



Extending TVA With Trial-to-Trial Variability in Processing Speed: A Monte Carlo Simulation Study

Aaron Khodami^{1,2}

Received: 20 April 2026 / Accepted: 17 September 2026
© The Author(s) 2026

Abstract

The Theory of Visual Attention (TVA) typically treats processing speed as fixed within the modelled condition structure. This assumption may be too restrictive when effective processing speed varies from trial to trial. We introduce a variability-augmented extension of a behavioural TVA count model in which trial-to-trial fluctuations in processing speed are represented explicitly through a gamma-distributed speed parameter, while preserving the classical interpretation of processing speed (C), visual short-term memory capacity (K), and perceptual threshold (t_0). A Monte Carlo simulation study evaluated both models, together with the trial-to-trial t_0 -variability model of Dyrholm et al. (2011), across whole-report scenarios with negligible, mild, moderate, and strong trial-to-trial variability in processing speed, and with variability in t_0 . The extended model improved model selection and parameter recovery under strong variability, with the advantage emerging when variability was large enough to induce systematic bias in the fixed-speed model's estimates of C and t_0 ; under moderate variability the evidence was transitional, and with fewer trials per dataset the precision lost to the additional parameter could outweigh the bias it removed. The two forms of variability left different traces in the data and the two variability models were distinguishable: the t_0 -variability model did not register speed variability and left the bias in C in place, whereas the extended model absorbed only part of the threshold variability. These findings suggest that the proposed extension is best understood as a conditional complement to standard behavioural TVA, appropriate when notable trial-to-trial variability in processing speed is present, rather than as a universal replacement.

Keywords Theory of visual attention · Processing speed · Trial-to-trial variability · Whole report · Model selection

Introduction

Bundesen's theory of visual attention (TVA) (Bundesen, 1990) is one of the most influential computational frameworks for the analysis of visual selection and short-term storage in behavioural data (Bundesen et al., 2005, 2015; Bundesen, 1998). In contrast to verbal or purely descriptive theories of attention, TVA provides an explicit quantitative account of how objects compete for representation under capacity limitations (Bundesen & Habekost, 2014; Vangkilde et al., 2013). This has made the framework especially

valuable in experimental settings in which the goal is not only to describe performance, but also to estimate theoretically meaningful latent parameters from observed behaviour. In practice, TVA has been particularly important because it links formal assumptions about attentional selection to measurable outcomes in whole-report and partial-report tasks, bridging theory, experimental design, and statistical estimation (Bundesen, 1990; Bundesen et al., 2015; Habekost & Starrfelt, 2009; Kyllingsbæk, 2006).

At the core of TVA is the idea that visual processing can be described as a race among categorisations, with the rate of each categorisation determined by the joint influence of sensory evidence, attentional weighting, and categorical bias (Bundesen, 1990). In standard behavioural applications, this framework gives rise to a set of parameters that have become central in the TVA literature, including processing speed (C), visual short-term memory capacity (K), and the perceptual threshold (t_0), with additional parameters estimated under specific task designs such as selectivity in

✉ Aaron Khodami
aaron.a.khodami@jyu.fi

¹ Faculty of Information Technology, University of Jyväskylä, Jyväskylä, Finland

² Department of General Psychology, University of Padua, Padua, Italy

partial report (Bundesen & Habekost, 2014; Bundesen et al., 2005; Kyllingsbæk, 2006). This parameterisation is one of the main strengths of TVA. Rather than reducing performance to a single summary score, it allows researchers to separate distinct components of attentional efficiency and storage and ask which part of the system changes across observers, tasks, or experimental manipulations (Bundesen, 1990; Chechlacz et al., 2015; Vangkilde et al., 2013).

This quantitative structure is also the reason TVA has remained useful across a broad range of behavioural contexts. Whole-report and partial-report paradigms have been used to study normal visual attention, individual differences, and attentional deficits, and the formal estimation framework has been refined to make such analyses more accessible and robust (Bundesen et al., 2005; Habekost, 2015; Kyllingsbæk, 2006). More generally, TVA has had enduring value because it offers a combination of theoretical interpretability and empirical tractability. Researchers can fit the model to trial-based or condition-based performance data, obtain psychologically interpretable parameter estimates, and compare competing explanations within a coherent likelihood-based framework (Dyrholm et al., 2011). For that reason, TVA remains one of the behavioural attention models that can support both substantive theoretical claims and explicit model-based inference.

At the same time, the strengths of TVA do not imply that every standard implementation is equally adequate for every data-generating process. In common behavioural applications, the fitted model is often effectively treated as if processing speed were fixed within the modelled condition structure, apart from random sampling variability (Bundesen, 1990; Kyllingsbæk, 2006). That assumption may be reasonable when the underlying process is relatively stable, but it becomes more restrictive when effective processing speed fluctuates from trial to trial. In such cases, a fixed-speed model may still produce acceptable descriptive fits, yet the latent variability may be absorbed indirectly into shifts in estimated C , K , or t_0 rather than being represented explicitly. Earlier methodological work has already pointed to the relevance of parameter variability in TVA-related estimation, suggesting that trial-to-trial heterogeneity is not a technical nuisance but a potentially important part of the behavioural process (Dyrholm et al., 2011; Habekost, 2015). The most developed example is the generalisation proposed by Dyrholm et al. (2011), in which the perceptual threshold t_0 varies from trial to trial. This formulation has become the standard choice in current TVA-based analyses of whole report, so that the fixed-parameter model is now rarely fitted on its own. Variability in processing speed has received less attention in whole-report analyses, but evidence suggests speed is not constant within a session. Dyrholm et al. (2015) modelled processing speed on single

trials in a single-item recognition task and found systematic variation over the course of the experiment, and Blurton et al. (2020) described trial-to-trial variation in processing rates with a gamma distribution in a TVA-based model of single-item recognition. Since the processing capacity C is the sum of the item-level processing rates v , variability in C can be understood as a common trial-level variation shared by all these rates. To our knowledge, this form of variability has not yet been built into the model that is used for the whole report. The present work, therefore, is not whether the standard TVA model is generally valid, but whether its usual fixed-parameter implementation may become too rigid under conditions of substantial trial-to-trial variability in processing speed.

The simulations developed in the present work are motivated by that specific problem. Our starting point is the possibility that a participant's effective processing speed may vary across trials even when the nominal task condition is unchanged. If such variability is present, then a model that assumes a single fixed processing-speed parameter may be misspecified, especially when the variability is moderate to strong. Conversely, when variability is negligible, the standard model should remain adequate and should not be displaced by the addition of extra flexibility. This leads to a clear methodological question: can an extension of the standard behavioural TVA model that accommodates trial-to-trial variability in processing speed improve model fit and parameter recovery when such variability is present, while preserving the adequacy of the classical formulation when it is not? A second question follows from the existence of the t_0 -variability model. Both formulations introduce trial-to-trial variability in the amount of processing that takes place during the exposure, and it is not obvious from the outset whether whole-report data can tell them apart. We therefore also ask whether the two forms of variability leave different traces in the data, whether a model that allows only t_0 to vary can absorb variability that actually resides in processing speed, and what happens to the estimates of C , K , and t_0 when the wrong form of variability is assumed.

The aim of the present paper is to address that question by developing and evaluating a variability-augmented extension of the standard behavioural TVA count model. Our goal is not to replace the classical TVA framework or to claim that the standard model is broadly deficient. Rather, the aim is to identify a specific regime in which the usual fixed-speed implementation may be too restrictive and to test whether an explicit variability term improves inference under those conditions. To do so, we compare the standard and extended models in a Monte Carlo framework that examines model selection and parameter recovery across scenarios with no variability, mild variability, moderate variability, and strong variability in processing speed. The t_0 -variability model

of Dyrholm et al. (2011) is fitted to the same data, and the design includes scenarios in which the generating process contains variability in t_0 rather than in processing speed, so that each model is evaluated on data produced by the other. Part of the design is also repeated with fewer trials. This last point matters because an additional variability parameter is not free. In the diffusion model literature, across-trial variability parameters were introduced for good theoretical reasons, and later work showed that data sets of ordinary size often fail to constrain them Boehm et al. (2018). Whether an extension of this kind is worthwhile therefore depends on whether the data can support it. The central hypothesis is deliberately narrow: the extended model should show an advantage when trial-to-trial variability in processing speed is sufficiently pronounced, whereas the standard model should remain appropriate when the generating process is effectively fixed-speed. For the comparison between the two variability models, the expectation is that each should be preferred on the type of variability it was built for, and that neither should be able to stand in for the other.

