Time is of the essence: Modelling explicit time estimation in instrumental learning

Stevenson, Niek

Researchmaster thesis
(Psychology)
Niek Stevenson
Studentnummer 1924486

Time is of the essence

Modelling explicit time estimation in instrumental learning

Niek Stevenson[1,2], Steven Miletić[2], Birte Forstmann[1,2]

[1]Leiden University, Department of Psychology, Leiden

[2]University of Amsterdam, Department of Psychology, Amsterdam

Abstract

Recent work has shown that we can achieve a better understanding of learning behavior by integrating reinforcement learning models with evidence accumulation models (RL-EAM). RL-EAM predict that as people learn they respond faster and more accurately. However, two recent experiments showed that when learning under speed pressure, people demonstrated a learning-related increase in accuracy, but not in response speed. We hypothesized that this might be caused by a proportion of responses resulting from a timing accumulation process that keeps track of time in parallel to the evidence accumulation process during a decision. We compared RL-EAM with and without the addition of time estimation on data from two independent experiments. We found no compelling evidence that the proposed mechanism of time estimation aid in decision-making in learning.

Introduction

Everyday life poses us with many small decisions in which we benefit from previous experiences. When you play chess for the first time, you will have no idea of what the right opening moves are. However, with more playing, you will learn the consequences of your moves and become better. This example illustrates that learning processes influence decisions, and the outcomes of these decisions in turn drive feedback-driven learning (Bogacz & Larsen, 2011; Miletić et al., 2021).

Although cognitive science has a rich history in studying both feedback-driven learning, and decision-making, these fields have so far remained largely separate. Error-driven learning behavior is often studied using probabilistic selection tasks, in which participants have to choose between multiple options that are each associated with a different probability of returning reward. In these tasks the aim is to maximize returns by learning to choose the most

rewarding option. Such choice behavior can be well understood using reinforcement learning (RL) models that posit that for each choice people maintain value representations that are updated based on feedback (Dayan & Daw, 2008; Sutton & Barto, 2018).

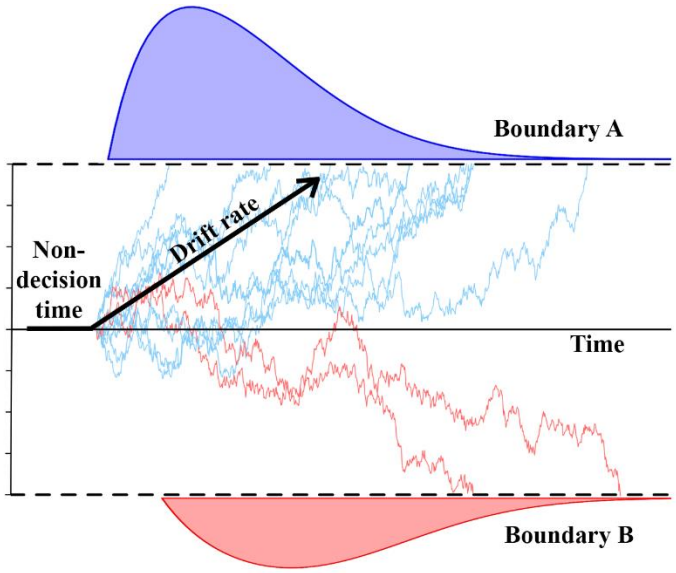
Decision-making processes are often studied using evidence accumulation models (EAMs). These models postulate that choices are made by gathering information for each choice alternative until sufficient evidence for one alternative has been accumulated to commit to a decision (Brown & Heathcote, 2008; Forstmann et al., 2016; Ratcliff & Smith, 2004). Although many different EAMs exist, most propose that decision-making is a concurrence of at least three latent cognitive mechanisms. First, the drift rate determines the speed of the evidence accumulation process. The drift rate depends jointly on the physical properties of the stimulus (Bode et al., 2018; Ratcliff & Smith, 2004) and on cognitive processes such as attention and information processing ability (Nunez et al., 2017; Smith & Ratcliff, 2009). Second, the threshold captures the amount of evidence needed to commit to a decision and represents response caution. Third, the non-decision time consists of the duration of sensory encoding of the stimulus and the motor processing of the necessary movements to execute the response.

Recent advances have furthered our understanding of the reciprocal influences of learning and decision-making, by integrating RL models and EAMs (RL-EAMs; Fontanesi, Gluth, et al., 2019; Fontanesi, Palminteri, et al., 2019; Frank et al., 2015; Miletić et al., 2021; Miletić, Boag, & Forstmann, 2020; Pedersen et al., 2017; Sewell et al., 2019). RL-EAMs propose that people make decisions by gradually integrating information of value representations associated with each available choice option. When enough evidence has been accumulated, we commit to a choice and the associated feedback is used to update the value representations. In turn these value representations adjust the speed of information integration the next time the decision-maker is faced with the same choice.

Most studies that investigated the influence of learning on decision-making have combined RL update rules with the most popular EAM, the diffusion decision model (DDM; Ratcliff, 1978; Ratcliff et al., 2016). The RL-DDM proposes one single accumulation process that is driven by the difference in value representations between two choice alternatives. There are two boundaries to the accumulation process, one for each choice (Figure 1).

Figure 1

The evidence accumulation process for the diffusion decision model.

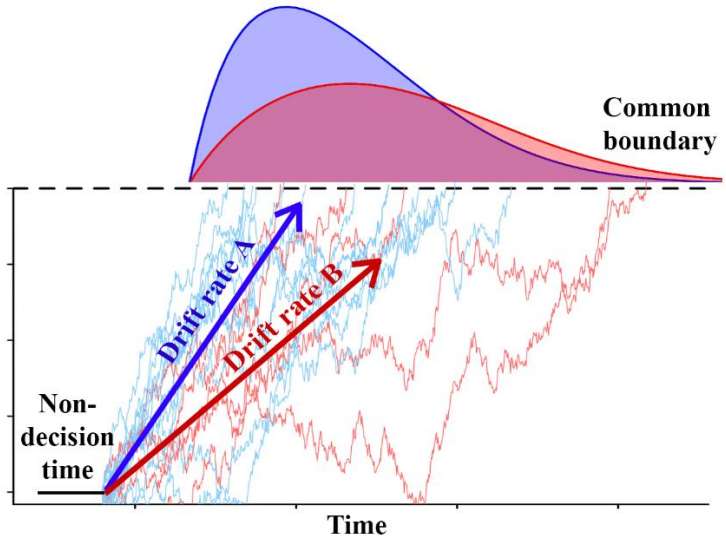


One accumulator accrues noisy evidence until either threshold is reached. The first threshold reached by the accumulator determines which choice is made. The response time is determined jointly by the time taken to reach the threshold and the non-decision time.

An alternative to the DDM is the advantage racing diffusion (ARD) framework that proposes a more neurally plausibly race between separate accumulators towards a common threshold (Figure 2; Forstmann et al., 2016; Miletić, Boag, Trutti, et al., 2021; Ratcliff et al., 2007; van Ravenzwaaij et al., 2019; Zandbelt et al., 2014). The first choice for which a threshold-level of evidence is accumulated wins the race and the corresponding response is made. Separate drift rates for each accumulator are jointly determined by a constant term, an advantage term that constitutes the difference between value representations of the choice alternatives, and a magnitude term that constitutes the sum of the value representations (van Ravenzwaaij et al., 2019). Compared to the RL-DDM, the RL-ARD has been found to better fit response time distributions and choice behavior for different instrumental learning tasks (Miletić et al., 2021).

Figure 2

The evidence accumulation process for the racing diffusion model.



Two separate noisy accumulators that are driven by separate drift rates race towards a common boundary. The first accumulator to reach the threshold determines the choice made. The response time is determined jointly by the time taken to reach the threshold and the non-decision time.

Recent studies have also focused on how timing and learning interact to influence decision-making. Some situations require more swift action than others and we flexibly balance the need for speed and accuracy. If I play a timed chess game, how much time should I spend on the next move? Such a problem is an example of a speed – accuracy trade-off (SAT). Two recent studies applied RL-EAMs to a learning task with a speed – accuracy manipulation to better understand how learning influences the balance between speed and accuracy (Miletić et al., 2021; Sewell & Stallman, 2020).

RL-EAMs predict that as people learn, the difference in value representations between the correct and incorrect choice increases. This difference in value representations does not only yield more accurate, but also faster responses as it drives the speed of evidence accumulation (Miletić, Boag, & Forstmann, 2020). However, observations in our own data (Miletić et al., 2021) and in the data of Sewell and Stallman (2020) suggest that whereas accuracy increases with learning under speed pressure, response times are relatively unaffected. Thus, accuracy increases with learning independently of response speed. We will refer to this observation as the speed – accuracy decoupling.

This speed – accuracy decoupling is incompatible with standard RL-EAMs since drift rates are thought to increase with learning, which simultaneously increases response speed and accuracy. This suggests that speed pressure affects decision-making in a way that cannot be explained with traditional accounts of decision-making in learning.

The influence of speed pressure on decision-making has been studied for decades (McElree & Dosher, 1989; Reed, 1973; Wickelgren, 1977), and one of the hallmark advantages of EAMs is that they provide a mechanistic understanding of the cognitive processes underlying the SAT (Bogacz et al., 2010; Kelly et al., 2020; Rae et al., 2014; Ratcliff & Rouder, 1998; Usher et al., 2002).

