UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Sensorimotor processes are not a source of much noise: Sensory-motor and decision components of reaction times

Permalink

https://escholarship.org/uc/item/1nj6m2n7

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

Authors

Meibodi, Neda
Schubö, Anna
Endres, Dominik M

Publication Date

2022

Peer reviewed

Sensorimotor processes are not a source of much noise: sensorimotor and decision components of reaction times

Neda Meibodi (meibodi@uni-marburg.de)

Anna Schubö (schuboe@uni-marburg.de)

Dominik Endres (dominik.endres@uni-marburg.de)

Department of Psychology, Philipps-University Marburg, Gutenbergstrasse 18,
35032 Marburg, Germany

Abstract

Statistical descriptions of reaction times are central components of quantitative attention models. It is often assumed that total reaction time is comprised of various components, e.g. sensory delays, decision making and motor execution contributions. We use machine learning to decompose observed total reaction times into sensorimotor and decision components, and evaluate which model assumptions maximize approximate Bayesian model evidence (free energy or evidence lower bound). We find that an inverse Gaussian decision time distribution combined with a very narrow Gaussian sensorimotor distribution can best explain human reaction time data. We also model outliers explicitly by a uniform background distribution. We find that the model assigns a small fraction of datapoints to this outlier distribution.

Keywords: Decision component; Expectation maximization; Free energy; Inverse Gaussian; Mixture models; Reaction time; Sensorimotor component; Visual attention.

Introduction

Reaction time has been widely used as a measure of cognitive processes, e.g. in attention research. It is believed that total observed reaction time (RT) is a sum of different time components. As Luce (1986) mentioned at least five processes may contribute to a total reaction time: physical input transduction into neural spikes, spike transmission to the brain, signal processing and motor programming for the target muscle group (we call this part *decision time*), signal transmission to the muscles and eventually muscle contraction. As it is hard to observe all these components separately, we stack them all – except decision time – and call them the *sensorimotor component* of reaction time. This component is commonly called residual latency (Luce, 1986) or non-decision time (Ratcliff & Tuerlinckx, 2002). During the last decades, there have been many proposals and investigations on how these components combine to yield the final RT distributions.

In some older research (Christie & Luce, 1956; Hohle, 1965), it is reported that RT is a sum of a Gaussian and a exponentially distributed component, where one represents the decision time and another represents the motor component. Consequently, the Ex-Gaussian distribution, which results from convolving these two distributions has been used for modelling RT distributions and cognitive processes (Ratcliff, 1978; Hohle, 1965; Fitousi, 2020; Meibodi, Abbasi, Schubö, & Endres, 2021b) and also psychological disorders (Hwang-Gu et al., 2019). Meibodi et al. (2021b) proposed a model

of visual attention which predicts parameters of RT distributions. In that study, an analysis of RTs showed that the ex-Gaussian is a better descriptor than other commonly used distributions, followed by an inverse Gaussian (Meibodi, Abbasi, Schubö, & Endres, 2021a). The authors of that study modelled total RTs without considering a decomposition into separate components, we would like to remedy this shortcoming here. However, the ex-Gaussian has several features that are theoretically not convincing (Schwarz, 2001): first, the Gaussian component has been linked to either the decision or the motor process. Since both processes must take a positive amount of time, a (wide) Gaussian is not a plausible distribution. Second, there is no compelling connection between the distribution parameters and theoretical accounts of the origin of reaction times. Third, the hazard function of the ex-Gaussian is increasing although the best descriptive RT distributions have been reported to have peaked hazard functions (Maddox, Ashby, & Gottlob, 1998).

To address these issues, Schwarz (2001) proposed the ex-Wald distribution for RTs, which is a convolution of an inverse Gaussian with an exponential. Here, the inverse Gaussian describes the decision time, whereas the non-decision component is distributed exponentially (Palmer, Horowitz, Torralba, & Wolfe, 2011). One appealing feature of the ex-Wald is the inverse Gaussian component which models the first passage time distribution of a random walk (Folks & Chhikara, 1978). Such random walks describe quasi-Bayesian sensory evidence accumulation, or drift-diffusion processes. On the other hand, the claim that non-decision time has an exponential distribution seems unjustified. Although the exponential component is commonly interpreted as the effect of a residual process, there are some controversial opinions mentioning that the exponential effect on RT distribution just reflects the search process in visual search tasks (Horowitz & Wolfe, 2003; Palmer et al., 2011).

In (Ratcliff & Tuerlinckx, 2002)'s drift diffusion model (DDM), the parameter T_{er} denotes the time that is spent on processes other than the decision making – such as stimulus encoding, response output and memory access. The parameter has variability to correct the model fits on different data sets under variety of conditions. In this model non-decision time is uniformly distributed (Ratcliff & Tuerlinckx, 2002; Hawkins, Forstmann, Wagenmakers, Ratcliff, & Brown, 2015) although it is mentioned that the true dis-

tribution might be normal (Wiecki, Sofer, & Frank, 2013) or skewed. Ratcliff claimed that the shape of the reaction time distribution is primarily determined by the shape of the decision component and the precise shape of the non-decision distribution has a small effect on that as the former has a very large standard deviation (Ratcliff & Tuerlinckx, 2002; Ratcliff & Smith, 2004; Ratcliff, 2006; Ratcliff & McKoon, 2008; Ratcliff & Childers, 2015). One drawback of that model is that the non-decision component happens before and after decision part (Ratcliff & McKoon, 2008) although the result of some studies indicated that these component are intertwined (Evans & Wagenmakers, 2020).