Method and Model

The present model is formulated for the behavioural design introduced above, in which participants are shown brief arrays of letters and are asked to report as many letters as they can correctly remember. In the current implementation, the task is a whole-report task without distractors, run at two display sizes, six letters and four letters. Accordingly, the present model targets the core behavioural TVA quantities associated with whole-report performance, namely processing speed, storage capacity, and perceptual threshold, while parameters that require different designs, such as selectivity in partial report with distractors or persistence under unmasked conditions, are not estimated here Bundesen (1990); Kyllingsbæk (2006). The model is therefore a count-based TVA implementation for whole-report data, extended to allow trial-to-trial variability in processing speed. In what follows, the fixed-speed formulation is called the old model, the extension with variability in processing speed is called the new model, and a third formulation, the trial-to-trial t_0 -variability model of Dyrholm et al. (2011), is specified at the end of this section as the second comparison model. Throughout, cv_C denotes the coefficient of variation of the trial-level processing speed, the realised value of the trial-level speed is written u , and the shape parameter of the gamma distribution is written η , so that none of these symbols is confused with the item-level processing rate v .

Let participant $i \in \{1, \dots, N\}$ complete trials indexed by $j \in \{1, \dots, T_i\}$. On each trial, the observed data are

$$D_{ij} = (n_{ij}, t_{ij}, y_{ij}), \quad (1)$$

where n_{ij} is the number of letters shown on trial j for participant i , t_{ij} is the exposure duration in seconds, and y_{ij} is the number of letters correctly reported. In the present design, $n_{ij} \in \{4, 6\}$. The response variable y_{ij} is the behavioural score obtained by comparing the participant's report with the set of letters presented on that trial, so that $y_{ij} \in \{0, \dots, n_{ij}\}$. If no letter is correctly reported, then $y_{ij} = 0$; if all shown letters are correctly reported, then $y_{ij} = n_{ij}$. The model treats this score as the observed outcome to be explained.

The starting point is the classical TVA idea that visual categorisations proceed in parallel and that performance is constrained by both processing speed and limited storage capacity (Bundesen, 1990). For the present whole-report task, we use a count formulation in which each displayed letter is assumed to be processed at the same average rate within a given trial. Let $C > 0$ denote the participant-level processing speed parameter, measured in letters per second; let $K > 0$ denote visual short-term memory capacity; and let $t_0 \geq 0$ denote the perceptual threshold, that is, the time before effective encoding begins. The effective exposure available for encoding on trial (i, j) is defined by

$$\tau_{ij} = \max(0, t_{ij} - t_0). \quad (2)$$

As shown in Eq. 2, τ_{ij} is zero whenever the physical exposure duration does not exceed the threshold t_0 , and otherwise equals the part of the exposure that is available for effective encoding.

Under the old fixed-speed model, the total processing speed C is distributed evenly across the n_{ij} displayed letters. The per-letter processing rate on trial (i, j) is therefore

$$v_{ij} = \frac{C}{n_{ij}}. \quad (3)$$

Given the exponential-race assumption, the probability that a particular letter finishes processing before the effective exposure ends is

$$p_{ij} = 1 - \exp(-v_{ij}\tau_{ij}) = 1 - \exp\left(-\frac{C}{n_{ij}}\tau_{ij}\right). \quad (4)$$

As shown in Eq. 4, completion probability increases with effective exposure τ_{ij} and with processing speed C , and decreases with display size n_{ij} because the available processing speed is shared across more letters.

Let M_{ij} denote the number of letters that complete encoding before the exposure ends, prior to any storage truncation. Because all letters are treated symmetrically within the

present task, and because each letter has completion probability p_{ij} as given in Eq. 4, the completed-count variable follows a binomial distribution,

$$M_{ij} \mid n_{ij}, t_{ij}, C, t_0 \sim \text{Binomial}(n_{ij}, p_{ij}). \quad (5)$$

Equivalently, the probability of m completed letters is

$$\Pr(M_{ij} = m \mid n_{ij}, t_{ij}, C, t_0) = \binom{n_{ij}}{m} p_{ij}^m (1 - p_{ij})^{n_{ij}-m}, \quad (6)$$

for $m = 0, \dots, n_{ij}$.

To connect the completed-count variable to the observed score, the model assumes that only the first K completed letters can be retained. Following standard TVA practice, non-integer estimates of K are interpreted as adjacent-integer mixtures rather than as literal fractional storage slots (Bundesen, 1990; Kyllingsbæk, 2006).

$$K = k_0 + \phi, \quad k_0 = \lfloor K \rfloor, \quad \phi = K - k_0, \quad 0 \leq \phi < 1. \quad (7)$$

The trial-specific effective capacity K_{ij}^* is then defined as

$$K_{ij}^* = \begin{cases} k_0, & \text{with probability } 1 - \phi, \\ k_0 + 1, & \text{with probability } \phi. \end{cases} \quad (8)$$

The observed score Y_{ij} is the number of correctly encoded letters that survive this storage limitation. Given the completed-count variable M_{ij} and the effective capacity K_{ij}^* , the score is

$$Y_{ij} = \min(M_{ij}, K_{ij}^*). \quad (9)$$

For a fixed integer capacity value k , the score distribution is therefore

$$\Pr(Y_{ij} = y \mid n_{ij}, t_{ij}, C, t_0, K_{ij}^* = k) = \begin{cases} \Pr(M_{ij} = y \mid \dots), & 0 \leq y < k, \\ \sum_{m=k}^{n_{ij}} \Pr(M_{ij} = m \mid \dots), & y = k, \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

For the continuous parameter K , the final score distribution is obtained by averaging over the two adjacent capacities in Eq. 8,

$$\Pr(Y_{ij} = y \mid n_{ij}, t_{ij}, C, K, t_0) = (1 - \phi) \Pr(Y_{ij} = y \mid K_{ij}^* = k_0) + \phi \Pr(Y_{ij} = y \mid K_{ij}^* = k_0 + 1). \quad (11)$$

The new model relaxes the assumption that processing speed is constant across trials. Instead, it assumes that each trial has its own latent processing speed, denoted C_{ij} , drawn from a gamma distribution with mean C and

coefficient of variation cv_C . We parameterize the gamma distribution by shape η and scale ω such that

$$\eta = \frac{1}{\text{cv}_C^2}, \quad \omega = C \text{cv}_C^2. \quad (12)$$

Under this parameterization, $C_{ij} \sim \text{Gamma}(\eta, \omega)$.

The gamma distribution was chosen for three reasons. First, it is a natural distribution for a positive rate parameter and has been used before for trial-to-trial variation of processing rates in a TVA-based model of single-item recognition (Blurton et al., 2020). Second, its Laplace transform has a closed form, so the marginal distribution of the completed count can be written without numerical integration, as shown below. Third, the mean and the coefficient of variation determine the distribution completely, so that variability is captured by a single additional parameter. Other positive distributions with a closed-form Laplace transform, such as the inverse Gaussian, could be substituted in the same derivation; a log-normal distribution has no closed-form Laplace transform and would require numerical integration at every evaluation of the likelihood.