Earlier studies suggested that people balance speed and accuracy by adjusting the amount of evidence needed to commit to a decision (Ratcliff & Rouder, 1998; Usher & McClelland, 2001). This birthed the selective influence account of the SAT, referring to the idea that speed pressure selectively lowers response caution, while the non-decision time and drift rate are unaffected. This account has long been used as a benchmark to test a new model's validity (Rae et al., 2014). However, selective influence was later questioned as more and more studies indicated that people performing under increased speed-pressure not only lower their response thresholds, but also increase drift rates (Bogacz et al., 2010; Rae et al., 2014).

The effect of SAT manipulations on drift rates could indicate various psychological effects. For one, it could suggest that speed pressure increases attention to the task, since people are aware they have limited time to inform their decisions. An alternative explanation could lie in the presence of urgency. Contrary to the traditional explanation of the SAT in terms of overall changes in a static threshold, urgency refers to *within-trial* adjustments of the thresholds (by means of *collapse* over time) or drift rates (by means of evidence-independent, but time-dependent increases; Cisek et al., 2009; Ditterich, 2006; Thura et al., 2012).

Urgency is thought to reflect the decision-makers urge to respond with passing time. Recent work suggests that within race models, such as the ARD, urgency can express itself as increases in drift rates, as long as these increases are shared across all accumulators (Miletić & van Maanen, 2019). In some models, such increases in drift rates are even equivalent to linearly collapsing bounds (Miletić & van Maanen, 2019). As such, the observed increases in drift rate as a result of SAT manipulations may indicate the presence of urgency-like strategies to ensure a decision is made before the deadline.

Nevertheless, the evidence for urgency in conventional decision-making tasks has been mixed (Boehm et al., 2016; Forstmann et al., 2016; Hawkins et al., 2015), owing partly to technical difficulties in estimating models that include urgency mechanisms (Evans et al., 2019; Voskuilen et al., 2016). Whether participants employ urgency strategies appears to depend on the specifics of the experimental design and the amount of training participants had (Evans & Hawkins, 2019).

Additionally, urgency mechanisms such as collapsing bounds and drift rate modulations have recently been criticized for their implicit dependence on time estimation ability (Hawkins & Heathcote, 2021). Research on time estimation suggests that how participants keep track of time is similar to how participants keep track of evidence in a decision-making task, namely via an accumulation-to-bounds process (Balci & Simen, 2014, 2016; Simen et al., 2016). Based on these considerations, Hawkins and Heathcote (2021) proposed the timed racing diffusion model (TRDM) which posits that a timing accumulation process occurs in parallel to the evidence accumulation process. If the timing accumulator reaches its threshold before the evidence accumulators, it prematurely ends evidence accumulation, leading to a hard time limit on the decision making process. Thus, the timing accumulation process speculates that people rely on an internal timer to respond adaptively when facing response deadlines.

We explored to what extent explicit internal time estimation can explain the SAT mechanisms of decision-making in learning, by augmenting the learning model as proposed in Miletić et al. (2021) with a timing accumulator as proposed by Hawkins and Heathcote (2021). The resulting model (RL-tARD) postulates a race between three accumulators: A timing accumulator, and two evidence accumulators collecting evidence for each choice option. The speed of the evidence accumulation processes is driven by the subjective value representations, which are updated on each trial as determined by an RL delta updating rule.

In their TRDM, Hawkins and Heathcote (2021) postulated that if the timer process ends evidence accumulation prematurely, a guess is made between the available response options. However, they did not explore the option that timer-generated responses favor the choice that has accumulated the most evidence so far, referred to as relying on partial information (Ratcliff, 1980, 1988, 2006). Although they show that the quality of fit for the TRDM was not affected by different guessing rules, we explored the possibility of reliance on partial information, since recent evidence in perceptual decision-making suggests decision-makers have access to partial information (McLean et al., 2020).

In the RL-tARD, the response speed is determined by the timing accumulation process, whereas the accuracy is either based on a guess as in the TRDM as proposed by Hawkins and Heathcote (2021), or on partial information from the evidence accumulators in our implementation. Compared to the standard RL-ARD, the response speed and accuracy in the RL-tARDs are to a greater degree decoupled since they can be products of different processes. We therefore predict that the RL-tARD can better explain the observed speed – accuracy decoupling. Additionally, we tested whether timer-generated responses rely on partial information (informed RL-tARD) or as previously proposed on a guess (random RL-tARD) between the available options.

Methods

We analyzed data from the instrumental learning task of Miletić et al. (2021) (their experiment 2) and from a similar instrumental learning task of Sewell and Stallman (2020).

Miletić et al., 2021

Participants

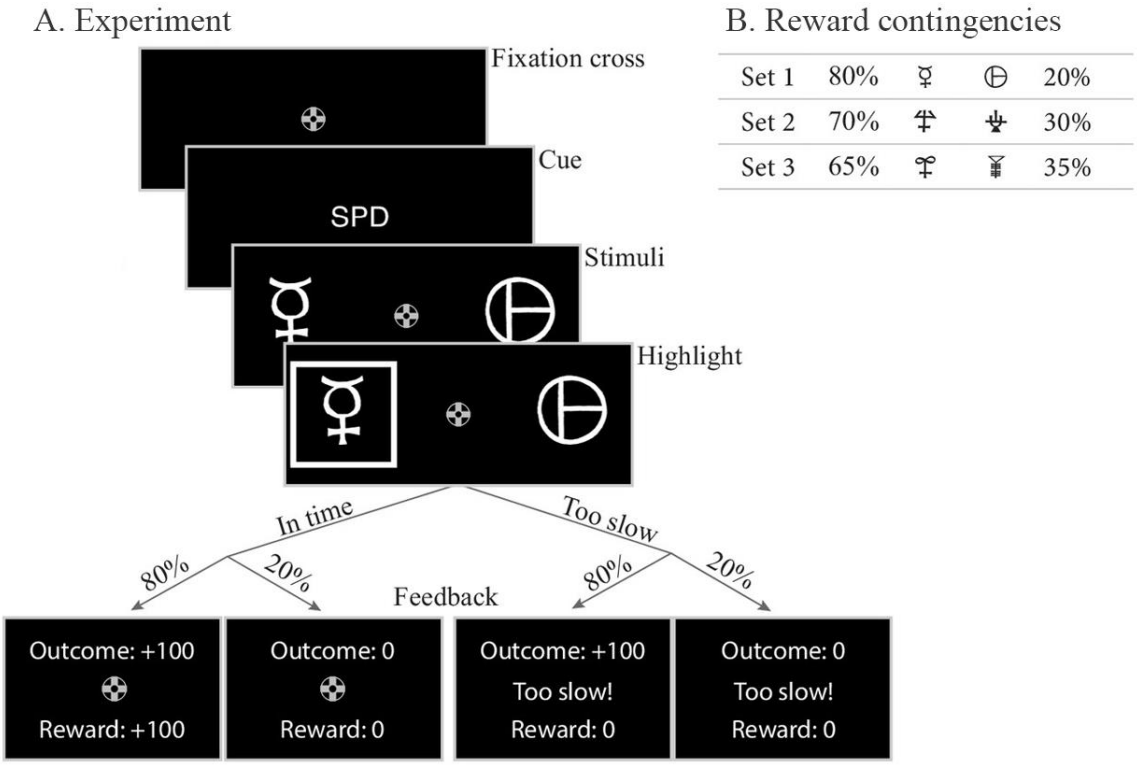
23 students from the subject pool of the department of psychology of the University of Amsterdam participated for course credits (16 female, 23 right-handed, age: M = 19 years, SD = 1.06 years). All participants had normal or corrected-to-normal vision and gave written informed consent prior to the experiment. The study was approved by the local ethics committee.

Task

Participants performed a probabilistic instrumental learning task (Frank et al., 2004). On each trial participants had to choose between two stimuli that were both associated with a fixed probability of returning reward (Figure 3). One choice alternative always had a higher probability of returning reward than the other. The goal of the task was to maximize rewards by learning through trial and error which choice alternative was most likely to return reward.

Figure 3

Example trial of the Miletić et al. (2021) experiment.



(A) On each trial participants were presented with a cue that indicated how much time they had to respond, followed by the presentation of the stimuli, a highlight of the chosen option, and the probabilistic feedback. On trials where the participants received the speed cue, they would get no reward if they were too slow to respond. They did get feedback in terms of the outcome if they had responded in time. (B) Reward contingencies for the different stimulus sets. The percentages represent the probability of receiving +100 reward if that option was chosen. The symbols used differed between participants. Adapted from Miletić et al., (2021).

Following a practice block, participants completed 324 trials divided over 3 blocks. On each trial, one of three different pairs of abstract symbols was presented. Within one block each pair was presented 36 times. Within each pair, one stimulus would be presented on the right side of a fixation cross just as often as on the left. Stimulus pairs differed in their associated reward probabilities: 0.8/0.2, 0.7/0.3, and 0.6/0.4. These reward probabilities were chosen such that the difference in reward probability between the two stimuli differed to vary difficulty, but the mean reward probability of both options combined was the same. Received reward was presented as +100 points or +0 points. Participants were instructed to earn as many points as possible.

In order to manipulate the speed – accuracy trade-off, a cue and a deadline manipulation were added to the task. Prior to each trial a cue instructed participants to emphasize response speed ('SPD') or accuracy ('ACC'). The stimuli were presented for 2000 ms, however, participants did not earn a reward in speed trials if they responded slower than 600 ms. Speed and accuracy trials were randomly interleaved.

