Supplementary Materials for "ParAcT-DDM: A diffusion-based framework for modelling systematic, time-varying cognitive processes"

Manikya Alister & Nathan J. Evans

Contents

1 Simulation Study: Investigating the measurement validity of time-varying functions

In order to determine the practical utility of this modelling framework, we need to assess the extent to which we can identify time-varying processes that we know exist in the data. This can be achieved by simulating a data set using a specific time-varying function and subsequently fitting various models with different time-varying functions – or potentially without any time-varying parameters – to the same data. We can then asses whether the model that generated the data outperforms the other models. This type of assessment is typically referred to as "model recovery", as the goal is to try and recover the generating model from a set of alternate candidate models. Performing model recoveries is crucial for quantifying the granularity of our approach, so that we can understand how similar different time-varying functions can be before we can no longer distinguish which function generated the data. For instance, distinguishing between a linear time-varying function and an exponential time-varying function might be feasible, but discerning between a power function and an exponential function, which are closely related, could be more challenging. Overall, these analyses help us to assess the viability of using these models in empirical settings, by determining the extent to which different time-varying functions are detectable within data, and how strong our conclusions can be about empirical data using these models (e.g., can we conclude in favour of a specific function for each parameter, or only that a specific parameter appears to be varying over time?).

Method

Data generating model parameters were sampled within a pre-specified range using Latin Hypercube Sampling (Evans & Servant, 2022; Evans et al., 2020; McKay et al., 1979). The generating ranges for each parameter are in Supplementary Material 3. For each generating model, we simulated 100 participants in a 1000 trial experiment (see the Results and Discussion section to see which models were simulated). Models were fit using the same method as described in the main manuscript. Because our simulations were fit assuming a 1,000 trial experiment, please note that any conclusions of this study only pertain to an experiment with this many trials per person. An experiment with fewer trials per person would likely have worse measurement properties, whereas an experiment with more would likely have better measurement properties.

Table 1: Number of participants best fit by the generating model, and the overall probability according to BIC (AIC) that the threshold (a) power and exponential models were the best fit to data generated from each model respectively.

		Fitted Model	
Generating Model	n Best Fit	a Exponential	a Power
a Power	81 (81)	0.2 (0.2)	**0.8 (0.8)**
a Exponential	100 (100)	**0.98 (0.98)**	0.02 (0.02)

Results and Discussion

Running these model recoveries was a computationally strenuous task, especially if we were to run recoveries with every time-varying function described in the main manuscript. Given this large computational cost, we decided to only run model recoveries on a subset of models. Specifically, we focused on simulations that we believed would help answer our key questions of interest about the measurement properties of our ParAcT-DDM variants, as well as situations with trial-varying functions (i.e., not the block–varying functions or trial+block-varying function), as trial-varying models assume the most fine-grained changes across time, and therefore, are likely to be the most difficult to recover and distinguish. Model performance was determined using both BIC and AIC. We chose to look at both BIC and AIC as each criterion strikes a different balance between flexibility (i.e., in this case, the number of free parameters in the model) and how well the model fits the data, with BIC tending to penalise flexibility more heavily than AIC (Evans, 2019). Therefore, using both criteria ensures that readers are aware of how the results may or may not change based on the choice of model selection criterion. We used these information criteria to determine which model best accounted for the data of each participant, both by comparing the raw criterion values and the relative probabilities that a given model was best accounting for the data according to BIC and AIC weights (Wagenmakers & Farrell, 2004) using the "modelProb" package (Alister, 2022).

Are Exponential and Power Processes Distinguishable from Each Other?

First, we performed a model recovery assessment between the standard power and exponential models, which we focused on for two reasons: (1) we predicted that these models would likely be the hardest to distinguish from one another, and (2) if these models are not distinguishable, then our subsequent analyses could focus on only one of these functions, reducing the computational cost. Here, we focused on assessments of drift rate and threshold separately, though note that a later simulation ("Are Drift Rate (v) and Threshold (a) Time-Varying Processes Distinguish-

Table 2: Probability according to BIC (AIC) that the drift rate (v) power and exponential models were the best fit to data generated from each model respectively, collapsed across participants.

		Fitted Model	
Generating Model	n Best Fit	v Exponential	v Power
v Power	56 (56)	0.45 (0.45)	**0.55 (0.55)**
v Exponential	83 (83)	**0.75 (0.75)**	0.25 (0.25)

able from Each Other?") addresses whether drift rate and threshold time-varying processes are distinguishable from one another.

The results for the threshold (a) and drift rate (v) trial-varying models are presented in Tables 1 and 2, respectively. When examining the a-varying models, both the exponential and power trial-varying models exhibited a greater than chance probability of being the best-performing model when they were the data-generating model. However, the exponential model appeared to be much more recoverable than the power model, as data generated by the power model still had a relatively high likelihood of being mis-classified as an exponential model, suggesting that the exponential model is able to better mimic the power model than vice versa (see Wagenmakers et al., 2004, for a more in-depth discussion of mimicry and model recovery). When examining the v-varying models, they appeared to be less recoverable overall. While both models had a greater than chance probability of being correctly identified as the data-generating model, the exponential model demonstrated better recovery between the two, again suggesting that the exponential model is able to better mimic the power model than vice versa. Due to the reasonably high probability of mis-classifying power and exponential time-varying processes as the other, especially when examining v, we decided to focus on the exponential model in subsequent analyses, mainly due to (1) the greater previous success of exponential functions in the literature (e.g., Evans et al., 2018; Heathcote et al., 2000), and (2) the greater mimicry of the exponential model of the power model than vice versa, as the exponential model would still likely be able to capture data that were truly a power function relatively well. This decision was further justified by the parameters in the power models having poorer measurement properties compared to the exponential models (see Supplementary Material 4C and 4G).

Are Different Trial-Varying Processes Distinguishable from Each Other

Next, we performed model recoveries for the standard diffusion model, as well as linear, exponential, and delayed-exponential trial-varying functions for threshold (a, see Table 6) and for drift rate (v, see Table 4). As with the previous simulation, we focused on assessments of drift rate and

Table 3: Probability according to BIC (AIC) that the threshold (a) trial-varying models were the best fit to data generated from each model respectively, collapsed across simulated participants and the number of participants best fit by each generating model. Models are ordered by complexity, with each proceeding model having one extra free parameter relative to the one before it.

		Fitted Model			
Generating Model	n Best	*Standard DDM*	*a Linear*	*a Exponential*	*a Delayed Exp.*
*S*tandard DDM*	100 (82)	**0.93 (0.44)**	0.01 (0.04)	<.01 (0.09)	<.01 (0.04)
a Linear	81 (68)	0.09 (0.04)	**0.79 (0.59)**	0.02 (0.09)	0.1 (0.28)
a Exponential	99 (91)	<.01 (<.01)	0.01 (<.01)	**0.95 (0.73)**	0.04 (0.27)
a Delayed Exp.	60 (78)	0.01 (<.01)	0.08 (0.05)	0.33 (0.16)	**0.58 (0.79)**

*Note. *The standard diffusion model data was fit by both a models and v models, which is why the probabilities in each respective table for that row do not add up to 1. Delayed exp. stands for delayed exponential. n best stands for the number of participants that were best fit to the generated data by the generating model.*

threshold separately, meaning that chance performance in each recovery would be a 0.25 probability of recovering the data generating model. However, as the standard diffusion model appeared in both recoveries, it gave the standard diffusion model the more difficult task of competing against all 6 other models in the assessment of its own data, meaning that chance performance for the standard diffusion model was 0.14.

All of the a-varying models were recoverable at well above chance, although recovery was more difficult for the a delayed exponential model. Recovery for the a delayed exponential model was better when using AIC, though using AIC also tended to increase the likelihood of false alarms for data generated by the other models (i.e., AIC generally preferred more flexible models). The v-varying models exhibited similar results, but the exponential and delayed-exponential models tended to be less recoverable compared to their a equivalents. Indeed, the v delayed exponential model performed *worse* than chance. Given that the a delayed exponential model performed reasonably well, and that the delayed models reflect a unique and empirically relevant theory of practice (Evans et al., 2018; Rickard, 1997, 2004), we decided to retain these models in our assessments of empirical data. However, readers should be aware that a poor performance of these models when assessed against empirical data may reflect their difficulty in being distinguished from their simpler counterparts, rather than their poor explanatory power of the data.

In summary, almost all of the data generating models could be identified at a reasonably high rate, but this tended to decrease with increased model flexibility, particularly for the v varying models, and most notably for the v delayed exponential model.

Table 4: Probability according to BIC (AIC) that the threshold (v) trial-varying models were the best fit to data generated from each model respectively, collapsed across participants.