Conditional on a specific trial speed value $C_{ij} = u$, the per-letter completion probability on trial (i, j) becomes

$$p_{ij}(u) = 1 - \exp\left(-\frac{u}{n_{ij}} \tau_{ij}\right). \quad (13)$$

To simplify notation, define $a_{ij} = \frac{\tau_{ij}}{n_{ij}}$. Using this definition, the conditional completed-count probability is

$$\Pr(M_{ij} = m \mid C_{ij} = u) = \binom{n_{ij}}{m} (1 - e^{-a_{ij}u})^m e^{-a_{ij}(n_{ij}-m)u}. \quad (14)$$

The marginal completed-count probability is then obtained by integrating over the gamma density:

$$\Pr(M_{ij} = m \mid n_{ij}, t_{ij}, C, t_0, \text{cv}_C) = \binom{n_{ij}}{m} \int_0^\infty (1 - e^{-a_{ij}u})^m e^{-a_{ij}(n_{ij}-m)u} f_{\Gamma}(u; \eta, \omega) du. \quad (15)$$

Expand the term $(1 - e^{-a_{ij}u})^m$ by the binomial theorem:

$$(1 - e^{-a_{ij}u})^m = \sum_{r=0}^m (-1)^r \binom{m}{r} e^{-a_{ij}ru}. \quad (16)$$

Substituting equation 16 into Eq. 15 yields

$$\Pr(M_{ij} = m \mid \dots) = \binom{n_{ij}}{m} \sum_{r=0}^m (-1)^r \binom{m}{r} \int_0^\infty e^{-a_{ij}(n_{ij}-m+r)u} f_{\Gamma}(u; \eta, \omega) du. \quad (17)$$

The remaining integral is the Laplace transform of a gamma distribution. For any $s \geq 0$, and with ω treated as the scale parameter (so that the gamma density is proportional to $u^{\eta-1}e^{-u/\omega}$),

$$\int_0^\infty e^{-su} f_{\Gamma}(u; \eta, \omega) du = (1 + \omega s)^{-\eta}. \quad (18)$$

As $cv_C \rightarrow 0$, the right-hand side of Eq. 18 approaches e^{-sC} , so that the old model is nested in the new model as the boundary case $cv_C = 0$. In the implementation, the right-hand side is evaluated as $\exp[-\eta \log(1 + \omega s)]$ with a numerically stable logarithm of $1 + \omega s$, which keeps the expression accurate for very small values of cv_C , where ωs is close to zero and η is very large.

Applying Eq. 18 gives the completed-count distribution for the new model:

$$\begin{aligned} \Pr(M_{ij} = m \mid n_{ij}, t_{ij}, C, t_0, cv_C) \\ = \binom{n_{ij}}{m} \sum_{r=0}^m (-1)^r \binom{m}{r} [1 + \omega a_{ij}(n_{ij} - m + r)]^{-\eta}, \end{aligned} \quad (19)$$

for $m = 0, \dots, n_{ij}$. Equation 19 is the central mathematical extension proposed in the present work.

The score distribution for the new model follows the same mixing logic:

$$\begin{aligned} \Pr(Y_{ij} = y \mid n_{ij}, t_{ij}, C, K, t_0, cv_C) &= (1 - \phi) \Pr(Y_{ij} = y \mid K_{ij}^* = k_0) \\ &+ \phi \Pr(Y_{ij} = y \mid K_{ij}^* = k_0 + 1). \end{aligned} \quad (20)$$

The t_0 -variability model keeps processing speed fixed at C and instead lets the perceptual threshold vary from trial to trial. Following (Dyrholm et al., 2011), the trial-level threshold $t_{0,ij}$ is drawn from a normal distribution with mean μ_{t_0} and standard deviation σ_{t_0} , and the effective exposure on trial (i, j) becomes

$$\tau_{ij}(t_{0,ij}) = \max(0, t_{ij} - t_{0,ij}), \quad t_{0,ij} \sim \text{Normal}(\mu_{t_0}, \sigma_{t_0}^2). \quad (21)$$

Conditional on $t_{0,ij}$, the completed count follows Eq. 6 with τ_{ij} replaced by $\tau_{ij}(t_{0,ij})$, and the marginal completed-count distribution is

$$\begin{aligned} \Pr(M_{ij} = m \mid n_{ij}, t_{ij}, C, \mu_{t_0}, \sigma_{t_0}) &= \int_{-\infty}^{\infty} \binom{n_{ij}}{m} p_{ij}(z)^m (1 - p_{ij}(z))^{n_{ij}-m} \\ &\times f_{\mathcal{N}}(z; \mu_{t_0}, \sigma_{t_0}) dz, \\ p_{ij}(z) &= 1 - \exp\left(-\frac{C}{n_{ij}} \max(0, t_{ij} - z)\right). \end{aligned} \quad (22)$$

This integral has no closed form and is evaluated by Gauss–Hermite quadrature with 48 nodes, which reproduces the

distribution obtained by direct simulation of the generating process to within Monte Carlo error. The score distribution is obtained from Eq. 22 by the same capacity mixing as in Eqs. 10 and 11. The model has the parameters $(C, K, \mu_{t_0}, \sigma_{t_0})$, and the old model is nested in it as the boundary case $\sigma_{t_0} = 0$. The same distributional form is used when data are generated under t_0 variability and when the model is fitted, so that draws of $t_{0,ij}$ below zero, which the normal distribution allows, are treated identically in both steps.

The corresponding log-likelihood is

$$\ell_{\text{new}}(\theta_{\text{new}}) = \sum_{i=1}^N \sum_{j=1}^{T_i} \log \Pr(Y_{ij} = y_{ij} \mid \theta_{\text{new}}). \quad (23)$$

The same expression defines the log-likelihood of the old model with $\theta_{\text{old}} = (C, K, t_0)$ and of the t_0 -variability model with $\theta_{t_0} = (C, K, \mu_{t_0}, \sigma_{t_0})$, with the score probability taken from Eq. 11 or from the t_0 -variability analogue of Eq. 20, respectively.

For computational efficiency, trials can be collapsed into unique cells:

$$\ell(\theta) = \sum_{g=1}^G w_g \log \Pr(Y = y_g \mid n_g, t_g, \theta), \quad (24)$$

where a cell g is a combination of display size, exposure duration, and score, w_g is the number of trials in that cell, and G is the number of cells.

Model selection is performed using AIC and BIC:

$$\text{AIC}_m = 2p_m - 2\ell_m(\hat{\theta}_m), \quad \text{BIC}_m = p_m \log(N_{\text{trials}}) - 2\ell_m(\hat{\theta}_m). \quad (25)$$

Here p_m is the number of free parameters, three for the old model and four for each of the two variability models, and N_{trials} is the number of trials in the data set. Because the two variability models have the same number of parameters, AIC and BIC lead to the same decision when comparing them. When either variability model is compared with the old model, the two criteria differ by a constant, $\Delta\text{BIC} = \Delta\text{AIC} + \log(N_{\text{trials}}) - 2$. A further consequence of the nesting is that the null values $cv_C = 0$ and $\sigma_{t_0} = 0$ lie on the boundary of the parameter space. Under a fixed-speed truth, the likelihood-ratio statistic between a variability model and the old model therefore does not follow a χ^2 distribution with one degree of freedom but a 50:50 mixture of a point mass at zero and that distribution (Self & Liang, 1987). This result is used in the Results section to derive the rate at which AIC and BIC are expected to prefer a variability model when the data contain no variability.

Finally, the expected score is calculated as:

$$\mathbb{E}[Y \mid n, t, \hat{\theta}] = \sum_{y=0}^n y \Pr(Y = y \mid n, t, \hat{\theta}). \quad (26)$$

Simulation Study

We evaluated the proposed variability-augmented model using a Monte Carlo simulation study designed to test two related questions. The first question was whether the new model would avoid spurious dominance when the data were generated by the standard fixed-speed formulation. The second question was whether the new model would be advantageous for model selection and parameter recovery when trial-to-trial variability in processing speed was present. A third question concerned the relation between the new model and the t_0 -variability model of Dyrholm et al. (2011): whether the two forms of variability can be distinguished from whole-report data, and what each model estimates when it is fitted to data generated by the other. The simulation was therefore constructed to include both old-model and new-model data-generating regimes, together with a t_0 -variability regime, so that the extension could be evaluated under conditions favourable and unfavourable to its additional flexibility.

The simulated task matched the behavioural structure used throughout the present work. Each simulated dataset consisted of one observer performing a whole-report letter task without distractors at two display sizes, six letters and four letters. Exposure duration varied across seven levels, namely 20, 40, 60, 80, 120, 180, and 260 ms. For each exposure duration and each display-size condition, 30 trials were generated. Because there were two display-size conditions and seven exposure durations, each simulated dataset contained

$$N_{\text{trials}} = 2 \times 7 \times 30 = 420 \quad (27)$$

trials. As shown in Eq. 27, each replication provided 420 trial-level observations for model fitting.