Response feedback consisted of reward and outcome. The outcome corresponded to the probabilistic outcome of the choice, whereas the reward corresponded to the actual points earned. Thus, the reward would always be +0 if the participant was too late. The outcome was included so that participants could still learn from their choices even if they responded too late. Fixation crosses with jittered durations were presented between each part of the trial, since the experiment also served as a pilot for an fMRI study. Fixation crosses were varied with steps of 500 ms. Pre-cue fixations lasted between 500-2000 ms. Post-cue pre-stimulus, post-stimulus pre-highlight and post-highlight pre-feedback fixation cues lasted between 500-1500 ms. Inter-trial fixation cues lasted between 500-2500 ms. Each trial always lasted 7500 ms in total and the experiment took approximately 45 minutes.

Sewell & Stallman, 2020

Participants

Six students from the University of Cleveland participated in eight sessions for $20 per session (5 female, age: M = 21.5, SD = 1.05). All participants gave written informed consent prior to the experiment and the study was approved by the local ethics committee. We only analyzed the first session to keep the experiments as similar in set-up as possible as Miletić et al. (2021),

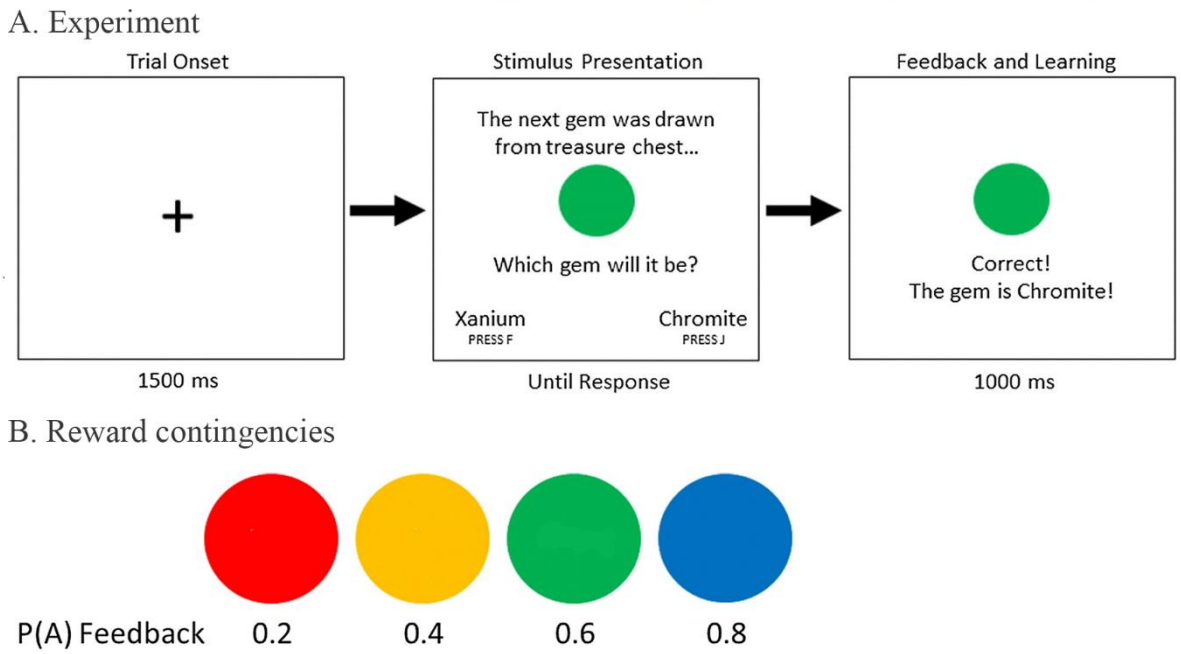
Task

Participants completed four runs of a probabilistic instrumental learning task each session. Each run comprised four stimuli to which the participant could respond by pressing left or right (Figure 4). Within one run, responding left was associated with a different reward probability for each stimulus (0.2/0.4/0.6/0.8). Responding right was associated with a reward probability of 1 – probability of reward for responding left. As in Miletić et al. (2021), the difference in reward probability between the two stimuli differed to vary difficulty, but the

mean reward probability was the same across stimuli. Reward consisted of fictitious gems, and participants were instructed to obtain as many gems as possible.

Figure 4

Example trial of the Sewell and Stallman (2020) experiment.



(A) On each trial participants were presented with a stimulus for which they had to choose left or right. They would obtain gems probabilistically based on their choice. Response window differed between speed and accuracy runs. (B) Reward contingencies for the different stimulus sets. P(A) represents the probability of receiving reward if option A (left) was chosen. The stimulus – reward probability mappings differed between runs and participants. Stimuli alternated between colored circles and math symbols between runs. Adapted from Sewell and Stallman, 2020.

Contrary to the Miletić et al. (2021) study, speed and accuracy trials were not randomly interleaved but manipulated block-wise. Half of the participants performed two speed runs first, followed by two accuracy runs, and the other half of the participants vice versa. Across runs, the four stimuli either consisted of colored circles (red, green, yellow, blue) or math symbols (+, −, ÷, and ×). Therefore, participants had to relearn the response associated with the highest reward probability on every run. The response-stimulus mappings were randomly determined per participant. The two stimulus sets were used once in both the speed emphasis and accuracy emphasis runs, however, not with the same stimulus-reward probability mappings.

Each run comprised four blocks of 40 trials. Stimuli were each presented 10 times per block. A 1500 ms fixation cross was presented prior to the presentation of the stimulus. The stimulus was presented until response and followed by probabilistic feedback for 1000 ms. In speed emphasis runs, additional timing feedback was presented on trials where participants responded slower than 800 ms. Timing feedback was presented as "TOO SLOW" for 3000 ms in order to incentivize participants to meet the response deadlines. Each session lasted approximately 50 minutes.

Cognitive modelling

We compared three variants of the RL-ARD: the standard RL-ARD (Miletić et al., 2021), a RL-ARD augmented with a timing accumulator as described in Hawkins & Heathcote (random RL-tARD; 2021), and a RL-tARD that relies on partial information (informed RL-tARD).

As in Miletić et al. (2021), we used a delta updating rule in all models to describe learning:

$$Q_{i,t+1} = Q_{i,t} + \alpha\,(r_t - Q_{i,t}) \qquad [1]$$

With $Q_{i,t}$ the reward representation of choice option i on trial t, learning rate α and r_t the reward received on trial t. With the current learning rule only the reward representation of the chosen option is updated. Q-values were initialized at 0.

RL-ARD

The RL-ARD assumes that two evidence accumulation processes, one for each choice option, race towards a common threshold a. The first accumulator to reach the threshold determines which choice is made. The time it takes to reach the threshold together with the non-decision time t_0 determine the response time. The RL-ARD assumes that the speed of evidence accumulation for each choice option is driven by three components: first, a constant term V_0; second, the difference in the reward representation of the available choice alternatives weighted by free parameter w_d; third, the sum of the reward representations of both options weighted by free parameter w_s (van Ravenzwaaij et al., 2019). The reward representations are modelled using Eq [1]. Additionally, the evidence accumulation process is subject to Gaussian noise W, with standard deviation s. For the two accumulators that correspond to each choice option this leads to:

$$dx_1 = [V_0 + w_d(Q_1 - Q_2) + w_s(Q_1 + Q_2)]dt + sW$$

$$dx_2 = [V_0 + w_d(Q_1 - Q_2) + w_s(Q_1 + Q_2)]dt + sW \qquad [2]$$

We fixed parameter s to 1 to satisfy scaling constraints (Donkin et al., 2009; van Maanen & Miletić, 2020). Since V_0 is an evidence-independent, but constant additive to the drift rate, we treated it as an urgency signal that varies between the speed and accuracy manipulation (Miletić & van Maanen, 2019). Additionally, we varied the evidence response threshold a between the speed and accuracy condition. These parameters were based on the best performing model from Miletić et al. (2021). In total the RL-ARD has 8 free parameters ($\alpha, V_{0,speed}, V_{0,accuracy}, w_d, w_s, a_{speed}, a_{accuracy}, t_0$).

In the race architecture for two choice options, the probability of response 1, given response time t, can be described as:

$$p_1(t) = PDF_1(t) \cdot [1 - CDF_2(t)] \qquad [3]$$

Where $PDF_1(t)$ is the probability density function of a Wald distribution for the first accumulator (Anders et al., 2019; Tillman et al., 2020). The PDF describes the likelihood of the response time t given the current parameters (irrespective of the second accumulator). $CDF_2(t)$ is the cumulative distribution function of a Wald distribution for the second accumulator, $[1 - CDF_2(t)]$ yields the probability that the second accumulator has not reached the threshold at t. The probability of the second accumulator finishing before the first accumulator, $p_2(t)$, can be described analogously.

Random RL-tARD

The random RL-tARD extends the RL-ARD with a timing accumulator, as proposed by Hawkins & Heathcote (2021). Specifically, the random RL-tARD formalizes a race between two evidence accumulators and one time accumulator. Subscript E refers to the evidence accumulation process and T to the timing accumulation process. The evidence accumulators race towards a common threshold a_E with drift rates given by Eq [2]. The time accumulator races towards an independent threshold a_T with drift rate v_T. For the time accumulator, accumulation noise *st* is fixed to 1 to satisfy scaling constraints. If the timing accumulator reaches its threshold before either of the evidence accumulators reach the evidence threshold, a random guess is made between the two responses. Following Hawkins and Heathcote (2021), we fixed the non-decision time of the time accumulator to 0.05 s.