		Fitted Model			
Generating Model	n Best Fit	Standard DDM	v Linear	v Exponential	v Delayed Exp.
*S*tandard DDM*	100 (82)	**0.93 (0.44)**	0.05 (0.26)	<.01 (0.09)	<.01 (0.05)
v Linear	87 (85)	0.12 (0.05)	**0.84 (0.65)**	0.04 (0.17)	<.01 (0.13)
v Exponential	53 (80)	0.33 (0.05)	0.12 (0.08)	**0.52 (0.59)**	0.02 (0.29)
v Delayed Exp.	11 (51)	0.06 (0.02)	0.15 (0.07)	0.6 (0.37)	**0.19 (0.54)**

*Note. *The standard diffusion model data was fit by both a models and v models, which is why the probabilities in each respective table for that row do not add up to 1. Delayed exp. stands for delayed exponential. n best stands for the number of participants that were best fit to the generated data bu the generating model.*

Are Drift Rate (v) and Threshold (a) Time-Varying Processes Distinguishable from Each Other?

Evidence accumulation models are often used to disentangle the cognitive processes responsible for behaviour in a particular task. For example, a researcher might be interested in whether a certain manipulation causes changes in drift rate across conditions, changes in threshold, both, or neither. Therefore, we believe that one of the most crucial questions for our ParAcT-DDM framework is whether we can distinguish between time-varying drift-rate processes and time-varying threshold processes. However, due to the computationally intensive nature of our simulations, our previous simulations only assessed whether a models could be distinguished from other a models, and v models could be distinguished from other v models. Given that our previous simulations have shown that different v-varying models seem to be less distinguishable from each other than different a-varying models are from each other, it is possible that a-varying processes may also provide better mimicry of v-varying processes than vice versa. To test whether trial-varying v processes and a processes can be distinguished from one another, we performed a model recovery that compared only the exponential trial-varying models for v and a. Specifically, we fit the a and v exponential models to each of their respective data sets.

The results can be seen in Table 5, and show that a-exponential trial-varying processes were almost never mis-identified as v-exponential trial-varying processes. Although, consistent with the previous simulations, v-varying processes were more difficult to recover, there was still a reasonably low probability of mis-classifying them as a-varying processes.

What are the Measurement Properties of the Trial-Varying Parameters?

The previous sections investigated whether different time-varying models can be distinguished from one another, which is important for justifying empirical assessments of whether data appear to exhibit time-varying processes, and which time-varying processes they appear to exhibit (i.e., "model comparison"; see Crüwell et al., 2019, for a discussion). However, a common goal of research is often to make inferences about the parameters themselves, and whether they differ between experimental conditions and/or groups (i.e., "model application"; see Crüwell et al., 2019, for a discussion). Similarly, a researcher applying our ParAcT-DDM framework might be interested in measuring particular parameters of these time-varying functions. For example, a researcher applying an exponential time-varying function might be interested in the participants' rate of change (η) to see how quickly the parameter changes, or their asymptote (α) to assess the parameter after learning has finished (see equations in main manuscript). In this section we describe how reliably the individual parameters in these time-varying functions can be measured.

The measurement properties of each parameter were assessed by evaluating the relationship between the parameters used to generate the simulated data for each participant and the parameters estimated[1] from the fit of the model to simulated data, quantitatively through correlations and qualitatively through scatterplots. Specifically, parameters that have good measurement properties should have high correlations between the generating and estimated parameters, since that would indicate that the model is correctly estimating the "true" parameters. The full parameter recovery assessments that show the correlations and scatterplots between the generated and estimated parameters for every model assessed in the previous model recovery simulations can be found in Supplementary Material 4.

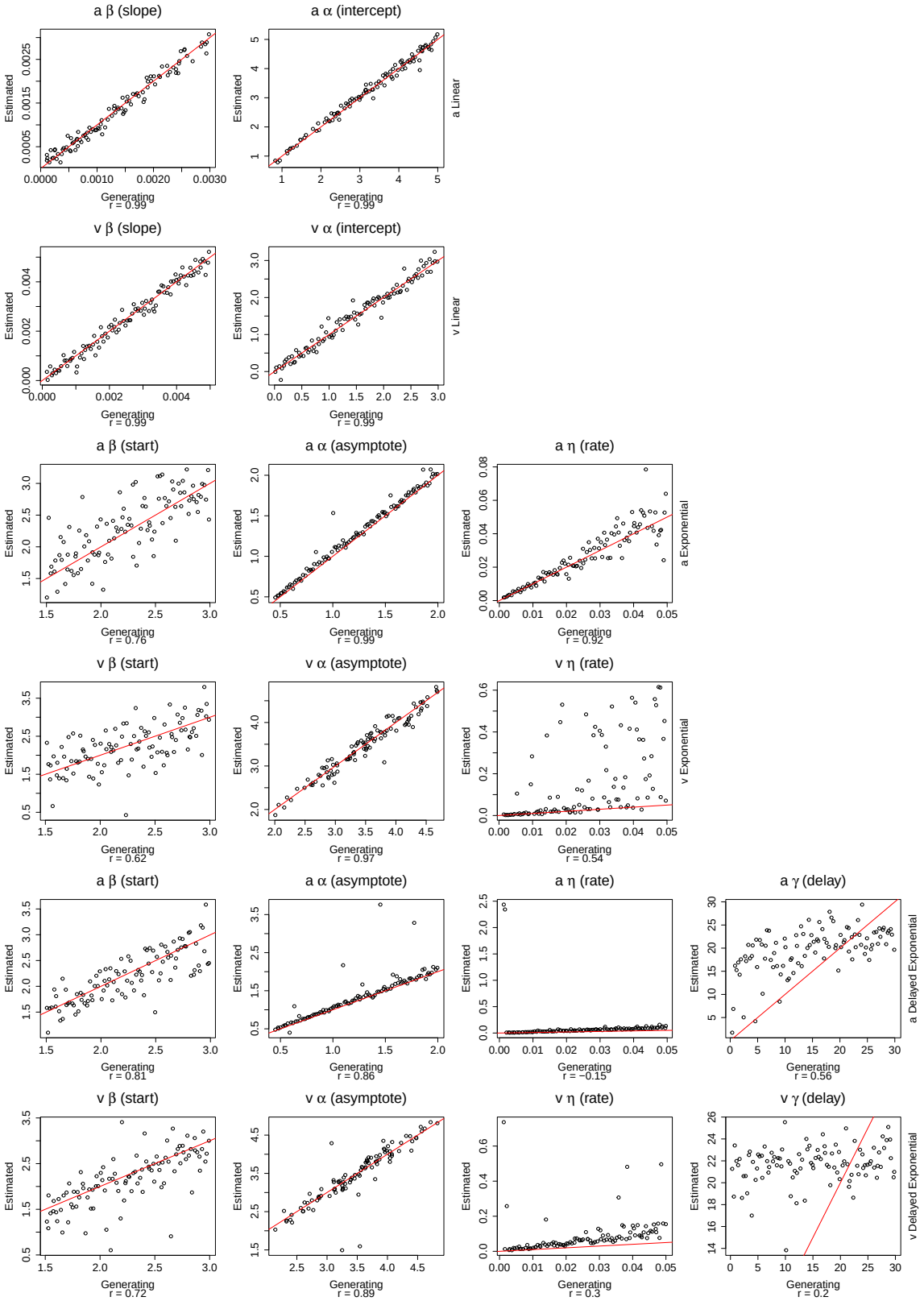
As expected based on previous research (e.g., Alister et al., 2023; Lerche & Voss, 2016; Lerche

[1]Specifically, the estimated parameters used for the parameter recovery assessment were those that provided the maximum likelihood.

Table 5: Probability according to BIC (AIC) that Exponential drift rate (v) and threshold (a) models were the best fit to data generated from each model respectively, collapsed across participants.

		Fitted Model	
Generating Model	n Best Fit	v Exponential	a Exponential
v Exponential	94 (94)	**0.91 (0.91)**	0.09 (0.09)
a Exponential	100 (100)	<.01 (<.01)	**>0.99 (>.99)**

Figure 1. Correlation between time-varying ParAcT parameters that were used to generated trial-varying data and the parameters that were estimated when the model was fit to the data. A high correlation indicates good measurement properties for that parameter. Note that this figure only includes the time-varying parameters, but the full parameters of each model can be found in Supplementary Material 4.



et al., 2017), all of the parameters for the standard diffusion model showed good measurement properties, with generating and estimating parameters having correlations of at least $r = .96$ (see Supplementary Material 3A). In all of the trial-varying models, the static diffusion model parameters (i.e., the parameters that assumed no change across trials) also showed good measurement properties, with only z in the a power model having a correlation lower than $r = .95$ ($r = .91$), meaning that including a time-varying function for a parameter did not adversely affect the measurement properties of the other, non time-varying parameters.