The mean of the non-decision time reported in DDM is about 300 ms with a standard deviation in range 3 to 10 (Ratcliff & Tuerlinckx, 2002). The reported range of non-decision time might differ in other studies based on the experiment, apparatus or participants' attributes. Using simple reaction time (SRT) experiments, previous research has tried to determine how response delay is influenced by features such as: colour of stimuli (Amini Vishteh, Mirzajani, Jafarzadehpour, & Darvishpour, 2019), participants' age (Jain, Bansal, Kumar, & Singh, 2015; Woods, Wyma, Yund, Herron, & Reed, 2015), gender (Dykiert, Der, Starr, & Deary, 2012; Jain et al., 2015), physical activities (Jain et al., 2015) or computer hardware and software (Dodonova & Dodonov, 2013). In a typical SRT study, participants have to press a key as soon as they see the stimulus on the screen (Ulrich & Stapf, 1984). We therefore assume that a SRT contains only a very short decision component and that it is dominated by the sensory and motor processing times. Hence, a SRT approximates the part of a RT which we call the SM component. See Table 1 for an overview of the reported results.

The importance of good RT distribution models is their applicability to statistical analysis and to the modelling of cognitive psychological processes. Most psychologists are interested in decision component of RT and look at the rest of it (commonly called residual latency) as a nuisance variable that should be subtracted from RT (Luce, 1986). However, in addition to decision component analysis, looking at residual latency is also informative. For instance the result of Pedersen, Frank, and Biele (2017) showed that longer RTs in medicated ADHD participants arose because of a strong increase in their non-decision (residual latency) time. Ratcliff, Thapar, and McKoon (2001) found that in some tasks, slower responses of older participants can also be the effect of longer non-decision time.

In this paper we try to disentangle the components of RTs and to recognize outlier responses using a machine learning approach derived from free energy minimization (Friston, Kilner, & Harrison, 2006). We do this with the aim of making attention models, e.g. the one presented by Meibodi et al. (2021b) more interpretable in terms of the underlying psychological processes. We investigate several proposals for the distribution of the SM component: Gaussian, gamma and Laplace. Our motivation for testing the Gaussian distribu-

tion is its popularity in previous research, e.g. Christie and Luce (1956); Hohle (1965); Ratcliff and Tuerlinckx (2002) as discussed above. The Laplace distribution has heavier tails rather than the Gaussian and might therefore be less sensitive to extreme SM variations. Both distributions are supported on $\mathbb{R}$ and assign non-zero probability to negative SM components, which is implausible. We therefore experimented with the gamma distribution that has a positive support. Furthermore, since motor output is driven by neuronal spiking activity, its timing would be determined by spike arrival at the neuromuscular synapses. The gamma distribution has been used before to model inter-spike intervals (Ostojic, 2011). In the next section, we will describe the models, followed by a short description of the database used for learning. We then present model comparison results, which indicate that an inverse Gaussian decision time distribution combined with a very narrow Gaussian sensorimotor distribution can best explain human reaction time data. We also model outliers explicitly by a uniform background distribution. We find that the model assigns only a small fraction of datapoints to this outlier distribution. Finally, we discuss the implications of our findings.

Methods

We model an RTs as mixtures of two models ($M = 0$ and $M = 1$) as shown in Figure 1. If $M = 1$ (the response model), then an RT has two components, namely 'decision' and 'sensorimotor (SM)'. If $M = 0$, then the RT is assumed to be an outlier which is drawn from a uniform distribution in range $(0, t_{max})$, i.e. an outlier response has no relationship to the task other than its occurrence before the trial's end at t_{max}. In this case, all we know about the response is that it may happen at any time point in $[t_{min}, t_{max}]$, which is captured by the uniform distribution. For $M = 1$, we assume that the decision component can be viewed as the first passage time in a Wiener diffusion process, which is a model of Bayesian evidence accumulation. Thus, the distribution of the decision component is an inverse Gaussian (Folks & Chhikara, 1978; Schwarz, 2001). The SM component's distribution precise shape has no clear theoretical motivation, hence we try to determine it

Table 1: Mean and standard deviation (SD) of some SRT experiments. The smallest and largest reported mean can be seen in the table for each study. The reported means vary based on between-group differences such as participants' age/gender or stimuli features.