To model the speed-accuracy trade-off manipulations, we allowed v_T to vary between conditions. Exploratory analyses showed that this was preferred over allowing a_T to vary. In the evidence accumulators, we followed the RL-ARD model described above, which allowed a_E and V_0 to vary between the speed and accuracy condition. However, we inspected posteriors of this model for the data of Miletić et al. (2021) and found overlapping $V_{0,speed}$ and $V_{0,accuracy}$, which indicates that separate V_0 terms were superfluous and suggests that the difference in urgency could instead be captured by the timing accumulator. In total the random RL-tARD has 10 free parameters ($\alpha, V_0,\ w_d,\ w_s,\ a_{E,speed},\ a_{E,accuracy}, t_{0,E}, v_{T,speed}, v_{T,accuracy},\ a_T$).

Given the race between the two evidence accumulators and the timing accumulator, the probability of response 1, given response time t, can be described as (Hawkins & Heathcote, 2021):

$$p_1(t) = PDF_{E,1}(t) \cdot \left[1 - CDF_{E,2}(t)\right] \cdot [1 - CDF_T] + p_{guess} \cdot PDF_T(t) \cdot \prod_{n=1}^{N}[1 - CDF_{E,n}(t)] \quad [4]$$

The rationale of Eq [4] is similar to the logic of Eq [3]. $PDF_{E,1}(t)$ is the probability of responding at time t for evidence accumulator 1, which is made defective by multiplying with the probability of *both* the evidence accumulator 2 *and* the timing accumulator not having finished. To obtain the final probability of response option 1 at time t, the probability of guessing 1 at *t* is added, given by the probability of the time accumulator reaching the threshold at *t* ($PDF_T(t)$), which is made defective by multiplying with the probability that *neither* of the evidence accumulators reached the threshold. The guessing probability p_{guess} was set to 0.5, meaning equiprobable outcomes for both choices. For timer-generated responses, N represents the number of evidence accumulators.

Informed RL-tARD

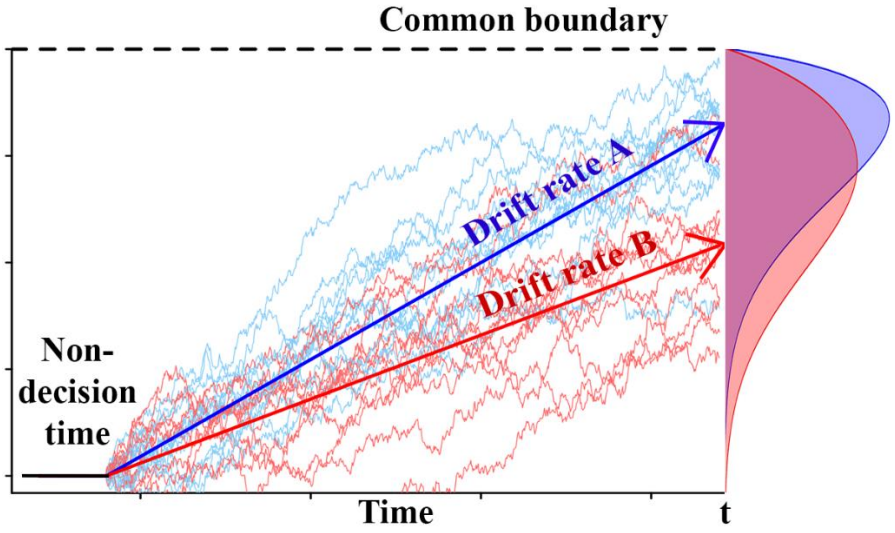
The informed RL-tARD proposes that if the timing accumulator wins the race, a probabilistic choice will be made weighted by the difference in accrued evidence at time t between the two evidence accumulators. If the timing accumulator reaches the threshold first, the choice relies on partial information (Ratcliff, 1980, 1988, 2006). The probability of an unfinished evidence accumulator having accrued x evidence can be described with the probability density function of an unfinished diffusion process with one absorbing boundary, which is derived in Cox and Miller (1965):

$$PDF(x,t) = \frac{1}{s\sqrt{2\pi t}}\left[\exp\left(-\frac{(x-v*t)^2}{2s^2t}\right) - \exp\left(\frac{2b}{s^2} - \frac{(x-2b-vt)^2}{2s^2t}\right)\right] \quad [5]$$

In order to find the probability that evidence accumulator 1 has accrued more evidence than accumulator 2 at time t, we sampled from both probability density functions (Eq [5]) using rejection sampling with a sample of 5000 from a uniform distribution. We then computed the probability of choosing 1 by calculating the proportion of samples from PDF_1 that are larger in x than PDF_2 (Figure 5). To model the speed-accuracy trade-off, we used the same set of parameters as in the random RL-tARD.

Figure 5

Partial information in the racing diffusion model.



Partial information is computed by sampling from the distribution of unfinished responses at time point *t* for both accumulators. The probability of choosing option 1 over option 2 is then calculated by comparing samples from both distributions.

Model estimation and comparison

We estimated group-level and subject-level posterior distributions of model parameters with differential evolution Markov-chain Monte Carlo with Metropolis-Hastings (Ter Braak, 2006), using the R package dynamic models of choice (DMC; Heathcote et al., 2019). We set the number of chains D to three times the number of estimated parameters. Cross-over probability was set at the optimal $2.38/\sqrt{D}$ at the subject-level (Ter Braak, 2006) and at

$U[0,1]$ at the group-level. We used migration only during burn-in and we used a migration probability of 0.05. We considered chains converged when the Gelman-Rubin diagnostic < 1.03, a measure of the relation between the between-chain and within-chain variances (Gelman & Rubin, 1992).

We fit the parameters for the hierarchical models assuming independent truncated Gaussian hyper distributions. Normal prior distributions for all hyper-mean parameters were broad (Table 1). Prior and posterior plots confirmed that these prior settings were not influential. Parameters that varied between conditions were estimated such that the parameter for the speed condition was proportional to the accuracy condition: $\theta_{spd} = (1 + m) * \theta_{acc}$

Table 1

Priors settings for all parameters

	α	$t_{0,E}$	a_E	V_0	w_D	w_s	v_T	a_T	m
Mean	-1.6	0.3	3	0	3	0	2	3	0
SD	5	0.5	5	3	5	3	5	5	5
Limits	$(-\infty, \infty)$	[0.025,1]	$[0, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$	$[0, \infty)$	$[-1, \infty)$

For the learning rate parameter α we transformed the normal prior to a probit scale to enforce limits [0,1].

To select the best model penalized for model complexity, we compared the three models on both data sets using the simplified Bayesian predictive inference criterion (BPIC; Ando, 2011). Lower BPIC values indicate better model fit.

To visually inspect the quality of fit, we simulated data using 100 samples from the posterior parameter distributions. We used these samples to create 2.5%-97.5% credible intervals of the model data that could be compared to the participant's data. For this comparison we summarized RT distributions using the 10th, 50th and 90th percentiles split on both the error and correct trials and the speed and accuracy emphasis trials. The 50th percentile shows the central tendency in the data, the difference between the 10th and 90th shows the variability, and the increased difference between the 90th and the 50th compared to the 50th and 10th percentile shows the positive skew common in RT distributions (Miletić et al., 2021). We visualized learning-related effects by splitting the data into 10 bins, and calculated the accuracy and RT percentiles per bin. Lastly for the RL-tARDs that included a timing

accumulator we calculated the proportion of timer-generated responses in the model-generated credible-intervals per bin.

Results

We compared the three models on both data sets using both the BPIC values and visual inspection of the posterior predictive distribution. We found that the RL-tARDs outperformed the standard RL-ARD for the data of Miletić and colleagues (2021; Table 2). Additionally, the random RL-tARD performed slightly better than the informed RL-tARD. We also examined deviance values to get a measure of absolute quality of fit and found the same order of preference for the three models.

Figure 6 shows that the RL-ARD provides the best fit to the accuracy data. The random RL-tARD underpredicts the learning-related increase in accuracy, and the informed RL-tARD over-estimates the increase in accuracy. Nevertheless, both RL-tARDs better fit the learning related decrease in error response times under accuracy emphasis especially in the early bins. Additionally, although all models capture the response time distributions under speed emphasis well, the RL-ARD over-estimates the decrease in response times towards the later bins, which is better captured by the RL-tARDs.