On every simulated trial, a set of letters was sampled without replacement from a fixed alphabet, so that no display contained repeated letters. The letters I and O were excluded from the stimulus pool in order to avoid trivial ambiguities in alphanumeric identity, leaving the set ABCDEFGHJKLMNPQRSTUVWXYZ. If the display size on trial (j) was n_j , then n_j distinct letters were drawn and treated as the set of shown items. The response on each trial was generated as a subset of the shown letters, so the simulation did not introduce commission errors in the form of reporting

letters that had not been presented. The observed score was then defined as the number of correctly reported letters, and the number of incorrect items was computed as the difference between the display size and the number correct. Thus, if no letter was reported on a six-letter trial, the score was 0, and the number incorrect was 6.

The underlying encoding mechanism followed the count-based TVA formulation described in the model section. For each trial, the effective exposure duration was defined as

$$\tau_j = \max(0, t_j - t_0), \quad (28)$$

where t_j is the physical exposure duration and t_0 is the perceptual threshold. Given a trial-specific processing speed C_j , the per-letter processing rate was

$$v_j = \frac{C_j}{n_j}, \quad (29)$$

so that the finishing times of the displayed letters were generated as exponential random variables beginning after t_0 . Letters whose finishing times exceeded the physical exposure duration were treated as not encoded. Among the letters that did finish in time, only the first K_j^* were retained, where K_j^* was the adjacent-integer realisation of the participant's capacity parameter K . The simulated response was the set of stored letters, presented in randomised order. In this way, the generated trial-level data were fully consistent with the count-based assumptions of the fitted models. In the scenarios with t_0 variability described below, the fixed threshold t_0 in Eq. 28 was replaced on every trial by a draw $t_{0,j}$ from a normal distribution with mean μ_{t_0} and standard deviation σ_{t_0} , and the finishing times were generated as exponential random variables beginning after $t_{0,j}$, as assumed by the t_0 -variability model.

The Monte Carlo design contained eight generating scenarios. We generated three scenarios under the old fixed-speed model. These scenarios were intended to test whether the new model would incorrectly outperform the standard model when no meaningful trial-to-trial variability existed. In these three scenarios, the coefficient of variation of trial-level processing speed was set to a near-zero value, $cv_C = 0.0001$, so that trial-specific processing speed was effectively constant. The first old-truth scenario used $C = 32$, $K = 3.4$, and $t_0 = 15$ ms. The second used $C = 42$, $K = 3.6$, and $t_0 = 18$ ms. The third used $C = 52$, $K = 3.8$, and $t_0 = 20$ ms. These three settings represent low, typical, and high processing-speed regimes while retaining plausible values of storage capacity and threshold.

We generated three scenarios under the new variability-augmented model. These scenarios were intended to test whether the new model could recover trial-to-trial

variability and improve model selection when such variability was present. In all three variability scenarios, the mean processing speed, capacity, and threshold were held constant at $C = 42$, $K = 3.6$, and $t_0 = 18$ ms. The coefficient of variation of trial-level processing speed differed across scenarios, set to $cv_C = 0.15$, $cv_C = 0.30$, and $cv_C = 0.45$, respectively. The first value represented mild variability, the second moderate variability, and the third strong variability. Trial-specific speed was generated from a gamma distribution with mean C and coefficient of variation cv_C , so that

$$C_j \sim \text{Gamma}(\eta, \omega), \quad (30)$$

with shape and scale chosen to satisfy

$$\eta = \frac{1}{cv_C^2}, \quad \omega = C cv_C^2. \quad (31)$$

As shown in Eqs. 30 and 31, the variability scenarios differed only in the amount of trial-to-trial heterogeneity around the same mean processing speed. In terms of the distribution of the trial-level speed, these three values imply that the central 95% of trials fall between 0.73 and 1.31 times the mean speed for $cv_C = 0.15$, between 0.50 and 1.67 times the mean for $cv_C = 0.30$, and between 0.32 and 2.06 times the mean for $cv_C = 0.45$ (Fig. 1a). The labels mild, moderate, and strong are descriptive only; the three values were chosen to bracket the range over which the additional parameter goes from undetectable to clearly detectable with 420 trials, rather than to represent empirically established magnitudes.

The final two scenarios were generated under the t_0 -variability model. Mean processing speed, capacity, and mean threshold were again $C = 42$, $K = 3.6$, and $\mu_{t_0} = 18$ ms, with no variability in processing speed, and the standard deviation of the trial-level threshold was set to $\sigma_{t_0} = 10$ ms and $\sigma_{t_0} = 20$ ms, respectively. We calibrated these two values so

the evidence against the old model would be comparable to that in the moderate and strong speed-variability scenarios. For this purpose, we computed, for each variability scenario, the expected log-likelihood gain of the true model over the best-fitting old model for a dataset of 420 trials. This gain is 1.80 for $cv_C = 0.30$ and 7.39 for $cv_C = 0.45$, and 1.64 for $\sigma_{t_0} = 10$ ms and 8.16 for $\sigma_{t_0} = 20$ ms, so that the two moderate scenarios and the two strong scenarios carry similar evidence for their respective form of variability. With $\mu_{t_0} = 18$ ms, a standard deviation of 20 ms implies that about 18% of the trial-level thresholds are negative; such draws were used as they were, in both the generating process and the fitted model, and the consequences are taken up in the Discussion.

For each of the eight scenarios, 25 independent replications were generated, yielding a total of

$$N_{\text{datasets}} = 8 \times 25 = 200 \quad (32)$$

simulated datasets. Each dataset was treated as one simulated participant for fitting purposes. Across all replications, the simulation therefore generated

$$N_{\text{all trials}} = 200 \times 420 = 84,000 \quad (33)$$

trial-level observations. The random-number generator was initialised from a fixed master seed, and each replication was assigned its own derived seed, so that the complete simulation could be reproduced while preserving independence across replications. To examine how the results depend on the number of trials, four of the scenarios were repeated with 10 instead of 30 trials per cell, that is, with 140 trials per dataset: the typical old-truth scenario ($C = 42$), the moderate and strong speed-variability scenarios ($cv_C = 0.30$ and 0.45), and the strong t_0 -variability scenario ($\sigma_{t_0} = 20$ ms). Each was replicated 25 times, adding 100 datasets and 14,000 trials, so the complete study comprised 300 datasets

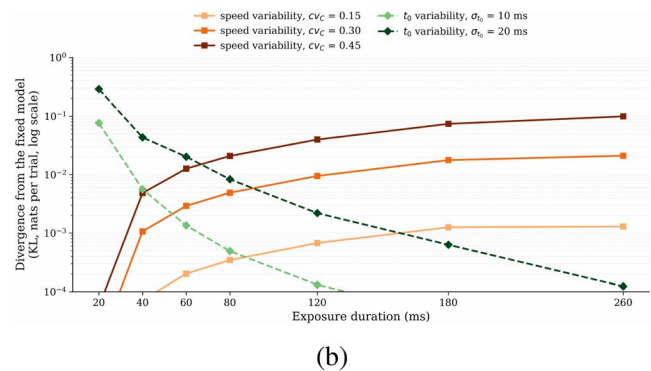
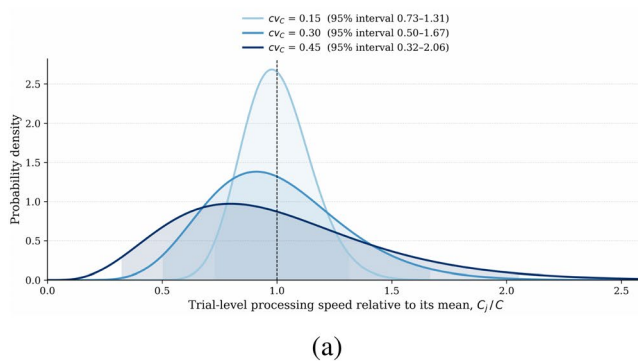


Fig. 1 (a) Gamma distributions of the trial-level processing speed C_j relative to its mean C for the three simulated variability levels; shaded regions cover the central 95% of trials. (b) Kullback–Leibler divergence per trial between the score distribution under each type of trial-

to-trial variability and the fixed-speed model with the same $C = 42$, $K = 3.6$, and $t_0 = 18$ ms (display size 6), as a function of exposure duration. Values below 10^{-4} are drawn at the axis floor

and 98,000 trials. The seeds of the reduced-trial datasets were derived after those of the main run, so that the main run is unaffected by this addition.