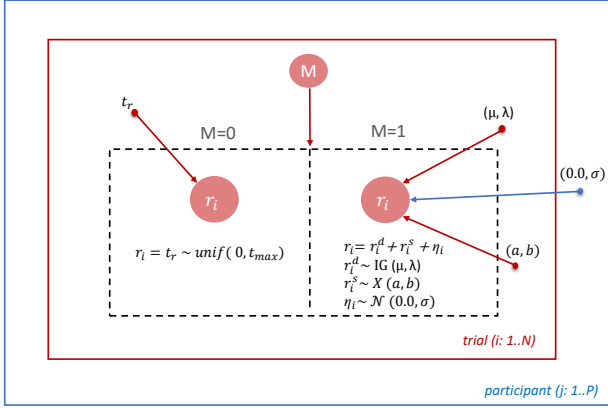| Study	Mean $\pm SD$ (ms)
Amini Vishteh et al. (2019)	207.88 ±7.14
	224.39 ±15.62
Jain et al. (2015)	217.13 ±12.60
	256.36 ±20.34
Woods et al. (2015)	217.9 ±19.5
	239.1 ±28.1

Figure 1: Mixture model of inverse Gaussian and uniform distributions. The blue and red boxes represent participants and trials respectively. P is the number of participants and N is the number of trials for each participant. Each trial's reaction time (r_i) can follow one of the models (M=0 or M=1). When M=0, r_i is a sample of uniform distribution $(0, t_{max})$ and t_{max} is maximum time window that participants had for each trial. In that case, the assumption is that RT can not be decomposed into SM and decision components and it is considered as an outlier. When M=1 then each r_i has 2 main components –decision component (r_i^d) and SM component (r_i^s)– and also noise (η_i) which is normally distributed. The assumption is that (r_i^s) can be a sample from a distribution (X) with parameters a and b. Here, X is either Gaussian, gamma or Laplace which we determine by model comparison.

by model comparison between a Gaussian, a Laplacian and a Gamma distribution. Note that the Gaussian can only be a suitable candidate if it is so narrow that the probability for a negative RT is virtually zero.

We assume that there is one SM distribution per participant, as shown in Figure 1, Thus, each reaction time (r_i) is a sum of a decision component (r_i^d), SM component (r_i^s) and measurement noise (η_i)

$$r_i = r_i^s + r_i^d + \eta_i$$
$$\eta_i \sim \mathcal{N}(0.0, \sigma)$$
$$r_i^s \sim X(a, b) \qquad (1)$$
$$r_i^d \sim IG(\mu, \lambda)$$

where IG is inv-Gaussian (inverse Gaussian) distribution and X is the SM distribution. We tested the model with gamma, Gaussian and Laplace as the X distribution (see the models comparison in result section). Normally distributed noise (η_i) describes the random effects on the measurement process.

Since exact inference is intractable in this model, we are instead maximizing a lower bound on the expected log likelihood, or evidence lower bound (ELBO) (Bishop, 2006) a.k.a.

free energy (Friston, 2003). The ELBO of our model is

$$L = \int dr_i^s \int dr_i^d \sum_{M_i=0}^{1} \sum_{i=1}^{N} \Bigg[M_i \big[\log p(r_i | r_i^s, r_i^d, \sigma) +$$
$$- \log \big(q(r_i^d | \ddot{\theta}_i^d) / p(r_i^d | \theta^d) \big) - \log \big(q(r_i^s | \ddot{\theta}_i^s) / p(r_i^s | \theta^s) \big) \big]$$
$$+ (1 - M_i) \log p(r_i) - \log \big(q(M_i) / p(M_i) \big) \Bigg]$$
$$q(M_i) \big(q(r_i^s | \ddot{\theta}_i^s) \, q(r_i^d | \ddot{\theta}_i^d) \big)^{M_i} \qquad (2)$$

where N is the number of trials for each participant, M is the model type, q is the variational posterior distribution and θ includes inv-Gaussian parameters for each r_i when $M_i = 1$. θ^s and θ^d are prior parameters on the decision component (μ, λ) and SM component (a, b) and $\ddot{\theta}_i^s$, $\ddot{\theta}_i^d$ are posterior parameters on the same components for each trial (i). When $M = 0$, the RT can not be decomposed, which is modelled by the uniform distribution $p(r_i)$ (see Figure 1).

Using the usual definition of the Kullback-Leibler divergence (KL) between distributions q and p

$$KL(q(x) | p(x)) = \int dx \, q(x) \big[\log(q(x) / p(x)) \big] \qquad (3)$$

we can rewrite Eq 2 as

$$L = \sum_{i=1}^{N} \Bigg[q(M_i = 1) \big[\langle \log p(r_i | r_i^s, r_i^d, \sigma) \rangle_{q(r_i^s | \ddot{\theta}_i^s) \, q(r_i^d | \ddot{\theta}_i^d)}$$
$$- KL(q(r_i^s | \ddot{\theta}_i^s) | p(r_i^s | \theta^s)) - KL(q(r_i^d | \ddot{\theta}_i^d) | p(r_i^d | \theta^d)) \big] \qquad (4)$$
$$+ q(M_i = 0) \, \log p(r_i) - KL(q(M_i) | p(M_i)) \Bigg]$$

where $\langle p(r_i | r_i^s, r_i^d, \sigma) \rangle_{q(r_i^s | \ddot{\theta}_i^s) \, q(r_i^d | \ddot{\theta}_i^d)}$ is the expectation of the conditional probability with respect to $q(r_i^s | \ddot{\theta}_i^s)$ and $q(r_i^d | \ddot{\theta}_i^d)$. See Appendix for the derivations of this term. We assume that the variational posteriors are from the same family of distributions as the respective prior.