Table 2

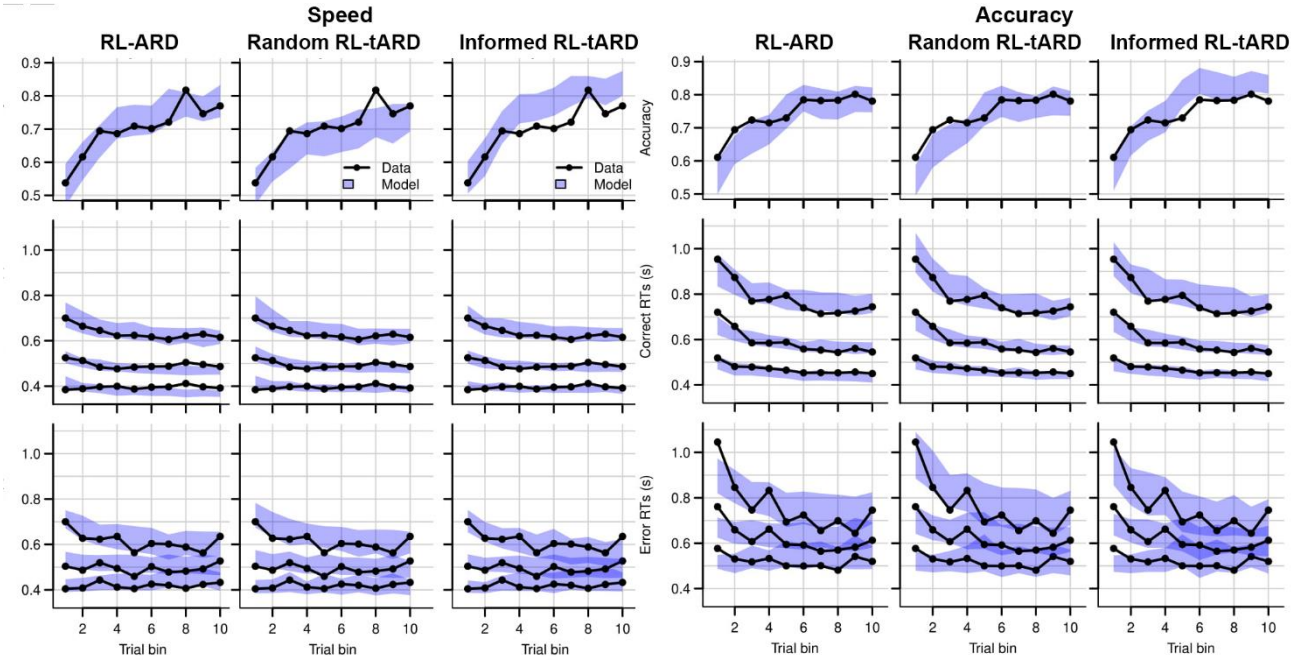
Across-subject mean and SD of the median posterior parameter estimates, and BPIC scores for both data sets

	Miletić et al. 2021								
	α	$t_{0,E}$	a_E (ACC/ SPD)	V_0 (ACC/ SPD)	w_D	w_s	v_T (ACC/ SPD)	a_T	BPIC
RL-ARD	0.12(0.05)	0.14(0.06)	1.83(0.32)/ 1.59(0.40)	2.52(0.50)/ 2.92(0.65)	2.21(0.50)	0.43(0.33)	-	-	-1071
Random RL-tARD	0.10(0.03)	0.23(0.01)	1.66(0.14)/ 1.48(0.05)	1.34(0.26)	3.62(0.49)	0.57(0.26)	2.59(0.30)/ 4.11(0.26)	2.13(0.12)	-1163
Informed RL-tARD	0.10(0.04)	0.32(0.02)	0.93(0.14)/ 0.88(0.10)	-0.30(0.38)	3.14(.39)	1.00(0.51)	3.55(0.26)/ 5.07(0.28)	2.37(0.11)	-1160
	Sewell and Stallman 2020								
	α	$t_{0,E}$	a_E (ACC/ SPD)	V_0 (ACC/ SPD)	w_D	w_s	v_T (ACC/ SPD)	a_T	BPIC
RL-ARD	0.10(0.03)	0.12(0.02)	1.66(0.08)/ 1.44(0.05)	1.29(0.11)/ 2.79(0.20)	2.69(0.52)	0.80(0.31)	-	-	1330

Random RL-tARD	0.09(0.02)	0.24(0.02)	1.14(0.07)/ 0.74(0.05)	0.23(0.09)	3.86(0.68)	0.71(0.35)	1.42(0.15)/ 3.76(0.23)	1.89(0.09)	1550
Informed RL-tARD	0.10(0.03)	0.13(0.01)	1.51(0.09)/ 1.80(0.22)	0.26(0.15)	2.38(0.29)	1.17(0.32)	1.77(0.17)/ 4.85/(0.28)	2.01(0.07)	1389

Figure 6

Posterior predictive distributions of the three RL-EAMs on the data of Miletić et al. 2021

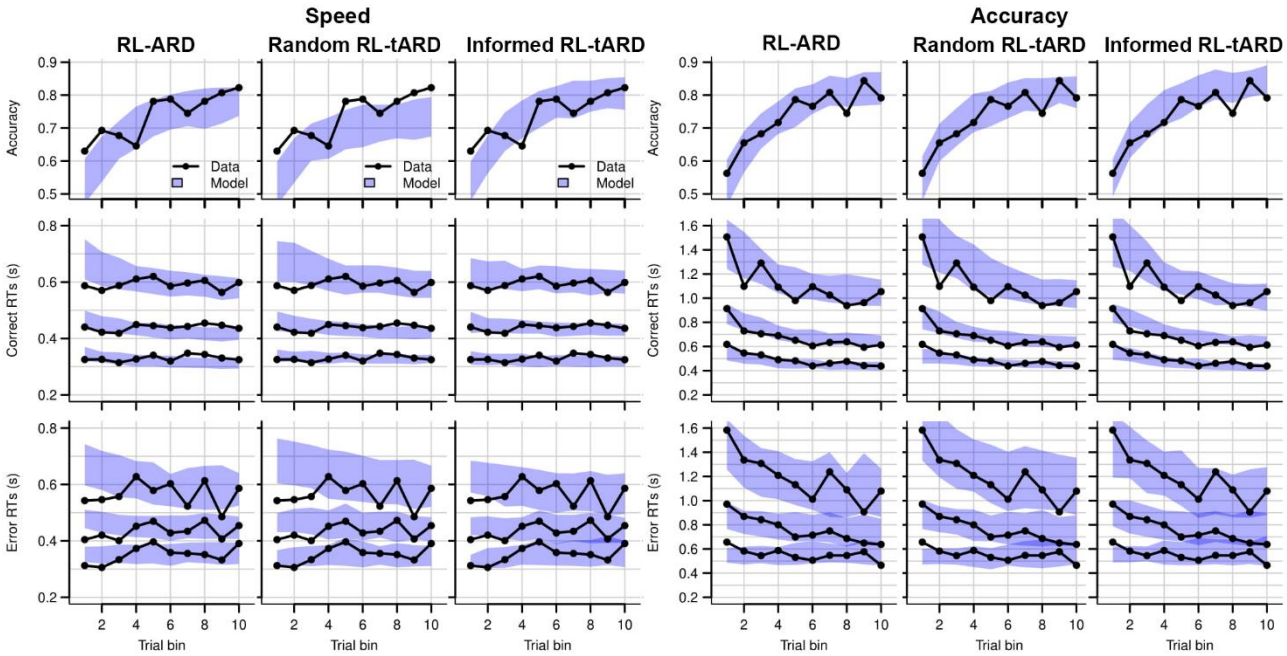


Data (black) and 95% credible interval of the posterior predictive distribution (purple) of the RL-ARD, random RL-tARD, and informed RL-tARD for the speed (left) and accuracy (right) manipulation. Accuracy (top row), correct response times (middle row), and error response times (bottom row) are depicted. The different lines in the response time distributions represent the 10^{th}, 50^{th} and 90^{th} percentile.

The data from Sewell and Stallman (2020), was best described by the RL-ARD without a timing accumulator (Table 2). The informed guess RL-tARD provided a slightly inferior account of the data and the random guess RL-tARD was a poorer fit. Again, deviance values suggested the same order in model preference.

Figure 7

Posterior predictive distributions of the three RL-EAMs on the data of Sewell and Stallman 2020



Data (black) and 95% credible interval of the posterior predictive distribution (purple) of the RL-ARD, random RL-tARD, and informed RL-tARD for the speed (left) and accuracy (right) manipulation. Accuracy (top row), correct response times (middle row), and error response times (bottom row) are depicted. The different lines in the response time distributions represent the 10th, 50th and 90th percentile. Note that the y-axes for the response time plots differ between the speed and accuracy columns to better visualize both distributions.