Both the old model and the new model were fitted to every simulated dataset. The t_0 -variability model was fitted to every dataset as well, so that each of the three models was evaluated on data generated by each of the three generating processes. In each case, fitting was performed by maximum likelihood on the trial-level score distribution, with repeated trials collapsed into unique likelihood cells defined by display size, exposure duration, and number correct in order to improve computational efficiency, using the collapsed log-likelihood of Eq. 24. Optimisation was carried out using the L-BFGS-B algorithm (Fei et al., 2014) with parameter bounds enforced directly. For the old model, C was constrained to the interval (0.01, 200) letters per second, K to (0.50, 6.00), and t_0 to (0, 150) ms. The new model added the constraint $cv_C \in (0.0001, 2.00)$; the lower bound determines the minimum detectable variability and should be kept in mind when interpreting estimates near zero. The t_0 -variability model used the same bounds for C and K , constrained μ_{t_0} to (0, 150) ms and σ_{t_0} to (0.1, 60) ms, and evaluated the integral in Eq. 22 by Gauss–Hermite quadrature with 48 nodes. No estimate of cv_C or σ_{t_0} reached its upper bound in any dataset; the largest estimates were $cv_C = 0.66$ and $\sigma_{t_0} = 36$ ms, both in the reduced-trial run. Estimates at the lower bound, by contrast, were common whenever the corresponding variability was absent, which is the expected behaviour of a variability parameter whose true value lies on the boundary of the parameter space. To reduce sensitivity to starting values, optimisation was repeated from multiple starting points. For the old model, a grid of 27 structured starts spanning plausible ranges of C , K , and t_0 was combined with 6 random starts drawn uniformly within the parameter bounds. For the new model, we constructed 99 structured starts by combining a range of cv_C values with perturbations around the old model's estimates, supplemented by 8 additional random starts. For the t_0 -variability model, 96 structured starts were constructed in the same way from a range of σ_{t_0} values and perturbations around the old model's estimates, supplemented by 8 random starts. The solution with the lowest negative log-likelihood across all starts was retained. The old model was always fitted first, and its parameter estimates were used to anchor part of the starting grid for the new model and for the t_0 -variability model, improving numerical stability for the extended models.

The numerical accuracy of the likelihood computations was verified before the simulation was run. The closed-form completed-count distribution of Eq. 19 agreed with direct numerical integration of Eq. 15 to within 5.3×10^{-8} in absolute probability over a grid of display sizes, exposures, speeds,

and values of cv_C between 0.05 and 2.0. At the lower bound $cv_C = 0.0001$, the new model reproduced the old model to within 8.0×10^{-9} , and at $\sigma_{t_0} = 0.1$ ms the t_0 -variability model reproduced the old model to within 1.4×10^{-5} . Both variability models also reproduced the score distributions obtained by direct simulation of their generating processes with 400,000 trials per condition to within Monte Carlo error. The vectorised likelihood used in the present run agreed with the scalar implementation of the original analysis to within 3.2×10^{-12} .

The old model was evaluated in terms of its estimates of C , K , and t_0 . The new model was evaluated on its estimates of C , K , t_0 , and cv_C . The t_0 -variability model was evaluated in terms of its estimates of C , K , μ_{t_0} , and σ_{t_0} , with μ_{t_0} compared with the generating value of t_0 . Model comparison was carried out with both the Akaike information criterion (Akaike, 2025) and the Bayesian information criterion (Neath & Cavanaugh, 2011),

$$\Delta AIC = AIC_{\text{new}} - AIC_{\text{old}}, \quad (34)$$

and

$$\Delta BIC = BIC_{\text{new}} - BIC_{\text{old}}. \quad (35)$$

The same differences were computed for the t_0 -variability model relative to the old model, and for the new model relative to the t_0 -variability model. In addition, the model with the lowest AIC and the model with the lowest BIC among the three were recorded for every dataset, which yields a cross-classification of generating truth against selected model. For the old-truth scenarios, the observed rates at which a variability model was preferred were compared with the rates expected from the boundary result described in the Method section. Under a fixed-speed truth, AIC prefers a variability model when the likelihood-ratio statistic exceeds 2, which happens with probability $\frac{1}{2} \Pr(\chi_1^2 > 2) = 0.079$; BIC prefers it when the statistic exceeds $\log N_{\text{trials}}$, which happens with probability 0.007 for 420 trials and 0.013 for 140 trials. In addition to information criteria, the fitted models were compared on condition-level predictive accuracy by computing the root mean squared error between observed and model-predicted mean scores across the combinations of display size and exposure duration. We included this additional measure because a model can sometimes improve descriptive accuracy without necessarily being preferred once parameter-count penalties are imposed. Accordingly, the simulation distinguished between descriptive gains and evidence that the additional variability parameter was warranted.

We assessed parameter recovery by comparing estimated values with the known generating values in each scenario. For each parameter, we summarised both bias and root mean squared error across replications. For the old model, we did

this for C , K , and t_0 . For the new model, the same quantities were computed, and the recovery of cv_C was additionally examined in the variability scenarios. For the t_0 -variability model, the same quantities were computed for C , K , and μ_{t_0} , together with the recovery of σ_{t_0} . Because every model was fitted to every dataset, these summaries also show what each model estimates when it is applied to data containing the other form of variability, or no variability at all. This allowed the simulation to address two separate but related questions. The first was whether the new model could correctly identify trial-to-trial variability when it existed. The second was whether, by modelling that variability explicitly, it could improve recovery of the core TVA parameters relative to the standard fixed-speed formulation.

Results

Before turning to the Monte Carlo results, Fig. 1 shows what the two forms of trial-to-trial variability assume and where they leave their trace in whole-report data. Panel A shows the gamma distributions of the trial-level processing speed for the three simulated values of cv_C . Panel B shows, for each exposure duration and display size 6, the Kullback–Leibler divergence between the score distribution under each type of variability and the score distribution of the fixed-speed model with the same $C = 42$, $K = 3.6$, and $t_0 = 18$ ms. The two forms of variability have opposite profiles. Speed variability leaves almost no trace at the shortest exposures and grows steadily with exposure duration, because the multiplicative fluctuation of the rate has more time to act and because the average of $1 - e^{-v\tau}$ over a variable rate falls below the value at the mean rate. Variability in t_0 is concentrated at the shortest exposures, where a jitter of a few milliseconds decides whether encoding starts at all, and becomes negligible at long exposures. Because the

present design spans both regions, the two forms of variability are in principle separable from the data, and the simulations below test whether this holds in practice.

Figure 2 shows the mean number of correctly reported letters as a function of exposure duration at display size 6 for one scenario of each generating truth: the typical fixed-speed scenario ($C = 42$), the strong speed-variability scenario ($cv_C = 0.45$), and the strong t_0 -variability scenario ($\sigma_{t_0} = 20$ ms), all with the same C , K , and mean t_0 . In all scenarios, performance increased monotonically with exposure duration, consistent with the capacity-limited race architecture of TVA. The three curves are close to each other, and the differences between them are of the size predicted by Fig. 1b. Speed variability lowered the mean score at the longer exposures by just under 0.2 letters at 260 ms, while leaving the shortest exposures unchanged. Variability in t_0 raised the mean score at the shortest exposure, from 0.07 to 0.34 letters at 20 ms, because on trials with a low threshold some encoding took place within 20 ms, and left the longer exposures almost unchanged. This similarity at the level of condition means is expected, because the gamma-distributed speed fluctuations average out across replications, leaving the mean score relatively stable. The distinguishing effects of variability are therefore not immediately visible in the average performance curves but instead manifest in model selection and parameter recovery, as shown in the following sections.