We then optimize the bound with respect to q parameters: $\ddot{\theta}_i^d$ and $\ddot{\theta}_i^s$. These optimizations, which we carry out in an alternating fashion, can be viewed as the E and M steps of a variational expectation maximization (EM) algorithm (Barber, 2012). In each E-step, for fixed parameters ($\ddot{\theta}_i^s$, $\ddot{\theta}_i^d$) we find the distribution $q(M)$ which maximizes Eq 4 and in each M-step, we find $\ddot{\theta}^d$ and $\ddot{\theta}^s$ that maximize Eq 4 while $q(M)$ is fixed. Additionally, we also update the prior parameters at the end of each M-step. We did not choose to equip the prior parameters with a hyperprior, because we expect them to be well determined by the data ($N > 1000$ trials per participant). Figure 2 shows a flow chart of the optimization steps. We implemented the model with Pytorch (1.10.1) in Python (3.9.7) using the Adam optimizer. For more information about learning rates and iteration steps, see the code at: http://dx.doi.org/10.17192/fdr/88.
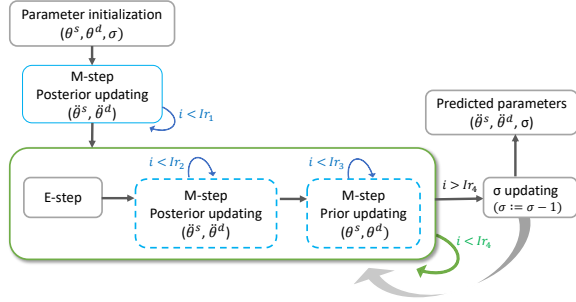
Figure 2: Flowchart of the variational E-M algorithm used for model optimization. θ^s, θ^d and σ are the parameters of SM component, decision component and the standard deviation of the noise, respectively (see Eq 4). Each blue box is an optimization step and blue arrows indicate that the parts are iterated to converge – with negative ELBO as the loss functions. *Ir* is iteration rate and its value differs among optimization parts ($Ir_1 \cdots Ir_4$). The whole green box is also iterated to converge. Additionally, after each σ updating (to seek the smallest possible standard deviation for the noise function) the whole green box is iterated again. $\ddot{\theta}^s$, $\ddot{\theta}^d$ and σ are predicted posterior parameters for SM component, decision component and standard deviation of the noise function, respectively. In the first M-step (the one out of the green box) we assume that all data points belong to $M = 1$ (the response model), to obtain initial parameter estimates, since we expect a only a small fraction of outliers.

Database

We optimized the model on a RT database of a visual attention experiment from (Feldmann-Wüstefeld, Uengoer, & Schubö, 2015) (the first experiment out of four) which includes two different types of intermixed tasks. Participants were presented with eight elements on an imaginary circle around the fixation point. In one task, they had to responded to either the shape or the colour singleton based on their group membership. In the other task, both groups of participants responded to the orientation of a line embedded in the shape singleton while they had to ignore a color distractor in half of the trials. The target of the experiment was investigating the role of selection history (Awh, Belopolsky, & Theeuwes, 2012) on selective visual attention. The participants (11 males and 17 females) were in 18-32 age range and all but two were right handed. Each participant responded to 1024 trials. For more information about the experiment, see the main reference (Feldmann-Wüstefeld et al., 2015). Meibodi et al. (2021b) proposed a model for these data which assumes an ex-Gaussian RT distribution, we aim to replace this model assumption by a more theoretically motivated one.

Results and discussion

As mentioned in previous section, we tested three versions of the model with different X distributions (Gaussian, gamma or Laplace) (see Eq 1 and Figure 1). We selected the prior

on these distributions' parameters in a way that matches the reported means and standard deviations in other SRT studies(see Table 1). We chose the prior on inv-Gaussian parameters based on a preliminary analysis of the data which assumed that the SM component is a constant. More specifically, priors parameter values are

$$
\begin{aligned}
r^d &\sim IG(\mu = 500.0, \lambda = 10000.0) \\
\eta &\sim \mathcal{N}(0.0, \sigma = 12.0) \\
r^s &\sim Laplace(m = 200.0, \ std = 9.8) \ or \\
&\quad Gaussian(m = 200.0, \ std = 10.0) \ or \\
&\quad Gamma(m = 200.0, \ std = 10.0).
\end{aligned}
\tag{5}
$$