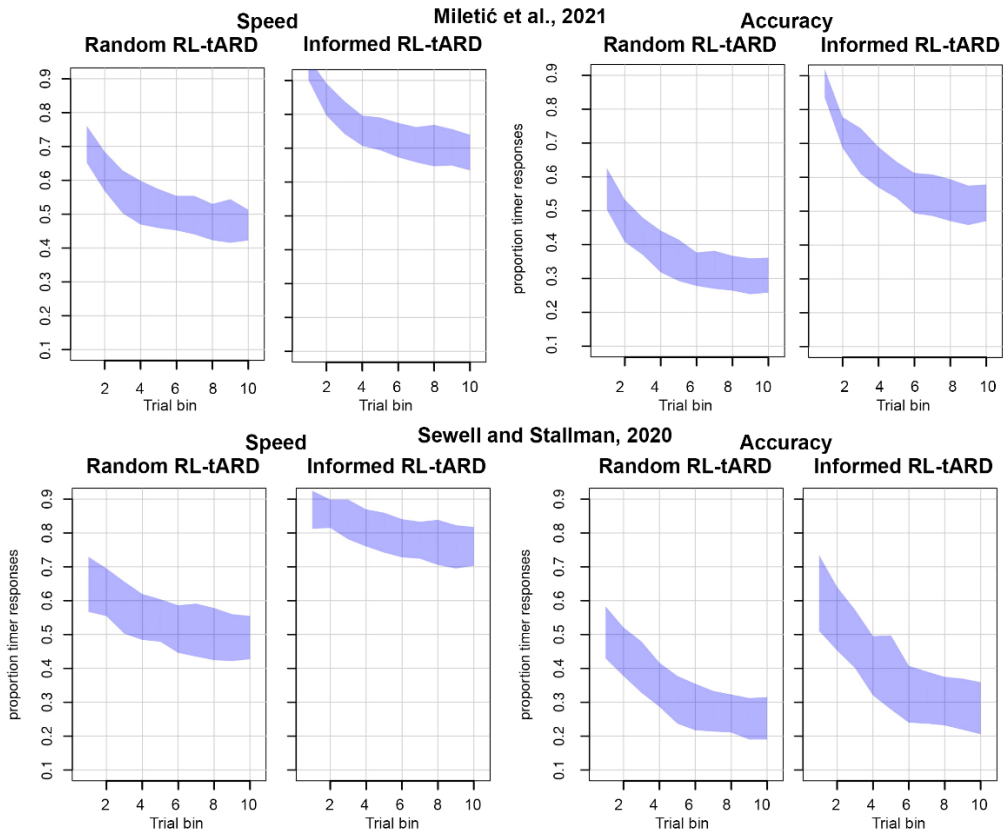
From Figure 7 we note that the speed-accuracy decoupling under speed emphasis was greater in the data from Sewell and Stallman (2020) than in the data from Miletić et al. (2021). Both the RL-ARD and the random RL-tARD cannot account for this decoupling under speed emphasis and over-estimate the learning-related decrease in response times. The informed RL-tARD better accounts for the constant response times under speed emphasis as the participant's data. Additionally, the random RL-tARD underestimates the learning-related increases in accuracy under speed pressure compared to the other two models. Both RL-tARDs slightly over-estimate the skew in the response time data of correct responses in the accuracy manipulation. We also note that compared to the data from Miletić et al. (2021), participants were slower to respond under accuracy emphasis, yet faster to respond under speed emphasis (note the different y-axes limits in Figure 6 and 7).

We also compared the proportion of timer-generated responses from the posterior predictive distributions of the two different RL-tARDs for both data-sets. From Figure 8 we note that each RL-tARD predicts a similar learning-related decrease in timing accumulator responses for both data sets. However, unsurprisingly the informed RL-tARD predicts a larger

proportion of timer-generated responses. Furthermore, both models predict a higher proportion of timer-generated responses for the data of Miletić et al. (2021). The differences in predicted timer responses between both data sets is especially large for the informed RL-tARD on the accuracy manipulation.

Figure 8

Posterior predictive distributions of the proportion of timer responses.



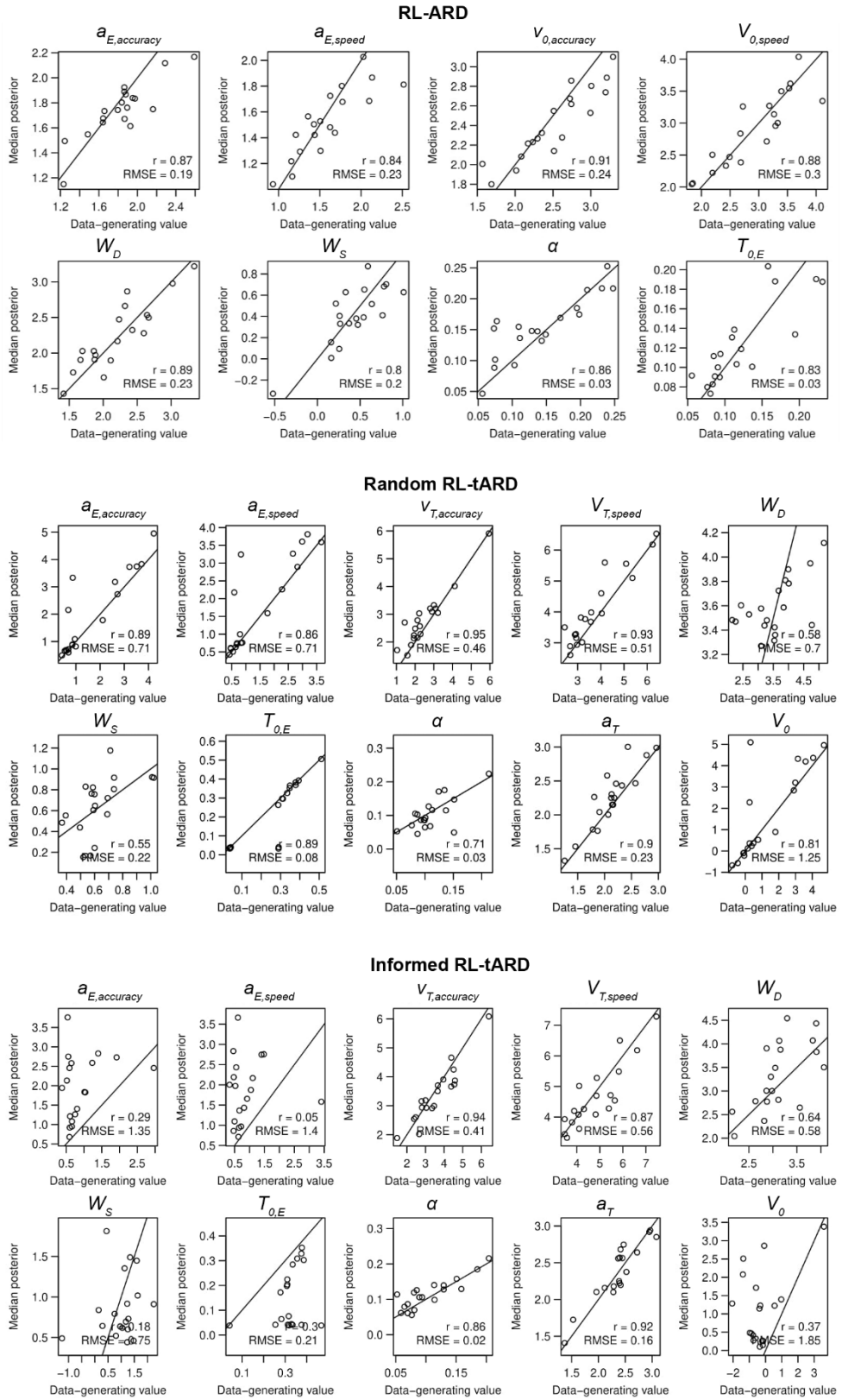
Proportion of timer responses based on the posterior predictive distribution of the random RL-tARD and informed RL-tARD for the speed (left) and accuracy (right) manipulation. Top row depicts the model for the data of Miletić et al. (2021), bottom row the model for the data of Sewell and Stallman (2020). Shaded areas correspond to the 95% credible interval of the posterior predictive distribution.

Lastly we performed a parameter recovery experiment for each model based on the data of Miletić et al., 2021 (Figure 9). We found that the RL-ARD and the random RL-tARD recovered their parameters well. The informed RL-tARD recovered some of its parameters more poorly. We observe a trade-off between the evidence threshold parameters and the non-decision time of the evidence accumulators, with higher recovered threshold values and lower

recovered non-decision time values. Poorer recovery has been a common problem for likelihood functions that are not analytically tractable (Evans et al., 2019; Voskuilen et al., 2016). Additionally, urgency models that rely on partial information are effectively underconstrained resulting in a trade-off between the non-decision time and the evidence threshold parameters. Even though not all responses are based on partial information in the informed RL-tARD, the high proportion of predicted timer responses, which do rely on partial information, has likely led to the same underconstraint.

Figure 9

Parameter recovery of the RL-ARD, random RL-tARD and the informed RL-tARD.



Parameter recovery for the three models. We fit each model to the data of Miletić et al. (2021), and then simulated the same design using the median parameter estimates of the posterior. Then we fit the model on the simulated data. Median posterior estimates of the simulated data (y-axis) are plotted against the data-generating values (x-axis). Each panel also shows Pearson's correlation coefficient r and the root mean square error (RMSE). Lines represent the diagonal x = y.

Discussion

In the current study we tested whether internal time estimation underlies the speed – accuracy trade-off (SAT) mechanisms of decision-making in learning. The SAT entails that people can flexibly balance response speed and response accuracy (Bogacz et al., 2010; Wickelgren, 1977). In previous work (Miletić et al., 2021; Sewell & Stallman, 2020), we observed that decision speed remains constant while learning under speed stress, while accuracy increases, which is at odds with current explanations of the SAT (Rae et al., 2014; Ratcliff & Rouder, 1998, Usher et al., 2002). We hypothesized that an internal timing mechanism could explain this decoupling between speed and accuracy.

To test this hypothesis we integrated the recently proposed reinforcement learning – advantage racing diffusion model (RL-ARD; Miletić et al., 2021), with a timing accumulator that is accumulating timing information in parallel to the evidence accumulators (Hawkins & Heathcote, 2021). The evidence accumulation process will stop prematurely if the timing accumulator reaches its threshold, resulting in a timer-generated response. The response speed of timer-generated responses is driven by the timing accumulation process.

We explored two options for the response made when the timing accumulator wins the race. In the first model we based the response made on a guess between the available options (random RL-tARD). In the second model we based the response made on partial information from the evidence accumulators (informed RL-tARD). The informed RL-tARD is similar to a collapsing bounds model, since both propose that time pressure forces responses to be made based on lesser amounts of evidence (Ditterich, 2006).