Perhaps most importantly, the estimated time-varying parameters also showed good measurement properties in most cases. As shown in the first two rows of Figure 1, the slope (β) and intercept (α) parameters for the a and v linear functions showed good measurement properties. As shown in the third row of Figure 1, both the asymptote (α) and rate (η) parameters for the a exponential function showed good measurement properties. Although the start (β) was more difficult to recover, the correlations suggest that researchers can be reasonably confident that estimated β parameter will be at least close to the "true" parameter the majority of the time, and that the estimates for α and η should be fairly precise. However, as shown in the fourth row of Figure 1, the v exponential parameters were more difficult to recover than the parameters for the a exponential function; while the asymptote (α) parameter still recovered very well, the recovery of the rate parameter (η) became much poorer with generating values beyond 0.025.

Apart from a few outliers, the start (β), rate (η), and asymptote (α) parameters in both the v and a delayed-exponential models could be recovered reasonably reliably. However, the delay (γ) parameter had very poor measurement properties, particularly for v, which could explain its poor performance in the model recovery. Most importantly though, we believe that our parameter recovery results show that researchers can be confident that both (1) the measurement properties of the parameters fixed over time should remain reasonably reliable when using the ParAcT-DDM models tested in this simulation, and (2) the measurement properties of the parameters that correspond to the value at the end of learning (α) and the rate of learning (η) are reliable in most circumstances.

Discussion

In this simulation study, we assessed the measurement properties of our ParAcT-DDM framework through model recovery simulations, where we generated data using a specific ParAcT-DDM variant, and assessed whether that variant was able to provide a better account of its own data than the standard diffusion model and a number of other candidate ParAcT-DDM variants. We observed

that the ParAcT-DDM framework could effectively distinguish between all of the different threshold (a/caution) trial-varying functions that we assessed, although delayed exponential models were much harder to distinguish compared to the other candidate functions. However, this could potentially be due to the parameter ranges that we chose for the simulations. For example, it is possible that the delayed model may be easier to separate from the regular exponential model with an extremely long delay and an extremely large – though practically unrealistic from an experimental perspective – number of trials.

On a related note, it is important to make clear that under certain extreme parameter values, all of the ParAcT models generated here can take the shape of simpler models. For example, with a fast enough or slow enough rate parameter, any asymptotic function (e.g., power, exponential) will essentially resemble a function that assumes no change over time (a straight line). This means that if the true generating curve resembles a straight line (even though it was technically generated with an asymptotic process), the start point parameter can take any value without affecting the likelihood, and minor variations in rate will also have little affect on the likelihood. These points mean that certain combinations of generating parameters will make both model and parameter recovery noisier, therefore offering an explanation for why some parameters were harder to recover at more extreme parameter values (e.g., the rate parameter).

There was a greater difficulty in distinguishing between different drift rate (v/task efficiency) functions, particularly for functions that made qualitatively similar predictions, such as exponential and power functions, as well as exponential and delayed-exponential functions. Specifically, the model comparison tended to overweight simpler models than those that generated the data, with v-exponential being the best performing model for data generated by v-delayed exponential. While it was clear that drift rate processes were more difficult to distinguish from one another than threshold processes in our simulations, it is unclear why this was the case. One possibility is that the estimates of the parameters of the drift rate functions are noisier than the estimates of the threshold functions, which is supported by the credible intervals in Figure 12 in the main manuscript; however, it is unclear why the estimates for drift rate are noisier than for threshold.

Despite the difficulty in distinguishing between different drift rate functions, our simulations suggested that our ParAcT-DDM framework is able to accurately distinguish between changes in drift rate and changes in threshold. Importantly, these simulations showcase that our ParAcT-DDM framework has the ability to determine whether changes in response time over the course of an experiment are the result of changes in drift rate, threshold, or both. However, it is important to note that we only tested *exponential* drift rate and threshold models against one another, meaning

it is possible that drift rate and threshold changes might potentially be harder to distinguish from one other if the changes occur according to a different function. Our simulations also lend support to the robustness of the only other study we are aware of to apply time-varying functions to diffusion models (Cochrane et al., 2023), which used exponential functions to identify whether drift rate or threshold drove learning effects across time.

Our simulations study also assessed the identifiability of the time-varying cognitive constructs in different ParAcT-DDM variants through parameter recovery simulations, where we generated data using specific parameter values from the ParAcT-DDM variant, and assessed whether that variant was able to accurately estimate the data generating parameters. We observed that the simpler ParAcT-DDM variants – such as the linear and exponential models – showed robust measurement properties for the key parameters relating to how people change over time – such as the rate and asymptote parameters of the exponential function – particularly for changes in threshold. Importantly, these robust estimation properties mean that future studies could investigate whether different experimental conditions result in changes in different parts of the learning process. For example, learning studies could use the ParAcT-DDM framework to compare how the rate of increase/decrease in a parameter, or the asymptote parameter of an exponential function, changes across time for different experimental manipulations. However, it is important to note that the "delay" parameter of the delayed exponential function, which is the only parameter that distinguishes delayed exponential functions from standard exponential functions, was poorly identified for both changes in drift rate and threshold, as were the parameters for the power functions (the full parameter recoveries are available in Supplementary Material 4). Furthermore, we wish to again note that our recoveries assumed a 1000 trial experiment, meaning that our inferences about the measurement properties of the parameters and the distinguishability of the models may not be applicable to situations with relatively few trials per person, where inferences might become noisier and less robust.

2 Model Priors

Definition of terms

TN = truncated normal distribution (mean, standard deviation, lower limit, upper limit)

N = normal distribution (mean, standard deviation)

Prior distributions

Standard Diffusion Model Parameters

$a = TN(2, 1, 0, \infty)$

$t0 = TN(0.3, 0.1, 0, \infty)$

$v = N(3, 1)$

$z = TN(0.5, 0.1, 0, 1)$

Linear Parameters

$\beta_a = TN(2, 1, 0, \infty)$

$\alpha_a = TN(2, 1, 0, \infty)$

$\beta_v = N(0.5, 0.5)$

$\alpha_v = N(2, 2)$

Exponential and Power Parameters

$\alpha_a = TN(2, 1, 0, \infty)$

$\beta_a = TN(2, 1, 0, \infty)$

$\eta_a = TN(2, 1, 0, \infty)$

$\alpha_v = TN(2, 2, 0, \infty)$

$\beta_v = TN(2, 2, 0, \infty)$

$\eta_v = TN(0.5, 0.5, 0, \infty)$

Delay Parameters

$\tau_a = TN(20, 10, 0, \infty)$

$\tau_v = TN(20, 10, 0, \infty)$

Exponential Trial + Block Parameters

$d = TN(2, 2, 0, \infty)$

Step Parameters

$step = TN(0.05, 0.5, 0, \infty)$

$initial = TN(0.05, 0.5, 0, \infty)$

3 Generating Parameter Ranges for the Model Recovery

This supplementary material shows the parameter ranges that were selected for the recovery simulations described in Supplementary Material 2. Ranges for the standard diffusion model parameters were chosen based on typical values found in previous research (Matzke & Wagenmakers, 2009) and used in previous simulation studies (Alister et al., 2023). Ranges for the time-varying functions were chosen in order to capture the diverse forms the time-varying functions could take, while trying to avoid too much overlap between functions (e.g., under certain parameter values an exponential function can resemble a linear function), though there was still some overlap, which can explain some of the poor differentiability at extreme values in some of the time-varying functions.

Table 6: Generating parameter ranges for model recovery.