Model Selection

The model-selection results are summarised in Table 1 and Fig. 3. In the three old-model scenarios, AIC rarely preferred the new model, and BIC never did. In Scenario 1 (old model, low processing speed), the new model was selected by AIC in only 1 of 25 replications and by BIC in 0 of 25 replications. The same pattern was observed in

Fig. 2 Mean number of correctly reported letters as a function of exposure duration at display size 6 for one scenario of each generating truth (Scenario 2: fixed speed; Scenario 6: strong speed variability; Scenario 8: strong t_0 variability). Error bars indicate ± 1 standard deviation of the per-dataset means across 25 replications; the dotted line marks the true $K = 3.6$

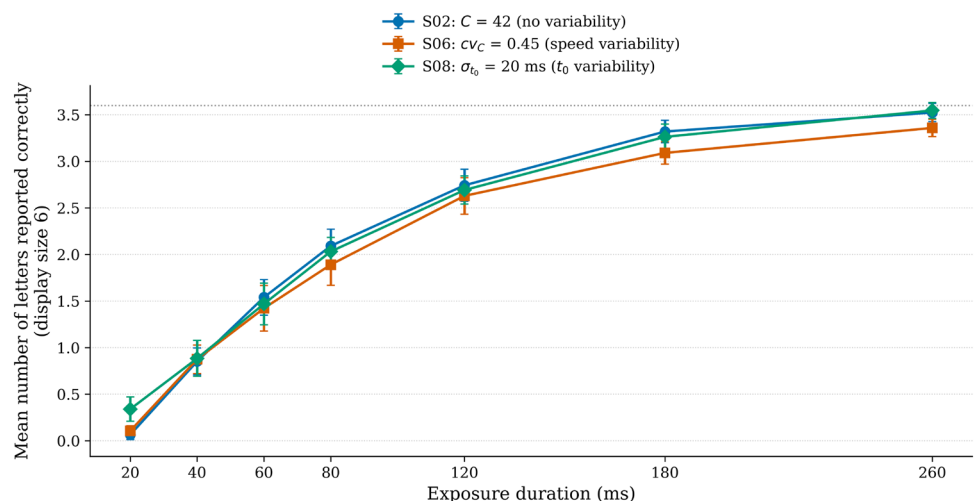


Table 1 Scenario-level model-selection summary for the new model relative to the old model. Negative values of ΔAIC and ΔBIC favour the new model. In the old-truth scenarios and in Scenarios 4 and 7, many of the RMSE comparisons are ties to within 10^{-6} because cv_C was estimated at its lower bound and the two models then make the same predictions; the RMSE win counts in those rows are therefore not informative

Scenario	Truth	New wins AIC	New wins BIC	New wins RMSE	Mean ΔAIC	Mean ΔBIC
Scenario 1: old, low C	old	1/25	0/25	13/25	1.7271	5.7674
Scenario 2: old, typical C	old	1/25	0/25	15/25	1.4963	5.5366
Scenario 3: old, high C	old	5/25	0/25	15/25	1.1021	5.1424
Scenario 4: new, mild variability	new	1/25	1/25	15/25	1.2959	5.3362
Scenario 5: new, moderate variability	new	14/25	7/25	13/25	-1.8246	2.2156
Scenario 6: new, strong variability	new	25/25	22/25	22/25	-13.5665	-9.5263
Scenario 7: t_0 var., moderate ($\sigma_{t_0} = 10$ ms)	t_0 var.	1/25	0/25	7/25	1.6861	5.7264
Scenario 8: t_0 var., strong ($\sigma_{t_0} = 20$ ms)	t_0 var.	10/25	5/25	4/25	-1.3875	2.6527

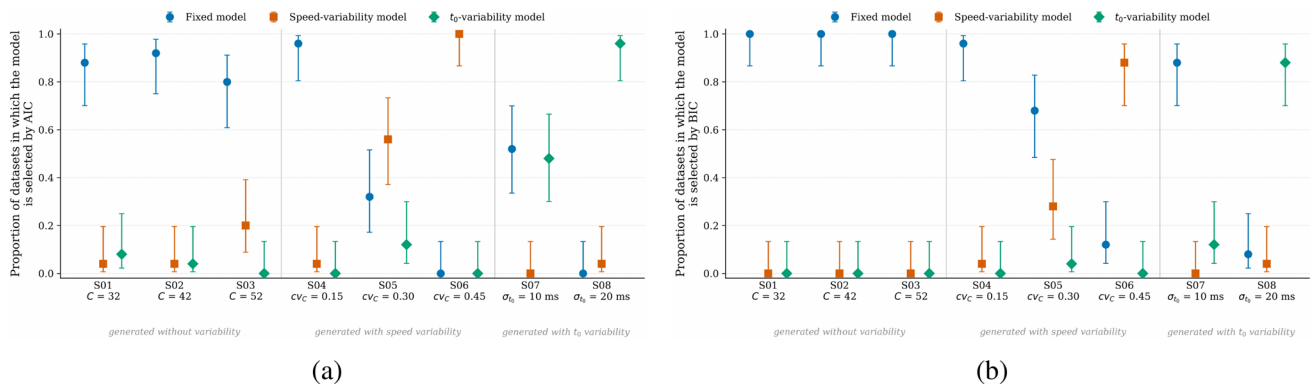


Fig. 3 Three-way model selection with 420 trials per dataset and 25 datasets per scenario: proportion of datasets in which each of the three models had the lowest AIC (a) or BIC (b). Error bars are 95% Wilson intervals

Scenario 2 (old model, typical processing speed), again with 1 of 25 AIC wins and no BIC wins. In Scenario 3 (old model, high processing speed), the new model was selected somewhat more often by AIC, but still only in 5 of 25 replications, while BIC again preferred the old model in every replication. Thus, when the generating process was effectively fixed-speed, the old model remained the appropriate description. The corresponding 95% confidence intervals on the AIC win rates for Scenarios 1–3 were (0.007, 0.195), (0.007, 0.195), and (0.089, 0.391) respectively, confirming that even the upper bound of plausible win rates remained modest in the old-truth scenarios. These rates are close to what is expected when the variability parameter lies on the boundary of its parameter space. Under a fixed-speed truth, AIC is expected to prefer the new model in 7.9% of datasets (about 2 of 25), and BIC in 0.7%, as derived in the Method section. Pooled over the 75 old-truth datasets, the new model was preferred by AIC in 7 (9.3%, 95% CI 0.046–0.180) against 5.9 expected, and by BIC in none against 0.5 expected. The 5 of 25 AIC wins in Scenario 3 are therefore within the range produced by sampling variation around the boundary null, and not a sign that the new model becomes systematically preferred at higher processing speeds.

As shown in Table 1 and Fig. 3, the mild-variability scenario did not yet provide meaningful support for the new model. In Scenario 4 (new model, mild variability), both AIC and BIC selected the new model in only 1 of 25 replications, and the mean values of both ΔAIC and ΔBIC remained positive. This indicates that, under the present design, mild trial-to-trial variability in processing speed was not sufficient to consistently justify the added parameter.

The pattern changed in Scenario 5 (new model, moderate variability). Here, the new model was selected by AIC in 14 of 25 replications (95% CI: 0.37–0.73), although BIC selected it in only 7 of 25 replications (95% CI: 0.14–0.48). The mean ΔAIC was negative (−1.8246), whereas the mean ΔBIC remained positive (2.2156). The wide confidence intervals on both win rates reflect the small number of replications and indicate that this scenario is genuinely transitional rather than providing consistent support for either model.

The clearest support for the new model appeared in Scenario 6 (new model, strong variability). In this condition, the new model was selected in all 25 replications by AIC (95% CI: 0.87–1.00) and in 22 of 25 replications by BIC (95% CI: 0.70–0.96). The corresponding mean values, $\Delta\text{AIC} = -13.5665$ and $\Delta\text{BIC} = -9.5263$, show that the gain in fit was sufficiently large to justify the additional

parameter even under stronger penalisation. The narrow confidence intervals here confirm that this result is stable across replications rather than driven by a small number of favourable outcomes.