For both RL-tARDs, response speed and accuracy are not necessarily byproducts of the evidence accumulation process such as in the standard RL-ARD. We therefore hypothesized that the RL-tARDs could better explain the speed-accuracy decoupling observed in the two previous studies (Miletić et al., 2021; Sewell & Stallman, 2020).

We found that the RL-tARDs better explained the data of Miletić et al. (2021), compared to the RL-ARD. In contrasts, the RL-ARD provided the best account of the data of Sewell and

Stallman (2020). All models captured the accuracy, skew, variability and central tendency of the response time data well, which are considered essential aspects of evaluating decision-making models (Forstmann et al., 2016; Voss et al., 2013).

Additionally, we found that for timer-generated responses, reliance on partial information provided a better account than guessing for the data of Sewell and Stallman (2020). However, the opposite was true for the data of Miletić et al. (2021). Older studies suggested that decision-makers do not have access to partial information and guess when they are forced to end evidence accumulation prematurely (De Jong, 1991; Ratcliff, 1988, 2006). Nevertheless, a more recent study found that second guesses in perceptual decision-making do rely on partial information (McLean et al., 2020). Still the accuracy of non-terminated diffusion processes increases only marginally as a function of time (Cox & Miller, 1965; Ratcliff, 1988). Therefore, it is difficult to dissociate between relying on partial information and guessing, which is again highlighted in the current study.

In general, the timing models better explained the speed – accuracy decoupling under speed emphasis compared to the RL-ARD. Nevertheless, the RL-tARDs provided a poorer account of the learning-related increases in accuracy. The fact that the timing models did not capture all aspects of the data suggests that these models do not provide a full account of the different cognitive processes underlying decision-making in learning.

Still, Hawkins and Heathcote (2021) suggested that a timing accumulator better explains SAT mechanisms in decision-making. In their study they varied both the timing drift rate and mean evidence drift rates between the speed and accuracy manipulation. Their mean drift rates are similar to the urgency term we used, since urgency in our models also drives the mean rate of evidence accumulation, but not the difference in evidence accumulation between the response options.

Even though Hawkins and Heathcote (2021) varied mean drift rates between the speed and accuracy manipulation, we did not include separate urgency terms for the speed and accuracy manipulation in the RL-tARDs. We aimed to provide a more parsimonious account and reasoned that if an explicit time estimation process is indeed at the foundation of the SAT, the RL-tARDs should capture urgency differences between the speed and accuracy manipulation through means of the timing accumulator. Exploratory analysis of the data of Miletić et al.

(2021) confirmed that there was indeed no support for varying evidence urgency between the speed and accuracy conditions in the random RL-tARD. We therefore did not include separate urgency terms in the RL-tARDs in subsequent analyses.

However, our analysis of the data of Sewell and Stallman (2020) indicated that the omission of separate urgency terms in the RL-tARDs may not have been warranted, since we found that the RL-ARD outperformed the RL-tARDs in terms of absolute fit (disregarding model complexity). The RL-ARD did include separate urgency terms to account for the speed – accuracy manipulation. The RL-tARD with separate urgency terms is a generalization of the RL-ARD if we prevent timer-generated responses by setting the timing threshold to infinity or the timing drift rate to minus infinity. Therefore, the RL-tARD with separate urgency terms would fit the data at least equally well compared to the RL-ARD in terms of absolute fit.

The fact that the RL-ARD with separate urgency terms outperformed the RL-tARDs without separate urgency terms, suggests the timing accumulation process did not capture all aspects of within-trial urgency. This again highlights that our proposed method of time estimation is unlikely to provide the full account of the SAT mechanisms of decision making in learning tasks.

Still, the observed speed – accuracy decoupling is incompatible with other evidence accumulation theories of decision-making in learning, since drift rates are thought to increase with learning, which simultaneously increases response speed and accuracy (Miletić, Boag, & Forstmann, 2020; Pedersen et al., 2017). Furthermore, even though our implementation of time estimation could not fully account for all aspects of the data, a recent study did find that time estimation ability aids in speeded decision-making (Van Maanen et al., 2019). This raises the question whether different forms of time estimation can capture the SAT mechanisms of decision-making in learning. An alternative model that could also explain the speed – accuracy decupling, is a threshold learning model in which the evidence threshold is updated adaptively throughout the experiment based on an estimate of time available.

To elaborate, RL-ARDs assume that the rate of evidence accumulation is low when people have not yet learned the reward probabilities associated with each choice. Therefore, a decision-maker should set their evidence threshold quite low in the beginning to still meet

time demands. However, with learning, the drift rate for the correct choice increases and therefore the decision-maker could also increase their response threshold and still meet time demands, thereby increasing the probability of making the right choice. Simply put, the evidence threshold is updated based on an estimation of time available. Such a process would allow people to adjust response accuracy based on the expected amount of time available, which could potentially explain the observed decoupling.

Although our RL-tARDs are fundamentally different then the above described process, they do share some similarities. Both the above proposed threshold learning model and the RL-tARDs rely on time estimation to meet time demands in the decision-process. Furthermore, the RL-tARDs and the threshold learning model both propose that with learning, choices are based on increasing amounts of evidence. However, they differ in that the RL-tARDs pose that within-trial time estimation can prematurely end evidence accumulation. Consequently, when time runs out a decision is made regardless of how much evidence has accumulated. In contrast, a threshold learning model poses that within trial time estimation updates evidence thresholds between trials. Although the thresholds are likely low in the beginning of the learning process, there is still a minimal amount of evidence needed to commit to a decision.

In summary, we studied the cognitive mechanisms of the SAT underlying decision-making and learning. We proposed that time estimation occurs in parallel to the evidence accumulation process, and poses a deadline after which a response must be made. We found no compelling evidence for such mechanisms of time estimation. Still, we observed a speed – accuracy decoupling that cannot be explained with traditional explanations of the SAT in learning. We suggest an alternative model of time estimation in decision-making and learning that future studies could explore.

Anders, R., Alario, F., & Maanen, L. Van. (2019). The shifted Wald distribution for response time data analysis. *Psychological Methods*, *21*(3), 309–327.

Ando, T. (2011). Predictive Bayesian model selection. American Journal of Mathematical and Management Sciences, 31(1-2), 13-38.

Balci, F., & Simen, P. (2014). Decision processes in temporal discrimination. *Acta Psychologica*, *149*, 157–168. https://doi.org/10.1016/j.actpsy.2014.03.005

Balci, F., & Simen, P. (2016). A decision model of timing. *Current Opinion in Behavioral Sciences*, *8*, 94–101. https://doi.org/10.1016/j.cobeha.2016.02.002

Bode, S., Bennett, D., Sewell, D. K., Paton, B., Egan, G. F., Smith, P. L., & Murawski, C. (2018). Dissociating neural variability related to stimulus quality and response times in perceptual decision-making. *Neuropsychologia*, *111*(October 2017), 190–200. https://doi.org/10.1016/j.neuropsychologia.2018.01.040

Boehm, U., Hawkins, G. E., Brown, S., van Rijn, H., & Wagenmakers, E. J. (2016). Of monkeys and men: Impatience in perceptual decision-making. *Psychonomic Bulletin and Review*, *23*(3), 738–749. https://doi.org/10.3758/s13423-015-0958-5

Bogacz, R., & Larsen, T. (2011). Integration of reinforcement learning and optimal decision-making theories of the basal ganglia. *Neural Computation*, *23*(4), 817–851. https://doi.org/10.1162/NECO_a_00103

Bogacz, R., Wagenmakers, E. J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed-accuracy tradeoff. *Trends in Neurosciences*, *33*(1), 10–16. https://doi.org/10.1016/j.tins.2009.09.002

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

Cisek, P., Puskas, G. A., & El-Murr, S. (2009). Decisions in changing conditions: The urgency-gating model. *Journal of Neuroscience*, *29*(37), 11560–11571. https://doi.org/10.1523/JNEUROSCI.1844-09.2009

Cox, D. R., & Miller, H. D. (1965). The theory of stochastic processes (Vol. 134). CRC press.