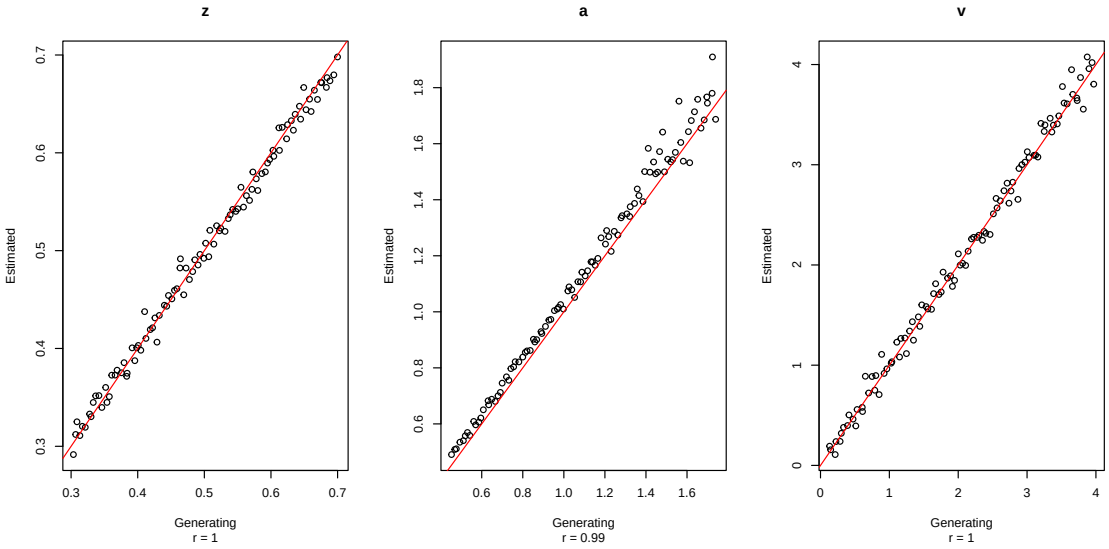
	Parameters							
Generating Model	z	$t0$	a	v	β	α	η	τ
*S*tandard DDM*	(0.3, 0.7)	(0.1, 0.6)	(0.45, 1.75)	(0.1, 4)	-	-	-	-
a Linear	(0.3, 0.7)	(0.1, 0.6)	-	(0.1, 4)	(0.0001, 0.003)	(0.5, 5)	-	-
v Linear	(0.3, 0.7)	(0.1, 0.6)	(0.1, 4)	-	(0.0001, 0.005)	(0.001, 3)	-	-
a Exponential		(0.1, 0.6)	-	(0.1, 4)	(1.5, 3)	(0.45, 2)	(0.001, 0.05)	-
v Exponential		(0.1, 0.6)	(0.45, 1.75)	-	(1.5, 3)	(0.45, 2)	(0.001, 0.05)	-
a Power		(0.1, 0.6)	-	(0.1, 4)	(1.5, 3)	(0.45, 2)	(0.001, 0.05)	-
v Power		(0.1, 0.6)	(0.45, 1.75)	-	(1.5, 3)	(0.45, 2)	(0.001, 0.05)	-
a Delayed Exp	(0.3, 0.7)	(0.1, 0.6)	-	(0.1, 4)	(1.5, 3)	(0.45, 2)	(0.001, 0.05)	(0.1, 30)
v Delayed Exp	(0.3, 0.7)	(0.1, 0.6)	(0.1, 4)	-	(1.5, 3)	(0.45, 2)	(0.001, 0.05)	(0.1, 30)

4 Correlation Between Generating Parameters and Estimated Parameters

This supplementary material presents the extent to which generating model parameters were estimated to be the same when fitting the different models. A high correlation between estimated and generated parameters means that that parameter has good measurement properties.

Standard Diffusion Model

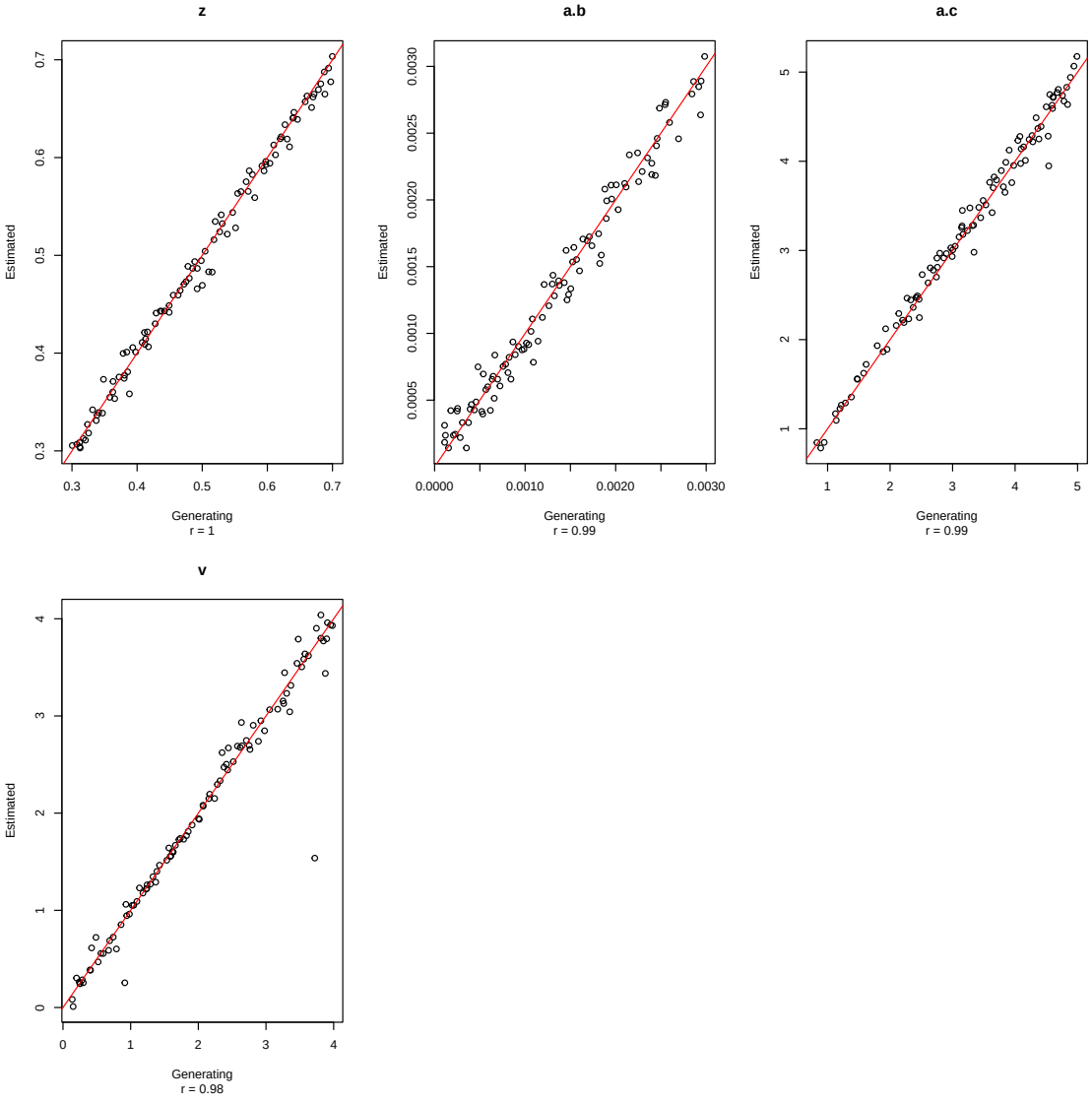
Figure 2. Correlation between generating parameters and estimated parameters for the standard diffusion model.



Red lines indicate a correlation of 1.

a Linear Trial-Varying Model

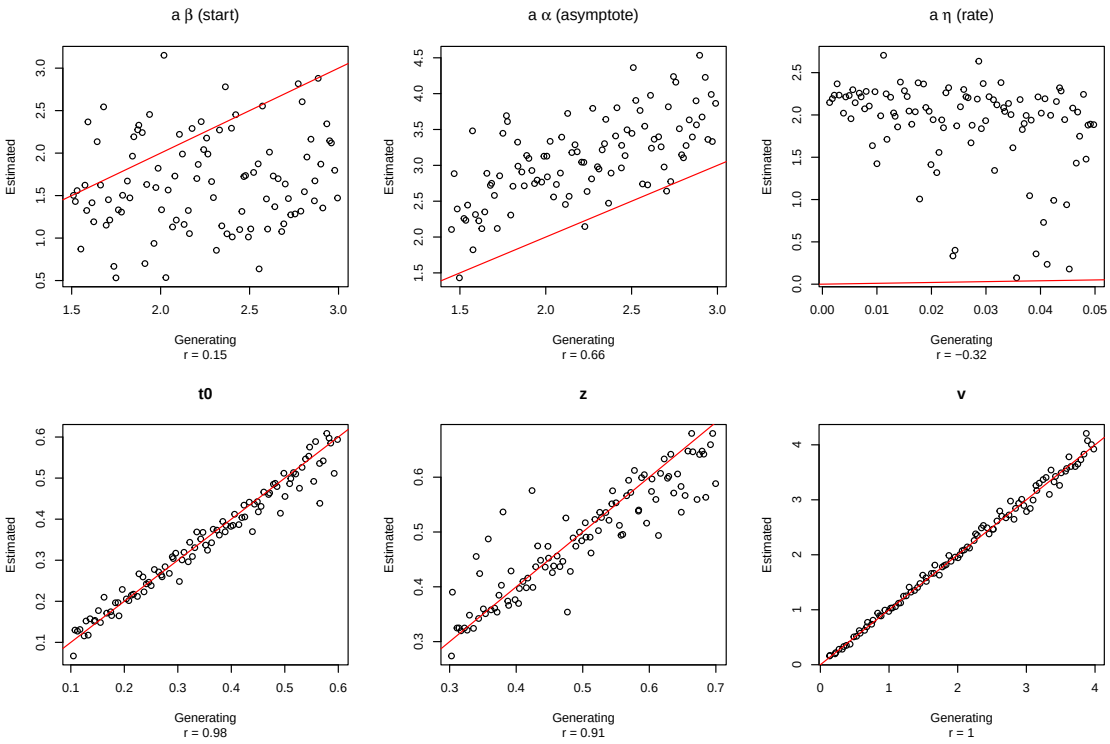
Figure 3. Correlation between generating parameters and estimated parameters for the *a*-linear trial-varying diffusion model.