The model was then fitted to the data of each of the 28 participants. The number of participants is within the range typically used in mixed model repeated measurement designs (Feldmann-Wüstefeld et al., 2015). For each participant the final free energy is computed (the results are plotted in Figure 3a for all models) and the sum over all participants is used for model comparison (see Table 2). Smaller free energy values indicate better fits (higher ELBO). Thus, the best model is the one with Gaussian SM distribution. The results of the Laplace model are very close to the Gaussian. The mean of outliers over all participants is also shown in Table 2. The models label a very similar proportion of trials as outliers, independent of the choice of SM distribution, as can be seen in Figure 3b. This closeness might be due to having similar decision components in all versions of the model. The decision component (inv-Gaussian) has a much bigger variance than the narrow SM components– and it is therefore driving the outlier determination. As explained in the methods section, outliers are RTs which can not be separated to decision and SM components by the model. So they are sampled from a uniform distribution (M=0). Thus, an outlier is either a very fast or a very slow response. For two example participants, Figure 4 illustrates that which part of the data is considered as an outlier by our model. The criterion in these plots is

Table 2: Model comparison results for different SM distributions. 'FE' is free energy (sum over all participants for each model), smaller values indicate a better RT database fit. 'Outliers' shows the mean fraction of outliers over all participants. For each participant the amount of outliers is the sum over the outlier posterior distribution ($q(M = 0)$). 'θ^s' includes the updated prior parameters (mean and standard deviation) of the SM distribution which are optimized by the model. The reported values are the grand means of the means and standard deviations over all participants.

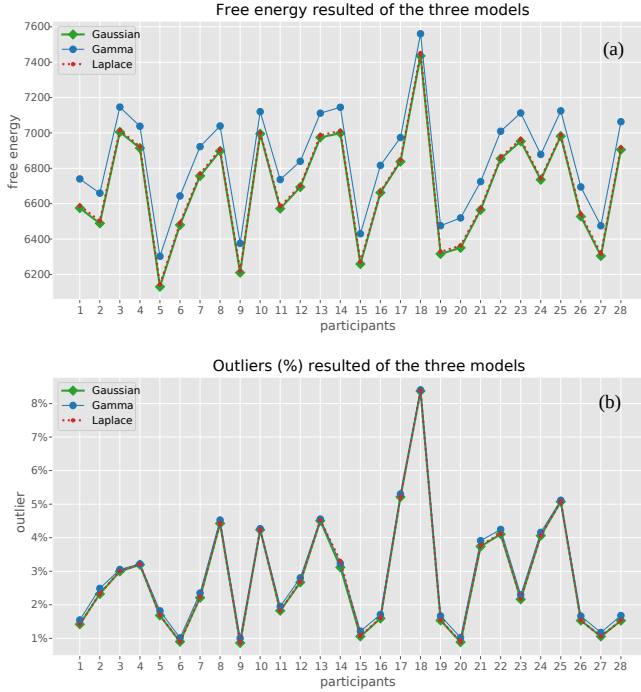| SM distribution	FE	Outliers	θ^s (mean, std)
Gaussian	187387.66	1.52%	199.58, 0.37
Laplace	187685.74	1.53%	199.59, 0.40
Gamma	191684.43	1.68%	199.50, 0.52

Figure 3: Results per participants. For 28 participants see the free energy values (a) and the percentage of outliers (b) resulting from the optimization of three different SM distribution models.
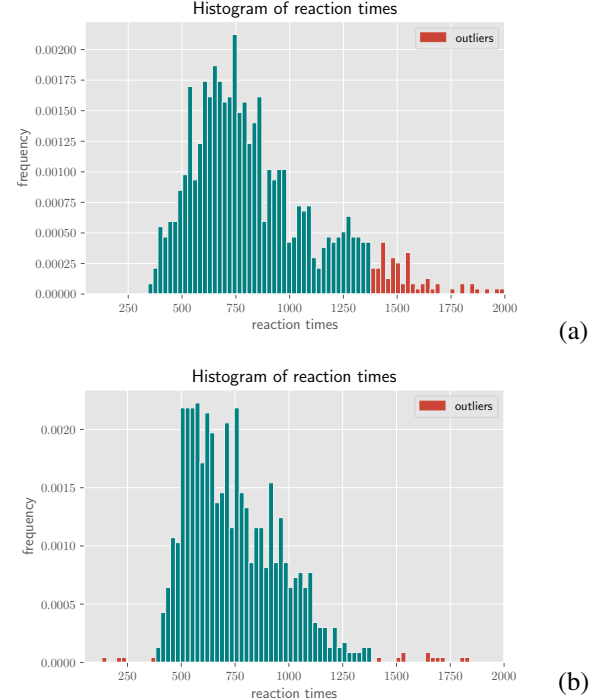

Figure 4: Outliers. Histograms of RTs with predicted outliers (marked in red) by the model for the 18th and the 3rd participant. These participants have relatively: the highest numbers of outliers (a) and an average numbers of outliers –with very short and long RTs (b).

$q_i(M = 0) > 0.6$ which means r_i is labelled as an outlier if it belongs to $M = 0$ with a probability of more than 0.6.

