three challenges for connecting model to mechanism in decision-making. *Current Opinion in Behavioral Sciences*, *11*, 74–80. URL: https://www.sciencedirect.com/science/article/pii/S2352154616301243. doi:10.1016/J.COBEHA.2016.06.008.

Churchland, A. K., Kiani, R., Chaudhuri, R., Wang, X.-J., Pouget, A., & Shadlen, M. (2011). Variance as a Signature of Neural Computations during Decision Making. *Neuron*, *69*, 818–831. URL: https://www.sciencedirect.com/science/article/pii/S0896627310010871http://linkinghub.elsevier.com/retrieve/pii/S0896627310010871. doi:10.1016/j.neuron.2010.12.037.

Diederich, A., & Oswald, P. (2014). Sequential sampling model for multiattribute choice alternatives with random attention time and processing order. *Frontiers in Human Neuroscience*, *8*, 697. URL: https://www.frontiersin.org/article/10.3389/fnhum.2014.00697. doi:10.3389/fnhum.2014.00697.

Diederich, A., & Oswald, P. (2016). Multi-stage sequential sampling models with finite or infinite time horizon and variable boundaries. *J. Math. Psychol.*, *74*, 128–145. doi:10.1016/j.jmp.2016.02.010.

Feng, S., Holmes, P., Rorie, A., & Newsome, W. T. (2009). Can Monkeys Choose Optimally When Faced with Noisy Stimuli and Unequal Rewards? *PLoS Comput. Biol.*, *5*, e1000284. URL: http://dx.plos.org/10.1371/journal.pcbi.1000284. doi:10.1371/journal.pcbi.1000284.

Forstmann, B., Ratcliff, R., & Wagenmakers, E.-J. (2016). Sequential Sampling Models in Cognitive Neuroscience: Advantages, Applications, and Extensions. *Ann. Rev. Psychol.*, *67*, 641–666. URL: http://www.annualreviews.org/doi/10.1146/annurev-psych-122414-033645. doi:10.1146/annurev-psych-122414-033645.

Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychol. Rev.*, *96*, 267–314. URL: http://psycnet.apa.org/doiLanding?doi=10.1037{%}2F0033-295X.96.2.267. doi:http://dx.doi.org/10.1037/0033-295X.96.2.267.

Gold, J. I., & Shadlen, M. N. (2007). The Neural Basis of Decision Making. *Annu. Rev. Neurosci.*, *30*, 535–574. doi:10.1146/annurev.neuro.29.051605.113038.

Hanks, T. D., & Summerfield, C. (2017). Perceptual Decision Making in Rodents, Monkeys, and Humans. *Neuron*, *93*, 15–31. URL: https://www.sciencedirect.com/science/article/pii/S0896627316309424https://linkinghub.elsevier.com/retrieve/pii/S0896627316309424. doi:10.1016/j.neuron.2016.12.003.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychon. Bull. Rev.*, *9*, 394–401. URL: http://www.springerlink.com/index/10.3758/BF03196299. doi:10.3758/BF03196299.

Heathcote, A., Wagenmakers, E.-J., & Brown, S. D. (2014). The falsifiability of actual decision-making models. *Psychol.Rev.*, *121*, 676–678. doi:10.1037/a0037771.

Holmes, P., & Cohen, J. D. (2014). Optimality and some of its discontents: Successes and shortcomings of existing models for binary decisions. *Topics Cogn. Sci.*, *6*, 258–278. doi:10.1111/tops.12084.

Hunt, L. T., Kolling, N., Soltani, A., Woolrich, M. W., Rushworth, M. F. S., & Behrens, T. E. J. (2012). Mechanisms underlying cortical activity during value-guided choice. *Nat. Neurosci.*, *15*, 470–476. URL: http://dx.doi.org/10.1038/nn.3017. doi:10.1038/nn.3017.

Jones, M., & Dzhafarov, E. N. (2014). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. *Psychological Review*, *121*, 1–32. URL: http://www.ncbi.nlm.nih.gov/pubmed/24079307http://doi.apa.org/getdoi.cfm?doi=10.1037/a0034190. doi:10.1037/a0034190.