The condition-level RMSE summaries in Table 1 should be interpreted more cautiously. In the old-model scenarios, the new model sometimes improved RMSE even though the information criteria did not prefer it. This indicates that the additional flexibility could occasionally improve descriptive fit without being justified once model complexity was accounted for. In addition, whenever cv_C was estimated at its lower bound, which happened in 14, 12, and 6 of the 25 datasets of Scenarios 1 to 3 and in 13 datasets of Scenario 7, the new model reduced to the old model and the two RMSE values were equal to within 10^{-6} ; which model is counted as the winner in those datasets is then decided by rounding. For inferential purposes, the AIC and especially the BIC patterns are therefore more informative than RMSE alone.

Figure 4a shows the dataset-level evidence behind these counts, for the new model and for the t_0 -variability

model relative to the old model. In the old-model scenarios, where $cv_C \approx 0$, replication-level ΔAIC values clustered around small positive values, confirming that the additional parameter was generally penalised. As the true variability increased through the mild, moderate, and strong regimes, the distribution of ΔAIC shifted progressively downward. At $cv_C = 0.45$, nearly all replications fell well below zero. Because the two variability models each add one parameter to the old model, ΔBIC equals $\Delta AIC + 4.04$ for 420 trials, and the dotted line in Figure 4a marks the value of ΔAIC below which BIC also prefers the variability model. This figure therefore provides a replication-level view of the transition from negligible to decisive evidence for the variability-augmented model, complementing the scenario-level summaries in Table 1 and Fig. 3.

Comparison with the t_0 -Variability Model

Table 2 summarises the same datasets from the point of view of the t_0 -variability model. In the three old-truth scenarios, the t_0 -variability model was preferred over the old

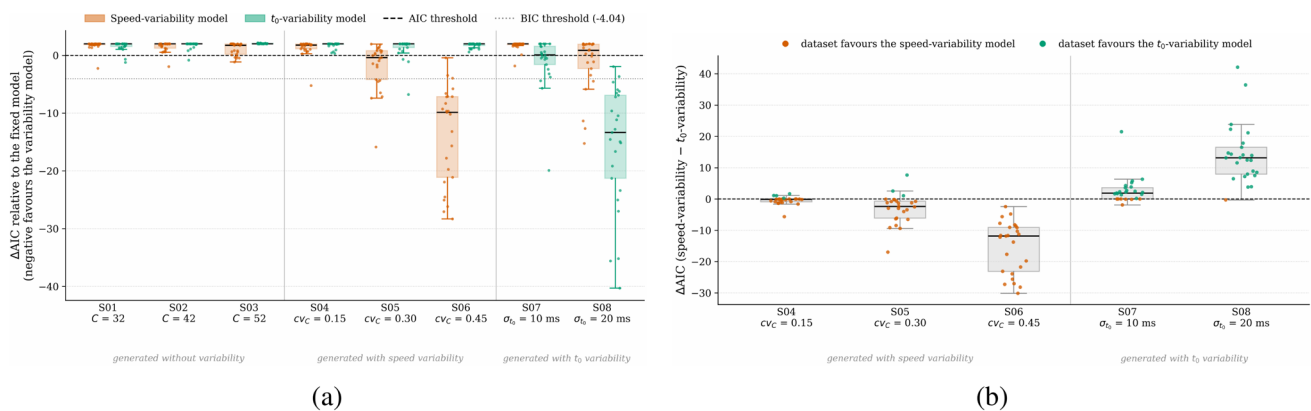


Fig. 4 (a) Dataset-level ΔAIC of each variability model relative to the old model (boxes: interquartile range and median; points: individual datasets). Values below the dashed line favour the variability model by AIC; values below the dotted line (-4.04) favour it by BIC as well. (b)

Direct comparison of the two variability models on the datasets generated with variability, $\Delta AIC = AIC_{\text{new}} - AIC_{t_0 \text{ var.}}$; both models have four parameters, so AIC and BIC decisions coincide

Table 2 Model selection involving the t_0 -variability model (420 trials per dataset, 25 datasets per scenario). Columns 3 to 6 compare the t_0 -variability model with the old model (negative values favour the t_0 -variability model). Columns 7 and 8 compare the new model with the t_0 -variability model (negative values favour the new model). The last two columns give the number of datasets in which the old, the new, and the t_0 -variability model had the lowest AIC and the lowest BIC among the three

Scenario	Truth	t_0 var. wins		Mean ΔAIC	Mean ΔBIC	New wins AIC	Mean ΔAIC	Selected by AIC	Selected by BIC
		AIC	BIC	(t_0 var. - old)	(t_0 var. - old)	vs. t_0 var.	(new - t_0 var.)	old / new / t_0 var.	old / new / t_0 var.
Scenario 1	old	2/25	0/25	1.5899	5.6302	16/25	0.1372	22 / 1 / 2	25 / 0 / 0
Scenario 2	old	1/25	0/25	1.7738	5.8140	21/25	-0.2775	23 / 1 / 1	25 / 0 / 0
Scenario 3	old	0/25	0/25	2.0369	6.0772	25/25	-0.9348	20 / 5 / 0	25 / 0 / 0
Scenario 4	new	0/25	0/25	1.7439	5.7841	21/25	-0.4479	24 / 1 / 0	24 / 1 / 0
Scenario 5	new	3/25	1/25	1.2353	5.2756	22/25	-3.0600	8 / 14 / 3	17 / 7 / 1
Scenario 6	new	0/25	0/25	1.7477	5.7879	25/25	-15.3142	0 / 25 / 0	3 / 22 / 0
Scenario 7	t_0 var.	12/25	3/25	-1.0555	2.9848	7/25	2.7416	13 / 0 / 12	22 / 0 / 3
Scenario 8	t_0 var.	25/25	23/25	-15.5413	-11.5011	1/25	14.1538	0 / 1 / 24	2 / 1 / 22

model by AIC in 2, 1, and 0 of 25 replications and by BIC in none, so that its false-alarm rate (3 of 75 by AIC, 4.0%, 95% CI 0.014–0.111) was, like that of the new model, in the range expected under the boundary null. In the three speed-variability scenarios, the t_0 -variability model gained nothing over the old model: it was preferred by AIC in 0, 3, and 0 of 25 replications, its mean Δ AIC relative to the old model stayed close to the penalty value of 2 (1.7439, 1.2353, and 1.7477), and in Scenario 6, where the new model was preferred in every dataset, the t_0 -variability model was preferred in none. The variability that the new model detected in these data was therefore not visible to a model that allows only the threshold to vary.

The two t_0 -variability scenarios mirror the two speed-variability scenarios to which they were calibrated. In Scenario 7 ($\sigma_{t_0} = 10$ ms), the t_0 -variability model was preferred over the old model by AIC in 12 of 25 replications (95% CI 0.30–0.67) and by BIC in 3 of 25, with a mean Δ AIC of -1.0555 ; this is the same transitional picture as in Scenario 5 (14 of 25 and 7 of 25). In Scenario 8 ($\sigma_{t_0} = 20$ ms), it was preferred by AIC in all 25 replications and by BIC in 23 of 25, with a mean Δ AIC of -15.5413 , matching the decisive result of Scenario 6. The new model, fitted to the same t_0 -variability data, behaved differently from the t_0 -variability model fitted to speed-variability data: in Scenario 8 it was preferred over the old model by AIC in 10 of 25 replications and by BIC in 5 of 25 (Table 1), so that part of the threshold variability was absorbed as apparent speed variability. This absorption was partial. In the direct comparison of the two variability models (Fig. 4b, Table 2), the t_0 -variability model was preferred in 24 of 25 datasets of Scenario 8, with a mean advantage of 14.15 AIC units, and the new model was preferred in all 25 datasets of Scenario 6, with a mean advantage of 15.31 AIC units. Because both models have four parameters, these decisions are the same under AIC and BIC. In Scenario 4 and Scenario 7, where the amount of variability was small, the two variability models were close to each other, and both were close to the old model.

Table 3 collects these outcomes in a cross-classification of generating truth against the model selected among the three. With 420 trials, misattribution between the two forms of variability was rare: of the 75 datasets generated with speed variability, 3 (4.0%) were assigned to the t_0 -variability model by AIC and 1 by BIC, and of the 50 datasets generated with t_0 variability, 1 (2.0%) was assigned to the new model by either criterion. The errors that did occur were almost entirely failures to detect variability at all, that is, selections of the old model, and they were concentrated in the mild and moderate scenarios. On data without variability, BIC selected the old model in all 75 datasets and AIC in 65 (86.7%).