Red lines indicate a correlation of 1.

a Power Trial-Varying Model

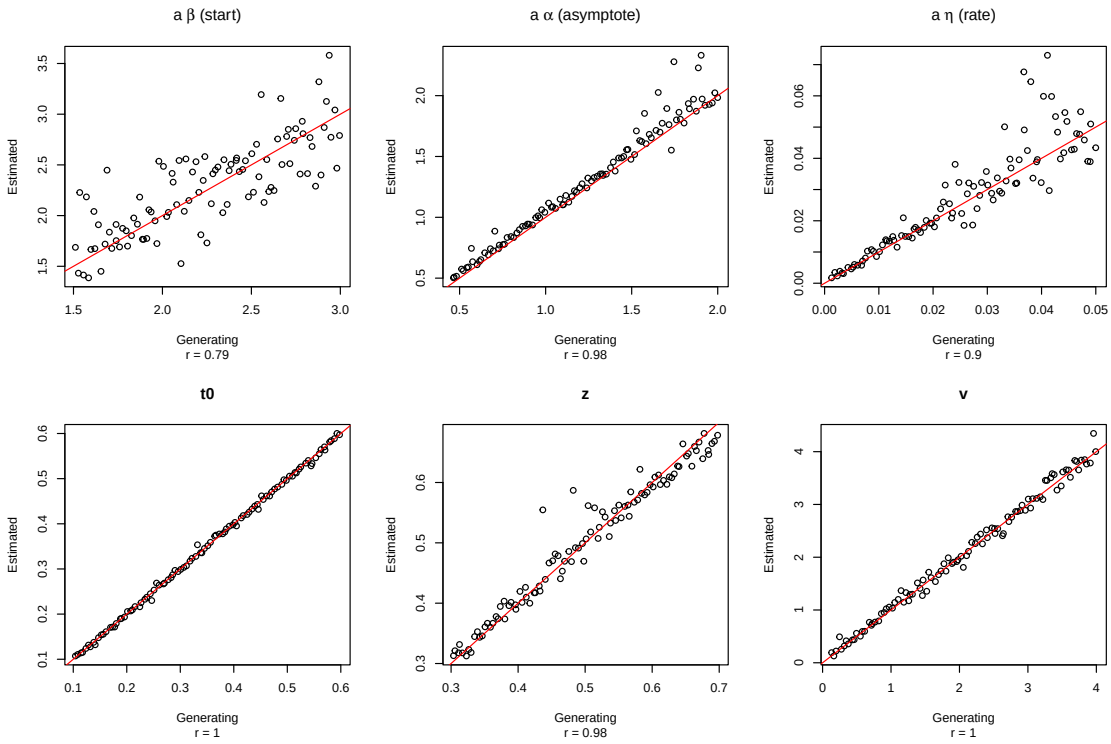
Figure 4. Correlation between generating parameters and estimated parameters for the *a* power trial-varying diffusion model.



Red lines indicate a correlation of 1.

a Exponential Trial-Varying Model

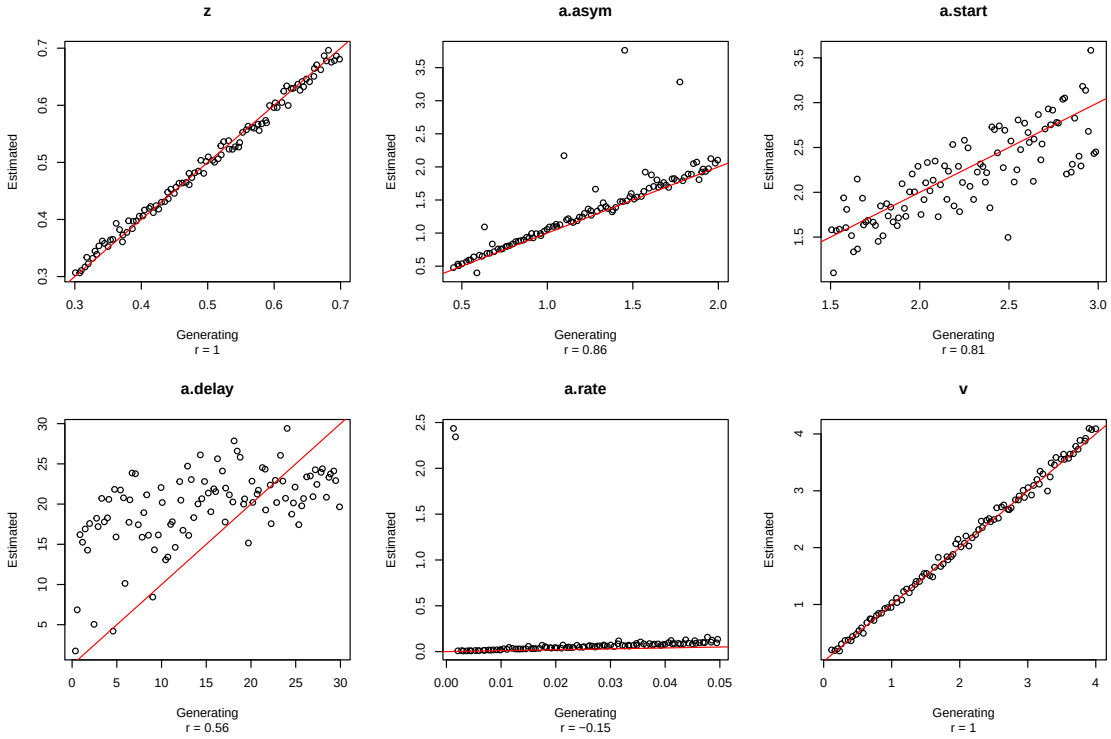
Figure 5. Correlation between generating parameters and estimated parameters for the *a* exponential trial-varying diffusion model.



Red lines indicate a correlation of 1.