Kacelnik, A., Vasconcelos, M., Monteiro, T., & Aw, J. (2011). Darwin's tug-of-war vs. starlings' horse-racing: how adaptations for sequential encounters drive simultaneous choice. *Behav. Ecol. Sociobiol.*, *65*, 547–558. URL: http://link.springer.com/10.1007/s00265-010-1101-2. doi:10.1007/s00265-010-1101-2.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nat. Neurosci.*, *13*, 1292–1298. URL: http://www.ncbi.nlm.nih.gov/pubmed/20835253http://www.nature.com/articles/nn.2635. doi:10.1038/nn.2635.

Krajbich, I., Hare, T., Bartling, B., Morishima, Y., & Fehr, E. (2015). A Common Mechanism Underlying Food Choice and Social Decisions. *PLoS Comput. Biol.*, *11*, e1004371. URL: http://www.econ.uzh.ch/faculty/fehr/http://dx.plos.org/10.1371/journal.pcbi.1004371. doi:10.1371/journal.pcbi.1004371.

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proc. Nat. Acad. Sci.*, *108*, 13852–13857. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.1101328108. doi:10.1073/pnas.1101328108.

van Maanen, L., Grasman, R. P. P. P., Forstmann, B. U., & Wagenmakers, E.-J. (2012). Piéron's Law and Optimal Behavior in Perceptual Decision-Making. *Front. Neurosci.*, *5*, 143. doi:10.3389/fnins.2011.00143.

Miletić, S., Turner, B. M., Forstmann, B. U., & van Maanen, L. (2017). Parameter recovery for the Leaky Competing Accumulator model. *J. Math. Psychol.*, *76*, 25–50. URL: https://www.sciencedirect.com/science/article/pii/S0022249616301663https://linkinghub.elsevier.com/retrieve/pii/S0022249616301663. doi:10.1016/j.jmp.2016.12.001.

Mulder, M., van Maanen, L., & Forstmann, B. (2014). Perceptual decision neurosciences A model-based review. *Neuroscience*, *277*, 872–884. URL: https://www.sciencedirect.com/science/article/pii/S0306452214006046https://linkinghub.elsevier.com/retrieve/pii/S0306452214006046. doi:10.1016/j.neuroscience.2014.07.031.

Nelder, J. A., & Mead, R. (1965). A Simplex Method for Function Minimization. *Computer Journal*, *7*, 308–313.

O'Connell, R. G., Shadlen, M. N., Wong-Lin, K., & Kelly, S. P. (2018). Bridging Neural and Computational Viewpoints on Perceptual Decision-Making. *Trends Neurosci.*, *in press*. URL: http://www.ncbi.nlm.nih.gov/pubmed/30007746. doi:10.1016/j.tins.2018.06.005.

Pais, D., Hogan, P. M., Schlegel, T., Franks, N. R., Leonard, N. E., & Marshall, J. A. R. (2013). A Mechanism for Value-Sensitive Decision-Making. *PLoS ONE*, *8*, e73216. URL: http://dx.plos.org/10.1371/journal.pone.0073216. doi:10.1371/journal.pone.0073216.

Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *J. Vision*, *5*, 1. URL: http://journalofvision.org/5/5/1/http://jov.arvojournals.org/article.aspx?doi=10.1167/5.5.1. doi:10.1167/5.5.1.

Pins, D., & Bonnet, C. (1996). On the relation between stimulus intensity and processing time: Piéron's law and choice reaction time. *Percept. Psychophys.*, *58*, 390–400. doi:10.3758/BF03206815.

Pirrone, A., Azab, H., Hayden, B. Y., Stafford, T., & Marshall, J. A. R. (2018a). Evidence for the speed-value trade-off: human and monkey decision making is magnitude sensitive. *Decision*, *5*, 129–142.

Pirrone, A., Stafford, T., & Marshall, J. A. R. (2014). When natural selection should optimize speed-accuracy trade-offs. *Front. Neurosci.*, *8*, 73. doi:10.3389/fnins.2014.00073.