For each trial, the model predicts posterior parameters on both decision and SM components through minimizing free energy and afterwards priors are updated (these steps are shown in Figure 2). The final prior parameters of SM component (θ^s) for each version of the model can also be seen in Table 2. These predicted parameters are close in mean and all are very narrow distributions (see Figure 5).

Subtracting both the expected SM component (r_i^s) and the noise component (η_i) from the RTs ($r_i^d = r_i - r_i^s - \eta_i$), we obtain the expected decision component. The inverse Gaussian distribution fits better on this component rather than on the total RT data, indicating that the Wiener process assumption might be justified. However, this assumption should be investigated more in future studies. For an illustration, see Figure 6: here, the inverse-Gaussian fits to the total RT data (orange histograms) are worse than the fits to the expected decision components only (in green) for two participants (6th and 18th). These participants have the smallest and the largest numbers of outliers (see and compare their outliers in Figure 3 b). The plots (Figure 6) show that the model works well in either case. In addition, the best distribution for the measurement noise is $\mathcal{N}(0.0, 2.0)$ which is obtained by updating σ at the end of the optimization iterations as shown in Figure 2. Finally, for each participant is possible to reverse the process and reconstruct the RTs from the posteriors. In this case, the mean of reconstruction error for each participant is

less than 8.5 ms.

Conclusion

The role and importance of different RT components in shaping the total RT distribution has long been a matter of question in cognitive modelling. Quantitative models, such as the ones proposed in this paper, can be helpful in comparing the predictions of different theoretical accounts of RTs objectively and disentangling the components. Moreover, different lines of research might be interested in different components such as the effect of brain disorders on decision making (Herz, Bogacz, & Brown, 2016) versus motor responses (Low, Miller, & Vierck, 2002).

The purpose of the modelling reported in this paper was to investigate if machine learning methods can help to disentangle a RT distribution into two main components of decision time and sensorimotor time. The motivation for our research was a previous study by Meibodi et al. (2021b) which presented an algorithmic model of selection history effects without a solid theoretical foundation for the chosen RT distribution. We are now in a position to remedy this issue. We expect that our proposed model will be useful whenever RT components need to be extracted in cognitive RT modeling.

The results showed that the final predicted SM distributions are very narrow which is comparable with the assumption in Ratcliff diffusion model: non-decision component might be
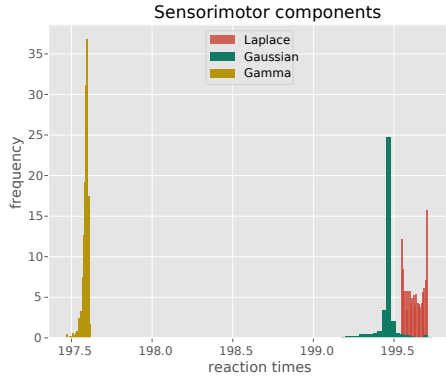
Figure 5: Posterior sensorimotor (SM) components in three versions of the model for a random participant. In these models, it was assumed that SM distributions might be Gaussian, gamma or Laplace. The results show very narrow distributions under all assumptions. Gaussian fits best, see also table 2.

sampled from any distribution and the shape of it can not influence the final RT distribution as the decision part has a very large standard deviation and the other one has a small one (Ratcliff & Childers, 2015). The predicted mean of this distribution is in range of 199.58 ± 0.37 by the best fitted model. The best fitted model is the version which assumes the SM component is Gaussian. The predicted mean has a close range to some simple reaction time experiments results (see Table 1).

The model can successfully label an acceptable number of extreme-valued RTs as outliers. Importantly, in our approach this labelling is driven by the model assumptions and the labels will therefore be internally consistent with the model's predictions, unlike more traditional methods for outlier labels based e.g. on standard deviation measurements. This property might be useful for the principled detection of inattentive participants, e.g. in ADHD or autism studies, where a larger proportion of outliers is to be expected.

Appendix

$$
\begin{aligned}
&\langle \log p(r_i|r_i^s,r_i^d,\sigma)\rangle_{q(r_i^s|\breve{\theta}_i^s)\,q(r_i^d|\breve{\theta}_i^d)} = \\
&-\frac{1}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\big\langle (r_i-(r_i^d+r_i^s))^2\big\rangle_q = \\
&-\frac{1}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\bigg[r_i^2 - 2r_i\langle r_i^d\rangle - 2r_i\langle r_i^s\rangle \\
&+ 2\langle r_i^s\rangle\langle r_i^d\rangle + \langle r_i^{d2}\rangle + \langle r_i^{s2}\rangle \bigg]
\end{aligned}
\tag{6}
$$