a Delayed Exponential Trial-Varying Model

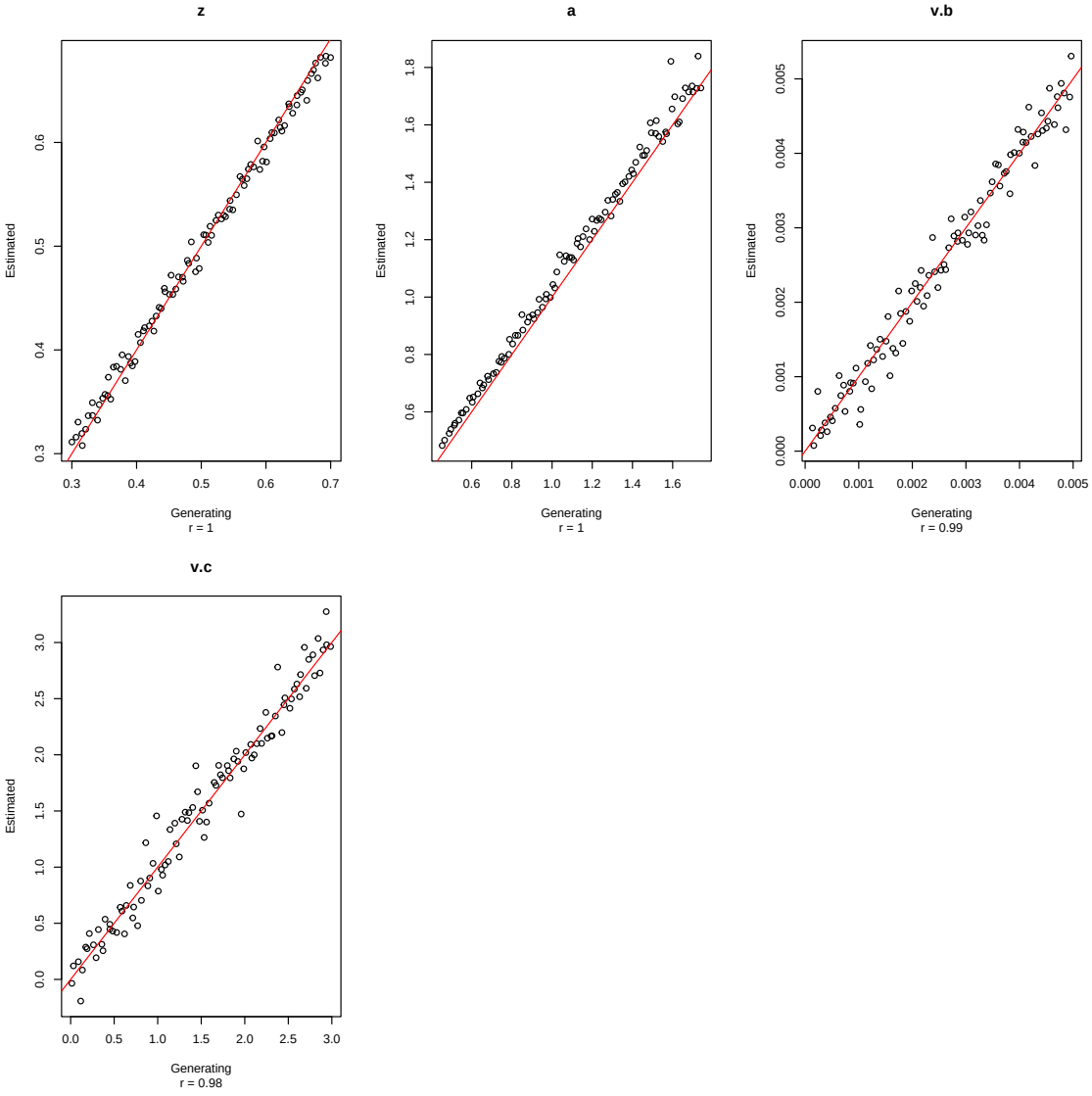
Figure 6. Correlation between generating parameters and estimated parameters for the *a* exponential trial-varying diffusion model.



Red lines indicate a correlation of 1.

v Linear Trial-Varying Model

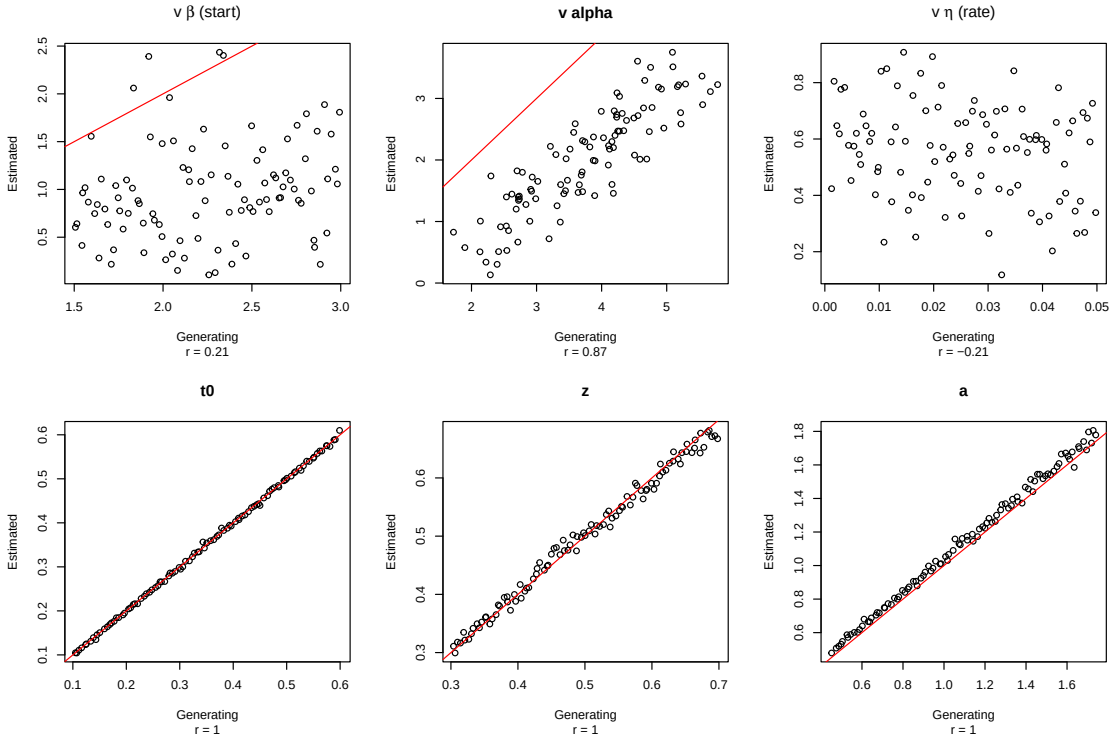
Figure 7. Correlation between generating parameters and estimated parameters for the *v* linear trial-varying diffusion model.



Red lines indicate a correlation of 1.

v Power Trial-Varying Model

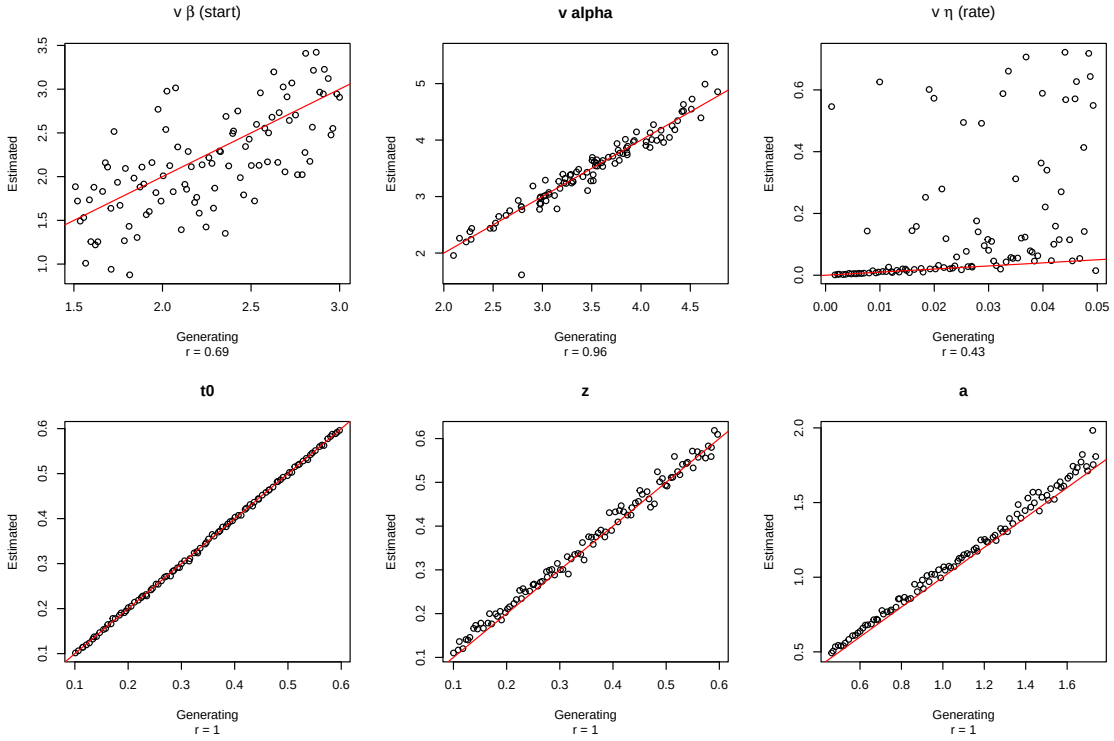
Figure 8. Correlation between generating parameters and estimated parameters for the *v* exponential trial-varying diffusion model.