Dayan, P., & Daw, N. D. (2008). Decision theory, reinforcement learning, and the brain. *Cognitive, Affective and Behavioral Neuroscience*, *8*(4), 429–453. https://doi.org/10.3758/CABN.8.4.429

De Jong, R. (1991). Partial information or facilitation? Different interpretations of results from speed-accuracy decomposition. *Perception & Psychophysics*, *50*(4), 333–350. https://doi.org/10.3758/BF03212226

Ditterich, J. (2006). Evidence for time-variant decision making. *European Journal of*

Neuroscience, *24*(12), 3628–3641. https://doi.org/10.1111/j.1460-9568.2006.05221.x

Donkin, C., Brown, S. D., & Heathcote, A. (2009). The overconstraint of response time models: Rethinking the scaling problem. *Psychonomic Bulletin and Review*, *16*(6), 1129–1135. https://doi.org/10.3758/PBR.16.6.1129

Evans, N. J., & Hawkins, G. E. (2019). When humans behave like monkeys: Feedback delays and extensive practice increase the efficiency of speeded decisions. *Cognition*, *184*(June 2018), 11–18. https://doi.org/10.1016/j.cognition.2018.11.014

Evans, N. J., Trueblood, J. S., & Holmes, W. R. (2019). A parameter recovery assessment of time-variant models of decision-making. *Behavior Research Methods*, *52*(1), 193–206. https://doi.org/10.3758/s13428-019-01218-0

Fontanesi, L., Gluth, S., Spektor, M. S., & Rieskamp, J. (2019). A reinforcement learning diffusion decision model for value-based decisions. *Psychonomic Bulletin and Review*, *26*(4), 1099–1121. https://doi.org/10.3758/s13423-018-1554-2

Fontanesi, L., Palminteri, S., & Lebreton, M. (2019). Decomposing the effects of context valence and feedback information on speed and accuracy during reinforcement learning: a meta-analytical approach using diffusion decision modeling. *Cognitive, Affective and Behavioral Neuroscience*, *19*(3), 490–502. https://doi.org/10.3758/s13415-019-00723-1

Forstmann, B. U., Ratcliff, R., & Wagenmakers, E.-J. J. (2016). Sequential Sampling Models in Cognitive Neuroscience: Advantages, Applications, and Extensions. *Annual Review of Psychology*, *67*(1), 641–666. https://doi.org/10.1146/annurev-psych-122414-033645

Frank, M. J., Gagne, C., Nyhus, E., Masters, S., Wiecki, T. V., Cavanagh, J. F., & Badre, D. (2015). FMRI and EEG predictors of dynamic decision parameters during human reinforcement learning. *Journal of Neuroscience*, *35*(2), 485–494. https://doi.org/10.1523/JNEUROSCI.2036-14.2015

Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By Carrot or by Stick : Cognitive Reinforcement Learning in Parkinsonism. *Science*, *306*(5703), 1940–1943.

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. https://doi.org/10.1523/JNEUROSCI.2410-14.2015

Hawkins, G. E., & Heathcote, A. (2021). Racing against the clock: Evidence-based versus time-based decisions. *Psychological Review*, *128*(2), 222–263. https://doi.org/10.1037/rev0000259

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

Kelly, S. P., Corbett, E. A., & O'Connell, R. G. (2020). Neurocomputational mechanisms of prior-informed perceptual decision-making in humans. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-020-00967-9

McElree, B., & Dosher, B. A. (1989). Serial Position and Set Size in Short-Term Memory: The Time Course of Recognition. *Journal of Experimental Psychology: General*, *118*(4), 346–373. https://doi.org/10.1037/0096-3445.118.4.346

McLean, C. S., Ouyang, B., & Ditterich, J. (2020). Second Guessing in Perceptual Decision-Making. *Journal of Neuroscience*, *40*(26), 5078–5089. https://doi.org/10.1523/JNEUROSCI.2787-19.2020

Miletić, S., Boag, R. J., & Forstmann, B. U. (2020). Mutual benefits: Combining reinforcement learning with sequential sampling models. *Neuropsychologia*, *136*. https://doi.org/10.1016/j.neuropsychologia.2019.107261

Miletić, S., Boag, R. J., Trutti, A. C., Forstmann, B. U., & Heathcote, A. (2020). *A new model of decision processing in instrumental learning tasks*. 1–28.

Miletić, S., Boag, R. J., Trutti, A. C., Stevenson, N., Forstmann, B. U., & Heathcote, A. (2021). A new model of decision processing in instrumental learning tasks. *ELife*, *10*. https://doi.org/10.7554/elife.63055

Miletić, S., & van Maanen, L. (2019). Caution in decision-making under time pressure is mediated by timing ability. *Cognitive Psychology*, *110*(January), 16–29. https://doi.org/10.1016/j.cogpsych.2019.01.002

Nunez, M. D., Vandekerckhove, J., & Srinivasan, R. (2017). How attention influences perceptual decision making: Single-trial EEG correlates of drift-diffusion model parameters. *Journal of Mathematical Psychology*, *76*, 117–130. https://doi.org/10.1016/j.jmp.2016.03.003

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. https://doi.org/10.3758/s13423-016-1199-y

Rae, B., Heathcote, A., Donkin, C., Averell, L., & Brown, S. (2014). The hare and the tortoise: Emphasizing speed can change the evidence used to make decisions. *Journal of Experimental Psychology: Learning Memory and Cognition*, *40*(5), 1226–1243. https://doi.org/10.1037/a0036801

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

Ratcliff, R. (1980). A note on modeling accumulation of information when the rate of accumulation changes over time. *Journal of Mathematical Psychology*, *21*(2), 178–184. https://doi.org/10.1016/0022-2496(80)90006-1

Ratcliff, R. (1988). Continuous Versus Discrete Information Processing: Modeling Accumulation of Partial Information. *Psychological Review*, *95*(2), 238–255. https://doi.org/10.1037/0033-295X.95.2.238

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

Ratcliff, R., Hasegawa, Y. T., Hasegawa, R. P., Smith, P. L., & Segraves, M. A. (2007). Dual diffusion model for single-cell recording data from the superior colliculus in a brightness-discrimination task. *Journal of Neurophysiology*, *97*(2), 1756–1774. https://doi.org/10.1152/jn.00393.2006

Ratcliff, R., & Rouder, J. N. (1998). Modeling Response Times for Two-Choice Decisions. *Psychological Science*, *9*(5), 347–356. https://doi.org/10.1111/1467-9280.00067

Ratcliff, R., & Smith, P. L. (2004). A Comparison of Sequential Sampling Models for Two-Choice Reaction Time. *Psychological Review*, *111*(2), 1–101. http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-295X.111.2.333%0Apapers3://publication/doi/10.1037/0033-295X.111.2.333

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends in Cognitive Sciences*, *20*(4), 260–281. https://doi.org/10.1016/j.tics.2016.01.007

Reed, A. V. (1973). Speed-Accuracy Trade-Off in Recognition Memory. *Science*, *181*(4099), 574–576. https://doi.org/10.1126/science.181.4099.574

Sewell, D. K., Jach, H. K., Boag, R. J., & Van Heer, C. A. (2019). Combining error-driven models of associative learning with evidence accumulation models of decision-making. *Psychonomic Bulletin and Review*, *26*(3), 868–893. https://doi.org/10.3758/s13423-019-01570-4

Sewell, D. K., & Stallman, A. (2020). Modeling the Effect of Speed Emphasis in Probabilistic Category Learning. *Computational Brain & Behavior*, *3*(2), 129–152. https://doi.org/10.1007/s42113-019-00067-6

Simen, P., Vlasov, K., & Papadakis, S. (2016). Scale (In)variance in a unified diffusion model of decision making and timing. *Psychological Review*, *123*(2), 151–181. https://doi.org/10.1037/rev0000014

Smith, P. L., & Ratcliff, R. (2009). An Integrated Theory of Attention and Decision Making in Visual Signal Detection. *Psychological Review*, *116*(2), 283–317. https://doi.org/10.1037/a0015156

Ter Braak, C. J. F. (2006). A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: Easy Bayesian computing for real parameter spaces. *Statistics and Computing*, *16*(3), 239–249. https://doi.org/10.1007/s11222-006-8769-1

Thura, D., Beauregard-Racine, J., Fradet, C. W., & Cisek, P. (2012). Decision making by urgency gating: Theory and experimental support. *Journal of Neurophysiology*, *108*(11), 2912–2930. https://doi.org/10.1152/jn.01071.2011

Tillman, G., Van Zandt, T., & Logan, G. D. (2020). Sequential sampling models without random between-trial variability: the racing diffusion model of speeded decision making. *Psychonomic Bulletin and Review*, 911–936. https://doi.org/10.3758/s13423-020-01719-6

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. In *Psychological Review* (Vol. 108, Issue 3, pp. 550–592). https://doi.org/10.1037/0033-295X.108.3.550

Usher, M., Olami, Z., & McClelland, J. L. (2002). Hick's law in a stochastic race model with speed-accuracy tradeoff. *Journal of Mathematical Psychology*, *46*(6), 704–715.

https://doi.org/10.1006/jmps.2002.1420

van Maanen, L., & Miletić, S. (2020). The interpretation of behavior-model correlations in unidentified cognitive models. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/s13423-020-01783-y

van Maanen, L., van der Mijn, R., van Beurden, M. H. P. H., Roijendijk, L. M. M., Kingma, B. R. M., Miletić, S., & van Rijn, H. (2019). Core body temperature speeds up temporal processing and choice behavior under deadlines. *Scientific Reports*, *9*(1), 1–12. https://doi.org/10.1038/s41598-019-46073-3

van Ravenzwaaij, D., Brown, S. D., Marley, A. A. J., & Heathcote, A. (2019). Accumulating Advantages: A New Conceptualization of Rapid Multiple Choice. *Psychological Review*, 1–44. https://doi.org/10.1037/rev0000166

Voskuilen, C., Ratcliff, R., & Smith, P. L. (2016). Comparing fixed and collapsing boundary versions of the diffusion model. *Journal of Mathematical Psychology*, *73*, 59–79. https://doi.org/10.1016/j.jmp.2016.04.008

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion models in experimental psychology: A practical introduction. *Experimental Psychology*, *60*(6), 385–402. https://doi.org/10.1027/1618-3169/a000218

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, *41*(1), 67–85. https://doi.org/10.1016/0001-6918(77)90012-9

Zandbelt, B., Purcell, B. A., Palmeri, T. J., Logan, G. D., & Schall, J. D. (2014). Response times from ensembles of accumulators. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(7), 2848–2853. https://doi.org/10.1073/pnas.1310577111