Using the definition of variance $\langle r_i^{d2}\rangle = \langle r_i^d\rangle^2 + var(r_i^d)$, the
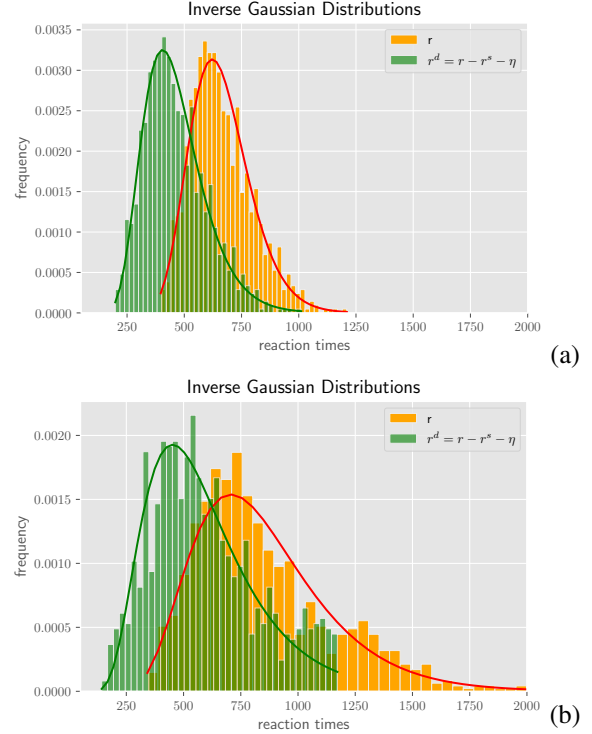

(a)

(b)

Figure 6: Inverse Gaussian distributions for the 6th participant (a) and the 18th participant (b). The orange histograms contain the total RTs (r), red curves fitted by maximizing log-likelihood. The green histograms show the expected decision components (r^d) after subtracting the SM components (r^s) and noises (η) and also discarding the outliers ($q(M = 0) > 0.6$). Parameters of the green densities are updated priors which are predicted by the model for each participant.

above equation can be rearranged

$$
\begin{aligned}
&\langle \log p(r_i|r_i^s,r_i^d,\sigma)\rangle_{q(r_i^s|\breve{\theta}_i^s)\,q(r_i^d|\breve{\theta}_i^d)} = \\
&-\frac{1}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\bigg[(r_i - \langle r_i^d\rangle - \langle r_i^s\rangle)^2 \\
&+ var(r_i^d) + var(r_i^s) \bigg]
\end{aligned}
\tag{7}
$$

References

Amini Vishteh, R., Mirzajani, A., Jafarzadehpour, E., & Darvishpour, S. (2019). Evaluation of simple visual reaction time of different colored light stimuli in visually

normal students. *Clinical Optometry*, *11*, 167–171. doi: 10.2147/OPTO.S236328

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: a failed theoretical dichotomy. *Trends in Cognitive Sciences*, *16*(8), 437–443. doi: 10.1016/j.tics.2012.06.010

Barber, D. (2012). *Bayesian Reasoning and Machine Learning*. Cambridge: Cambridge University Press. doi: 10.1017/cbo9780511804779

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Christie, L., & Luce, R. D. (1956). Decision structure and time relations in simple choice behavior. *The bulletin of mathematical biophysics*, *18*, 89–112.

Dodonova, Y. A., & Dodonov, Y. S. (2013). Is there any evidence of historical slowing of reaction time? No, unless we compare apples and oranges. *Intelligence*, *41*(5), 674–687. doi: 10.1016/j.intell.2013.09.001

Dykiert, D., Der, G., Starr, J. M., & Deary, I. J. (2012). Sex differences in reaction time mean and intraindividual variability across the life span. *Developmental Psychology*, *48*(5), 1262–1276. doi: 10.1037/a0027550

Evans, N. J., & Wagenmakers, E.-J. (2020). Evidence Accumulation Models: Current Limitations and Future Directions. *The Quantitative Methods for Psychology*, *16*(2), 73–90. doi: 10.20982/tqmp.16.2.p073

Feldmann-Wüstefeld, T., Uengoer, M., & Schubö, A. (2015). You see what you have learned. Evidence for an interrelation of associative learning and visual selective attention. *Psychophysiology*, *52*(11), 1483–1497. doi: 10.1111/psyp.12514

Fitousi, D. (2020). Linking the Ex-Gaussian Parameters to Cognitive Stages: Insights from the Linear Ballistic Accumulator (LBA) Model. *The Quantitative Methods for Psychology*, *16*(2), 91–106. doi: 10.20982/tqmp.16.2.p091

Folks, J. L., & Chhikara, R. S. (1978). The Inverse Gaussian Distribution and Its Statistical Application-A Review. *Journal of the Royal Statistical Society*, *40*(3), 263–289. doi: 10.1111/j.2517-6161.1978.tb01039.x

Friston, K. (2003). Learning and inference in the brain. *Neural Networks*, *16*(9), 1325–1352. doi: 10.1016/j.neunet.2003.06.005

Friston, K., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology Paris*, *100*(1-3), 70–87. doi: 10.1016/j.jphysparis.2006.10.001

Hawkins, G. E., Forstmann, B. U., Wagenmakers, E. J., Ratcliff, R., & Brown, S. D. (2015). Revisiting the evidence for collapsing boundaries and urgency signals in perceptual decision-making. *Journal of Neuroscience*, *35*(6), 2476–2484. doi: 10.1523/JNEUROSCI.2410-14.2015