Red lines indicate a correlation of 1.

v Exponential Trial-Varying Model

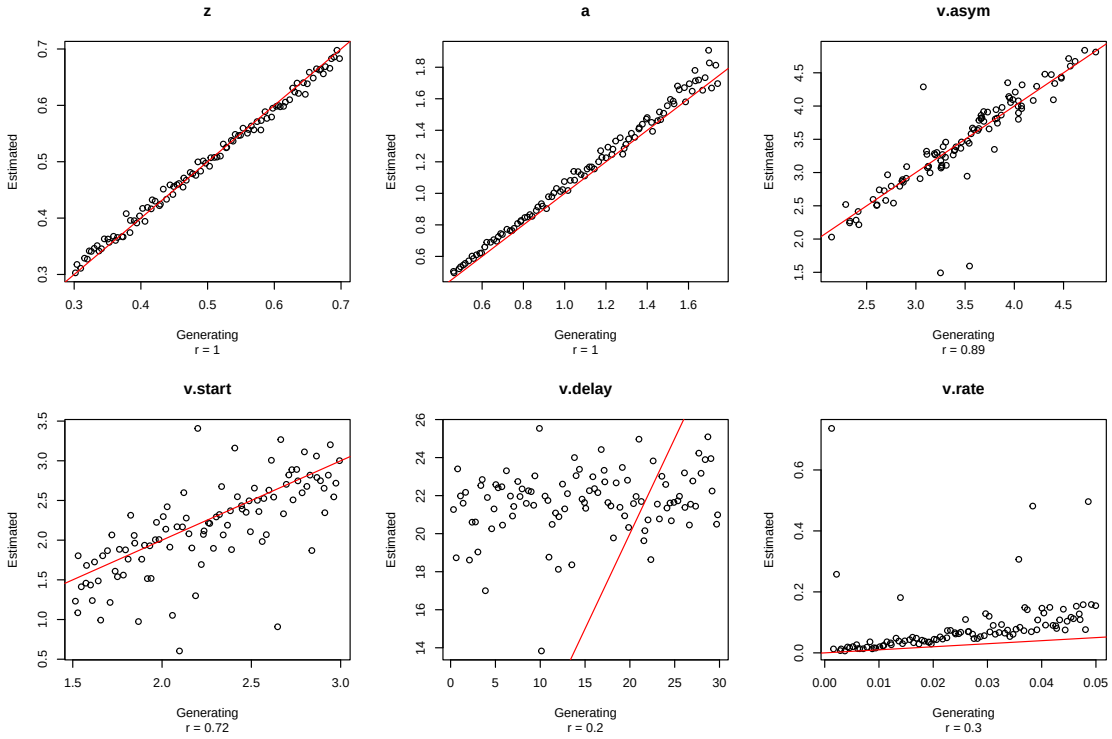
Figure 9. Correlation between generating parameters and estimated parameters for the *v* exponential trial-varying diffusion model.



Red lines indicate a correlation of 1.

v Delayed Exponential Trial-Varying Model

Figure 10. Correlation between generating parameters and estimated parameters for the *v* delayed exponential trial-varying diffusion model.



Red lines indicate a correlation of 1.

5 Can across trial variability explain ParAcT processes?

In our study, we used a simple version of the Standard Diffusion Decision Model (DDM), which only included the main cognitive parameters (*v, a, z, t0*). The full version of the diffusion model includes extra parameters that assume random variability across trials for three of these parameters (*v, z, t0*). We did not include these parameters because 1) they often have poor measurement properties, 2) numerical integration required to implement them would make computation time in our ParAcT framework infeasible as every trial has a different set of parameters, and 3) although these parameters help improve the fit of the DDM, they do not allow for any theoretical inferences about the source of the random variability nor predictions about systematic variability.

Table 7: Probability according to BIC (AIC) that the drift rate (v) exponential model and Standard DDM + v across-trial variability model were the best fit to data generated from each model respectively, collapsed across participants.

		Fitted Model	
Generating Model	n Best Fit	Standard DDM + v Variability	v Exponential
Standard DDM + v Variability	100 (100)	**1 (0.98)**	<.01 (0.02)
v Exponential	0.92 (0.97)	0.08 (0.03)	**94 (97)**

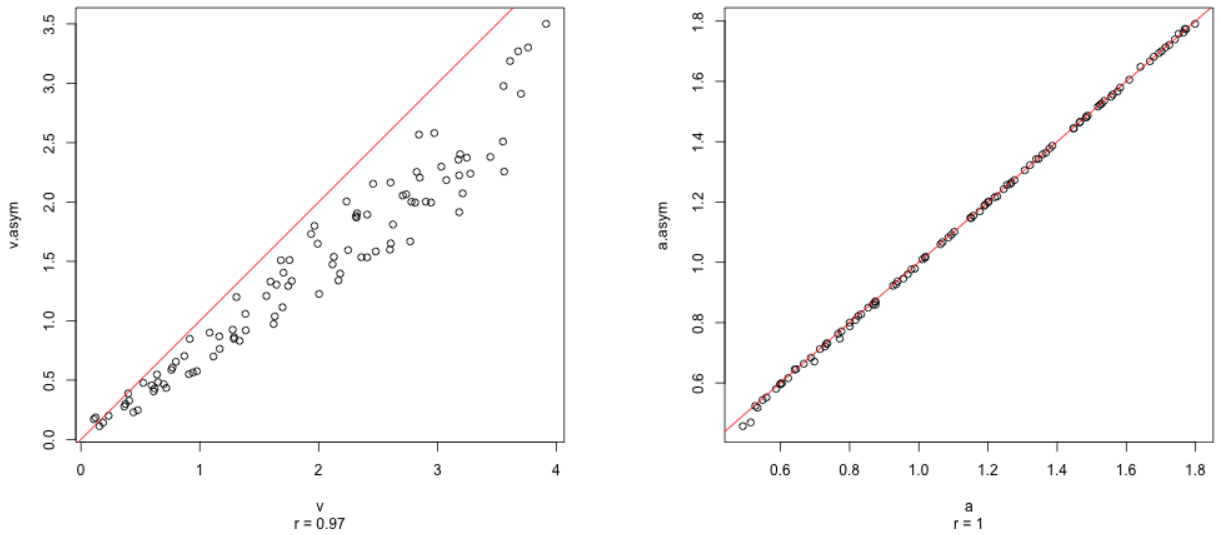
However, it is possible that the standard DDM would still be able to account for ParAct processes had it included across-trial variability parameters. To check whether this explanation was plausible, we ran model recovery simulations where we generated data with across-trial variability models and ParAcT models respectively, then fit both types of models to each data set, to see whether the generating model performed best.

Table 8: Probability according to BIC (AIC) that the drift rate (a) exponential model and standard DDM + z across-trial variability model were the best fit to data generated from each model respectively, collapsed across participants.

		Fitted Model	
Generating Model	n Best Fit	Standard DDM + z Variability	a Exponential
Standard DDM + z Variability	99 (97)	**0.97 (0.79)**	0.03 (0.21)
a Exponential	100 (100)	<.01 (<.01)	**1 (1)**

Specifically, since exponential ParAct models seemed to perform best in our empirical analyses (and have good measurement properties), we compared v-Exponential trial-varying model to the standard DDM that also included random across trial variability for v, and the a-exponential

Figure 11. Correlation between generating parameters and estimated parameters between-trial variability parameters and their asymptote parameter of the equivalent exponential ParAcT model.



(a) Data generated from the standard DDM with across-trial variability for v

(b) Data generated from the standard DDM with across-trial variability for z

Red lines indicate a correlation of 1.

trial-varying model to the standard DDM that also included random across trial variability for z[2] .

Tables 7 and 8 suggest that systematic, exponential changes in a and v as generated by the ParAcT models cannot be accounted for by the standard DDM that includes the relevant across-trial variability parameters, nor can the ParAcT processes account for random-variability. These simulations therefore suggest that random, across-trial variability parameters and ParAct models are explaining different processes.

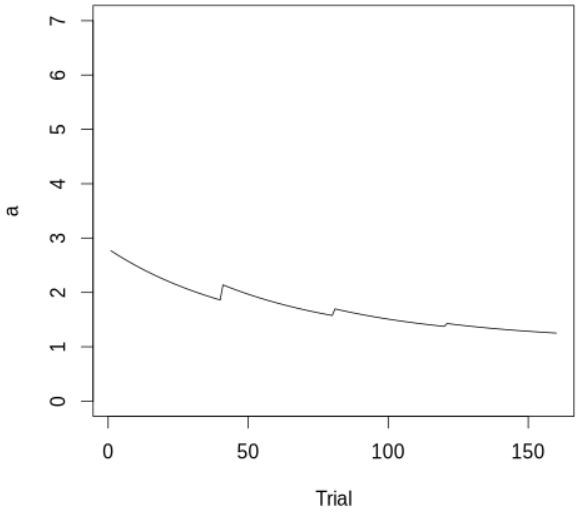
Another question is whether the presence of between-trial variability could affect the measurement properties of the ParAcT functions. To test this, we correlated the asymptote parameters of v and a in the ParAcT models respectively with the standard v and a parameters in the between-trial variability models. The logic behind this analysis was that the asymptote parameters should align with the standard DDM parameters when there are no systematic time-varying effects. As is shown in Figure 12, estimation of the asymptote parameter remained unaffected for the most part, with a having perfect identifiability, whereas v had a very high correlation, but the v-exponential model tended to systematically underestimate the true value of v as v increased.