Pirrone, A., Wen, W., & Li, S. (2018b). Single-trial dynamics explain magnitude sensitive decision making. *BMC Neuroscience*, *19*, 54. URL: https://bmcneurosci.biomedcentral.com/articles/10.1186/s12868-018-0457-5. doi:10.1186/s12868-018-0457-5.

Pirrone, L. S., A. (2018). Lateral inhibition between evidence accumulators explains magnitude sensitivity in perceptual decision making. Talk presented at the 41st European Conference on Visual Perception. Trieste, Italy.

Polanía, R., Krajbich, I., Grueschow, M., & Ruff, C. C. (2014). Neural Oscillations and Synchronization Differentially Support Evidence Accumulation in Perceptual and Value-Based Decision Making. *Neuron*, *82*, 709–720. doi:10.1016/j.neuron.2014.03.014.

Purcell, B. A., & Palmeri, T. J. (2017). Relating accumulator model parameters and neural dynamics. *J. Math. Psychol.*, *76*, 156–171. URL: https://www.sciencedirect.com/science/article/pii/S0022249616300499http://linkinghub.elsevier.com/retrieve/pii/S0022249616300499. doi:10.1016/j.jmp.2016.07.001.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychol. Rev.*, *85*, 59–108. URL: http://content.apa.org/journals/rev/85/2/59. doi:10.1037/0033-295X.85.2.59.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for decisions between two choices. *Psychol. Sci.*, *9*, 347–356.

Ratcliff, R., & Smith, P. L. (2004). A Comparison of Sequential Sampling Models for Two-Choice Reaction Time. *Psychol. Rev.*, *111*, 333–367. doi:10.1037/0033-295X.111.2.333. arXiv:NIHMS150003.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends Cogn. Sci.*, *20*, 260–281. doi:10.1016/j.tics.2016.01.007.

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychon. Bull. Rev.*, *9*, 438–481. doi:10.3758/BF03196302. arXiv:NIHMS150003.

Ratcliff, R., Voskuilen, C., & Teodorescu, A. (2018). Modeling 2-alternative forced-choice tasks: Accounting for both magnitude and difference effects. *Cogn. Psychol.*, *103*, 1–22. doi:10.1016/j.cogpsych.2018.02.002.

Reina, A., Bose, T., Trianni, V., & Marshall, J. A. R. (2018). Psychophysical laws and the superorganism. *Scientific Reports*, *8*, 4387. URL: http://www.nature.com/articles/s41598-018-22616-y. doi:10.1038/s41598-018-22616-y.

Reina, A., Marshall, J. A. R., Trianni, V., & Bose, T. (2017). Model of the best-of-n nest-site selection process in honeybees. *Phys. Rev. E*, *95*, 052411. URL: https://link.aps.org/doi/10.1103/PhysRevE.95.052411. doi:10.1103/PhysRevE.95.052411.

Roxin, A., & Ledberg, A. (2008). Neurobiological models of two-choice decision making can be reduced to a one-dimensional nonlinear diffusion equation. *PLoS Comput. Biol.*, *4*, e1000046. URL: http://dx.plos.org/10.1371/journal.pcbi.1000046. doi:10.1371/journal.pcbi.1000046.

Schall, J. D. (2004). On Building a Bridge Between Brain and Behavior. *Ann. Rev. Psychol.*, *55*, 23–50. URL: http://www.annualreviews.org/doi/10.1146/annurev.psych.55.090902.141907. doi:10.1146/annurev.psych.55.090902.141907.

Seeley, T. D., Visscher, P. K., Schlegel, T., Hogan, P. M., Franks, N. R., & Marshall, J. a. R. (2012). Stop signals provide cross inhibition in collective decision-making by honeybee swarms. *Science*, *335*, 108–11. URL: http://www.ncbi.nlm.nih.gov/pubmed/22157081. doi:10.1126/science.1210361.

Shadlen, M., & Newsome, W. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J. Neurophysiol.*, *86*, 1916–1936. URL: http://www.ncbi.nlm.nih.gov/pubmed/11600651.

Shadlen, M. N., & Kiani, R. (2013). Decision Making as a Window on Cognition. *Neuron*, *80*, 791–806. doi:10.1016/j.neuron.2013.10.047.