Herz, D. M., Bogacz, R., & Brown, P. (2016). Neuroscience: Impaired Decision-Making in Parkinson's Disease. *Current Biology*, *26*(14), R671–R673. doi: 10.1016/j.cub.2016.05.075

Hohle, R. H. (1965). Inferred components of reaction times as functions of foreperiod duration. *Journal of Experimental Psychology*, *69*(4), 382–386. doi: 10.1037/h0021740

Horowitz, T. S., & Wolfe, J. M. (2003). Memory for rejected distractors in visual search? *Visual Cognition*, *10*(3), 257–298. doi: 10.1080/13506280143000005

Hwang-Gu, S. L., Chen, Y. C., Liang, S. H. Y., Ni, H. C., Lin, H. Y., Lin, C. F., & Gau, S. S. F. (2019). Exploring the variability in reaction times of preschoolers at risk of attention-deficit/hyperactivity disorder: An ex-Gaussian analysis. *Journal of Abnormal Child Psychology*, *47*(8), 1315–1326. doi: 10.1007/s10802-018-00508-z

Jain, A., Bansal, R., Kumar, A., & Singh, K. (2015). A comparative study of visual and auditory reaction times on the basis of gender and physical activity levels of medical first year students. *International Journal of Applied and Basic Medical Research*, *5*(2), 124. doi: 10.4103/2229-516x.157168

Low, K. A., Miller, J., & Vierck, E. (2002). Response slowing in Parkinson's disease: A psychophysiological analysis of premotor and motor processes. *Brain*, *125*(9), 1980–1994. doi: 10.1093/brain/awf206

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. Oxford University Press.

Maddox, W. T., Ashby, F. G., & Gottlob, L. R. (1998). Response time distributions in multidimensional perceptual categorization. *Perception and Psychophysics*, *60*(4), 620–637. doi: 10.3758/BF03206050

Meibodi, N., Abbasi, H., Schubö, A., & Endres, D. (2021a). Distracted by previous reward: integrating selection history, current task demands and saliency in a computational model. *PsyArXiv*. doi: 10.31234/OSF.IO/MBE5A

Meibodi, N., Abbasi, H., Schubö, A., & Endres, D. (2021b). A model of selection history in visual attention. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*. Retrieved from https://escholarship.org/uc/item/3m33h9h7

Ostojic, S. (2011). Interspike interval distributions of spiking neurons driven by fluctuating inputs. *Journal of Neurophysiology*, *106*(1), 361–373. doi: 10.1152/jn.00830.2010

Palmer, E. M., Horowitz, T. S., Torralba, A., & Wolfe, J. M. (2011). What are the shapes of response time distributions in visual search? *Journal of Experimental Psychology: Human Perception and Performance*, *37*(1), 58–71. doi: 10.1037/a0020747

Pedersen, M. L., Frank, M. J., & Biele, G. (2017). The drift diffusion model as the choice rule in reinforcement learning. *Psychonomic Bulletin and Review*, *24*(4), 1234–1251. doi: 10.3758/s13423-016-1199-y

Ratcliff, R. (1978). A theory of memory retreival. *Psychological Review*, *85*(2), 59–108. doi: 10.1037/h0021465

Ratcliff, R. (2006). Modeling response signal and response time data. *Cognitive Psychology*, *53*(3), 195–237. doi: 10.1016/j.cogpsych.2005.10.002

Ratcliff, R., & Childers, R. (2015). Individual Differences and Fitting Methods for the Two-Choice Diffusion Model

of Decision Making. *Decision*, *2*(4), 237–279.

Ratcliff, R., & McKoon, G. (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Computation*, *20*(4), 873–922. doi: 10.1016/j.biotechadv.2011.08.021.Secreted

Ratcliff, R., & Smith, P. L. (2004). A Comparison of Sequential Sampling Models for Two-Choice Reaction Time. *Psychological Review*, *111*(2), 333–367. doi: 10.1037/0033-295X.111.2.333

Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology and Aging*, *16*(2), 323–341. doi: 10.1037/0882-7974.16.2.323

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin and Review*, *9*(3), 438–481. doi: 10.3758/BF03196302

Schwarz, W. (2001). The ex-Wald distribution as a descriptive model of response times. *Behavior Research Methods, Instruments, and Computers*, *33*(4), 457–469. doi: 10.3758/BF03195403

Ulrich, R., & Stapf, K. H. (1984). A double-response paradigm to study stimulus intensity effects upon the motor system in simple reaction time experiments. *Perception & Psychophysics*, *36*(6), 545–558. doi: 10.3758/BF03207515

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical bayesian estimation of the drift-diffusion model in Python. *Frontiers in Neuroinformatics*, *7*. doi: 10.3389/fninf.2013.00014

Woods, D. L., Wyma, J. M., Yund, E. W., Herron, T. J., & Reed, B. (2015). Factors influencing the latency of simple reaction time. *Frontiers in Human Neuroscience*, *9*. doi: 10.3389/fnhum.2015.00131