[2]There is no across-trial variability for a, however the z across trial variability corresponds to a, since z relates to the difference between the two decision thresholds.

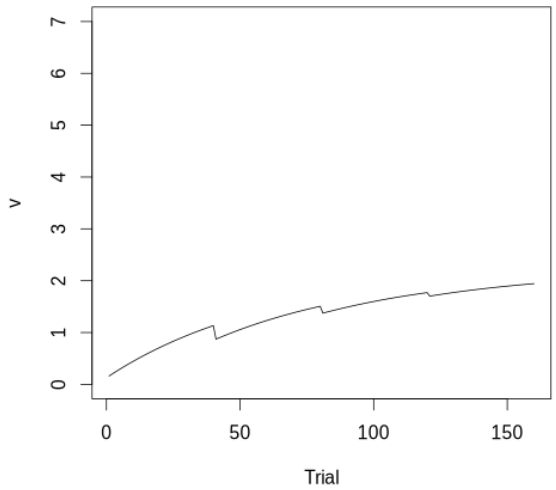
6 Visualisation of Exponential Trial change with Block Bump Model

Here, we show a visualisation of exponential trial change with a block bump model for both drift rate and threshold. Importantly, the formula for these models may seem somewhat counterintuitive, as the 'bump' parameter (d) appears to provide a constant regression to the start over blocks, which might suggest that any learning occurring within blocks is quickly wiped out at the start of the new block. However, as the 'bump' term is added to the parameter for the difference between the initial value and the asymptote (β), which decays over time based on the learning rate, the 'bump' is quickly overwhelmed by learning, causing it to have a lessening influence as greater learning occurs, and for it to do little to prevent the function from reaching its asymptote.

Figure 12. Simulated trajectory of the "exponential trial change with block bump" function.



(a) Threshold (a). Parameter values: $d_a = 0.67$, $\alpha_a = 1.12$ $\beta_a = 1.69$ $\eta_a = 0.02$



(b) Drift Rate (v). Parameter values: $d_v = 0.55$, $\alpha_v = 0.13$ $\beta_v = 2.08$ $\eta_v = .02$

7 Alternative parameter trajectory plots

In Figures 3, 7, 9, & 11 of the main manuscript, we presented the trajectory of the median parameter estimates for the most common functions for each parameter. While these plots are useful for seeing how the parameters changed across time, taking the median of the parameter estimates can appear differently to any one single estimate, as well as if one were to take the central tendency of multiple individual trajectories.

In this supplementary material, we present alternate parameter trajectory plots that show each individual trajectory (thin, light lines) and the median central tendency of those individual trajectories (thick black lines), as opposed to the median of the parameter estimates. The points are still the same as the original plots: the median block-level estimates.

Figure 13. Data set 1.

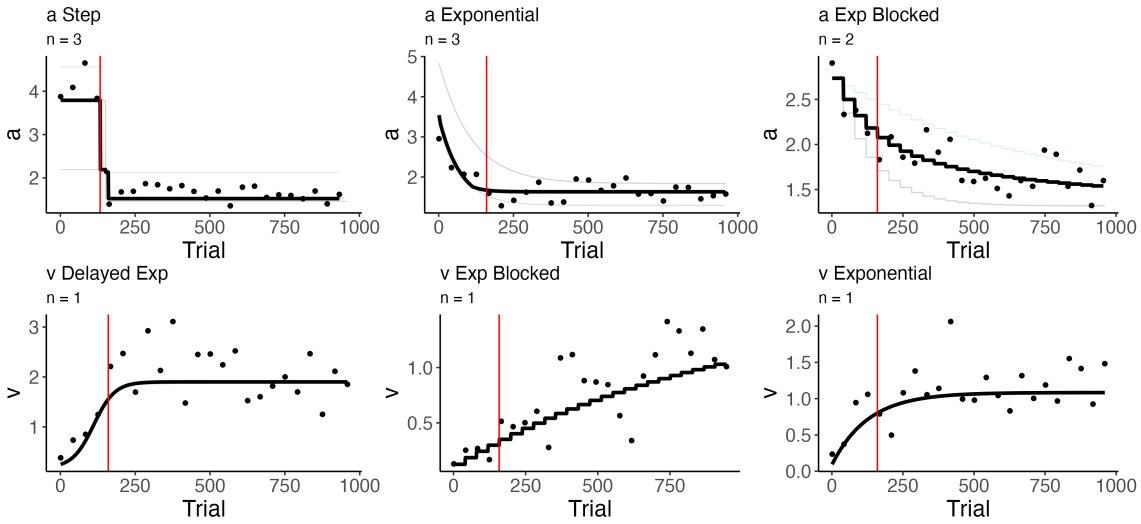


Figure 14. Data set 2.

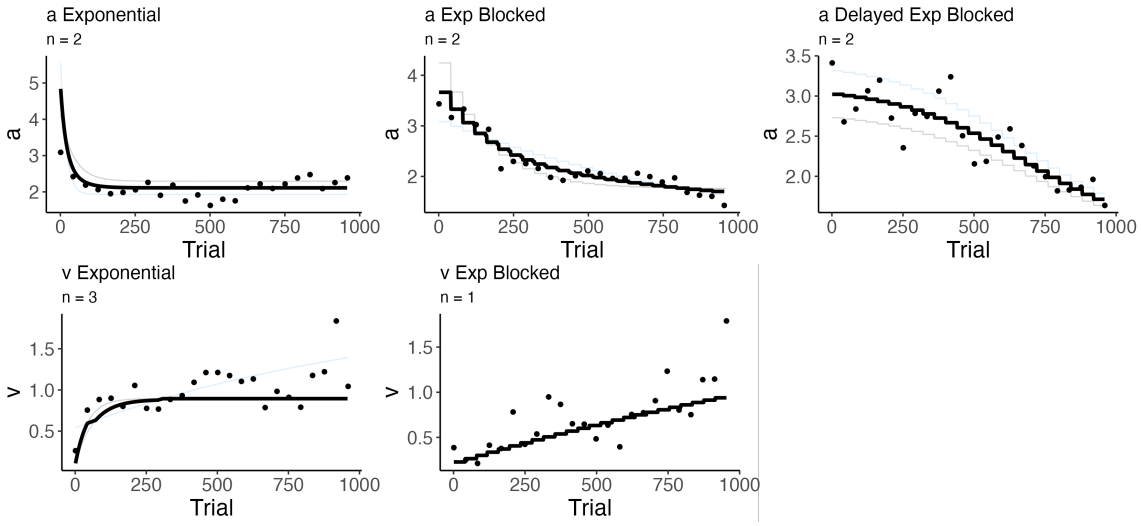


Figure 15. Data set 3.

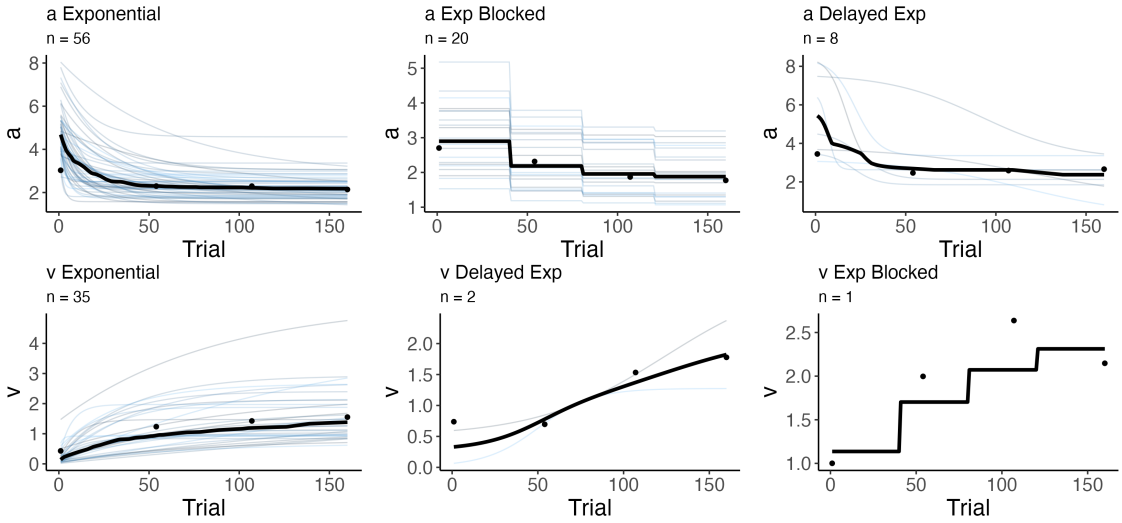


Figure 16. Data set 4.