Shadlen, M. N., & Newsome, W. T. (1996). Motion perception: seeing and deciding. *Proc. Natl. Acad. Sci.*, *93*, 628–633. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.93.2.628. doi:10.1073/pnas.93.2.628.

Simen, P., Vlasov, K., & Papadakis, S. (2016). Scale (in)variance in a unified diffusion model of decision making and timing. *Psychol. Rev.*, *123*, 151–181. URL: http://psycnet.apa.org/doiLanding?doi=10.1037{%}2Frev0000014.

Smith, P. L., Ratcliff, R., & McKoon, G. (2014). The diffusion model is not a deterministic growth model: Comment on Jones and Dzhafarov (2014). *Psychol. Rev.*, *121*, 679–688. URL: http://psycnet.apa.org/doiLanding?doi=10.1037{%}2Fa0037667.

Smith, S. M., & Krajbich, I. (2019). Gaze amplifies value in decision making. *Psychological Science*, *30*, 116–128. URL: https://doi.org/10.1177/0956797618810521. doi:10.1177/0956797618810521. arXiv:https://doi.org/10.1177/0956797618810521. PMID: 30526339.

Stafford, T., & Gurney, K. N. (2004). The role of response mechanisms in determining reaction time performance: Piéron's law revisited. *Psychon. Bull. Rev.*, *11*, 975–87. URL: http://www.ncbi.nlm.

nih.gov/pubmed/15875968http://www.springerlink.com/index/10.3758/BF03196729. doi:10.3758/BF03196729.

Tajima, S., Drugowitsch, J., & Pouget, A. (2016). Optimal policy for value-based decision-making. *Nat. Commun.*, *7*, 12400. URL: http://dx.doi.org/10.1038/ncomms12400http://www.nature.com/doifinder/10.1038/ncomms12400. doi:10.1038/ncomms12400.

Teodorescu, A. R., Moran, R., & Usher, M. (2016). Absolutely relative or relatively absolute: violations of value invariance in human decision making. *Psychon. Bull. Rev.*, *23*, 22–38. URL: http://link.springer.com/10.3758/s13423-015-0858-8http://www.ncbi.nlm.nih.gov/pubmed/26022836. doi:10.3758/s13423-015-0858-8.

Teodorescu, A. R., & Usher, M. (2013). Disentangling decision models: From independence to competition. *Psychol. Rev.*, *120*, 1–38. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/a0030776. doi:10.1037/a0030776.

Turner, B. M., Forstmann, B. U., Love, B. C., Palmeri, T. J., & Van Maanen, L. (2017). Approaches to analysis in model-based cognitive neuroscience. *J. Math. Psychol.*, *76*, 65–79. URL: https://www.sciencedirect.com/science/article/pii/S0022249616000031https://linkinghub.elsevier.com/retrieve/pii/S0022249616000031. doi:10.1016/j.jmp.2016.01.001.

Turner, B. M., Rodriguez, C. A., Liu, Q., Molloy, M. F., Hoogendijk, M., & McClure, S. M. (2018). On the neural and mechanistic bases of self-control. *Cerebr. Cort.*, *29*, 732–750. URL: https://academic.oup.com/cercor/article-abstract/29/2/732/4823219?redirectedFrom=fulltext. doi:10.1093/cercor/bhx355.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychol. Rev.*, *108*, 550–592. URL: http://dx.doi.org/10.1037/0033-295X.108.3.550. doi:10.1037/0033-295X.108.3.550.

Wang, X. J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. *Neuron*, *36*, 955–968. doi:10.1016/S0896-6273(02)01092-9.

Wong, K.-F., & Wang, X.-J. (2006). A Recurrent Network Mechanism of Time Integration in Perceptual Decisions. *Journal of Neuroscience*, *26*, 1314–1328. doi:10.1523/JNEUROSCI.3733-05.2006.

Zandbelt, B., Purcell, B. A., Palmeri, T. J., Logan, G. D., & Schall, J. D. (2014). Response times from ensembles of accumulators. *Proc. Natl. Acad. Sci.*, *111*, 2848–53. URL: http://www.ncbi.nlm.nih.gov/pubmed/24550315http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3932860. doi:10.1073/pnas.1310577111.