Table 3 Cross-classification of generating truth against the model selected among the three (number of datasets, with proportions in parentheses). Rows pool all scenarios of the same truth type: 75 datasets without variability (Scenarios 1 to 3), 75 with speed variability (Scenarios 4 to 6), and 50 with t_0 variability (Scenarios 7 and 8) in the main run; 25, 50, and 25 datasets in the reduced-trial run

Trials per dataset, criterion	Generating truth	Old selected	New selected	t_0 var. selected
420, AIC	no variability	65 (0.87)	7 (0.09)	3 (0.04)
	speed variability	32 (0.43)	40 (0.53)	3 (0.04)
	t_0 variability	13 (0.26)	1 (0.02)	36 (0.72)
420, BIC	no variability	75 (1.00)	0 (0.00)	0 (0.00)
	speed variability	44 (0.59)	30 (0.40)	1 (0.01)
	t_0 variability	24 (0.48)	1 (0.02)	25 (0.50)
140, AIC	no variability	19 (0.76)	1 (0.04)	5 (0.20)
	speed variability	21 (0.42)	27 (0.54)	2 (0.04)
	t_0 variability	3 (0.12)	1 (0.04)	21 (0.84)
140, BIC	no variability	24 (0.96)	1 (0.04)	0 (0.00)
	speed variability	38 (0.76)	12 (0.24)	0 (0.00)
	t_0 variability	8 (0.32)	0 (0.00)	17 (0.68)

Parameter Recovery

The parameter-recovery results are shown in Table 4 and Figure 5. In the three old-model scenarios, the old model generally recovered the generating parameters at least as well as, and often better than, the new model. In Scenario 1 (old model, low processing speed), the old model produced lower RMSE for C and t_0 , whereas the two models were nearly identical for K . In Scenario 2 (old model, typical processing speed), the same general pattern held: the old model showed smaller RMSE for C and K , whereas the two models were effectively equal for t_0 . In Scenario 3 (old model, high processing speed), the difference became more pronounced, with the old model clearly outperforming the new model for C and t_0 . These results are consistent with the model-selection findings in showing that the extended model did not offer a general advantage when the generating process was fixed-speed. The t_0 -variability model behaved in the same way on these data for C and K , but it recovered t_0 less well than the old model (RMSE of 2.2562, 1.3514, and 1.6742 ms in Scenarios 1 to 3, against 1.4519, 0.8699, and 1.6645 ms), because its estimate of the mean threshold moves together with its estimate of σ_{t_0} when the latter is pushed away from zero by sampling noise.

In Scenario 4 (new model, mild variability), the new model still did not show a clear overall recovery advantage. It yielded a slightly smaller RMSE for K than the old model (0.0571 vs 0.0589), but performed worse for C (2.4488 vs 2.0483) and was essentially indistinguishable for t_0 (0.8726 vs 0.8694 ms). Thus, the mild-variability regime did not yet provide strong evidence that explicitly modelling trial-to-trial variability improved inference.

Table 4 Scenario-level parameter-recovery RMSE for the three fitted models (420 trials per dataset, 25 datasets per scenario). For the new model, recovery of cv_C is reported only for the extended model because the old model does not estimate this parameter. For the t_0 -variability model, t_0 refers to the estimated mean threshold μ_{t_0} , and the RMSE of σ_{t_0} is given in the last column. The true value of cv_C is 0 in Scenarios 1 to 3 (generated with 0.0001) and in Scenarios 7 and 8; the true value of σ_{t_0} is 0 in Scenarios 1 to 6

Scenario	Truth	C_{old}		C_{new}		C_{t_0}		K_{old}		K_{new}		K_{t_0}		$t_{0,old}$		$t_{0,new}$		t_{0,t_0}		cv_C_{new}		σ_{t_0,t_0}	
		RMSE		RMSE		RMSE		RMSE		RMSE		RMSE		RMSE		RMSE		RMSE		RMSE		RMSE	
Scenario 1: old, low C	old	1.7932		1.9459		1.9453		0.0663		0.0662		0.0659		1.4519		1.5184		2.2562		0.1146		5.4384	
Scenario 2: old, typical C	old	2.3461		2.7354		2.5969		0.0614		0.0637		0.0629		0.8699		0.8635		1.3514		0.1424		2.4529	
Scenario 3: old, high C	old	2.5054		4.3310		2.5128		0.0588		0.0627		0.0587		1.6645		2.5184		1.6742		0.1773		0.4258	
Scenario 4: new, mild variability	new	2.0483		2.4488		2.0259		0.0589		0.0571		0.0569		0.8694		0.8726		1.7868		0.0966		3.4046	
Scenario 5: new, moderate variability	new	4.2137		3.0167		3.8052		0.0681		0.0732		0.0665		1.0494		0.8314		1.6830		0.0864		5.2830	
Scenario 6: new, strong variability	new	6.8408		2.4605		6.7629		0.0929		0.0574		0.0920		2.0121		0.8627		2.0045		0.0698		3.8938	
Scenario 7: t_0 var., moderate	t_0 var.	2.9330		2.5217		2.8101		0.0567		0.0546		0.0540		3.1115		2.9504		2.9919		0.1218		6.0070	
Scenario 8: t_0 var., strong	t_0 var.	5.0400		3.2421		2.9868		0.0741		0.0627		0.0610		6.4530		5.6924		3.3853		0.2841		4.2105	

The results changed in Scenario 5 (new model, moderate variability). Here, the new model reduced RMSE for C from 4.2137 to 3.0167 and for t_0 from 1.0494 to 0.8314 ms, although it was slightly worse for K (0.0732 vs. 0.0681). This pattern suggests that the added flexibility began to recover the generating structure more effectively, but that the advantage was still selective rather than uniform across all parameters.

The clearest recovery advantage again appeared in Scenario 6 (new model, strong variability). In this condition, the new model markedly outperformed the old model for all three core parameters. RMSE for C decreased from 6.8408 to 2.4605, RMSE for K decreased from 0.0929 to 0.0574, and RMSE for t_0 decreased from 2.0121 to 0.8627 ms. Figure 5 shows these differences directly. The t_0 -variability model provided no such improvement on speed-variability data. Its RMSE for C in Scenarios 5 and 6 (3.8052 and 6.7629) was almost the same as that of the old model (4.2137 and 6.8408), and its RMSE for t_0 in Scenario 6 (2.0045 ms) was nearly identical to that of the old model (2.0121 ms). Conversely, in the two t_0 -variability scenarios the t_0 -variability model had the lowest RMSE for C (2.8101 and 2.9868 against 2.9330 and 5.0400 for the old model) and for t_0 (2.9919 and 3.3853 ms against 3.1115 and 6.4530 ms), whereas the new model removed only part of the error in C (2.5217 and 3.2421) and little of the error in t_0 (2.9504 and 5.6924 ms). The recovery results therefore support the same conclusion as the information criteria: the extended model becomes useful when variability in processing speed is sufficiently large, but not when it is absent or only weakly expressed, and each variability model improves recovery only for the type of variability it was built for.

Although RMSE summarises the magnitude of estimation error, it does not reveal the direction of that error. Figure 6 and Table 5 show the signed mean bias, defined as the mean of estimated minus true values, for C and t_0 under the three models. In the old-model scenarios, where no meaningful variability existed, all three models exhibited only small biases for C and t_0 , consistent with generally adequate recovery. The biases of the two variability models were, however, not smaller than those of the old model. For C , the new model was biased upward by 0.36, 1.23, and 2.93 letters per second in Scenarios 1 to 3 (1.1% to 5.6% of the true value), compared with -0.01 , 0.43 , and 0.62 for the old model; for t_0 , its bias was 0.51 , 0.20 , and 1.78 ms, compared with 0.41 , 0.13 , and 1.01 ms. The reason is the coupling between $\hat{C}$ and $\hat{cv}_C$ noted in the section on the recovery of trial-to-trial variability below: whenever sampling noise is absorbed as spurious variability, the estimate of mean speed rises with it. The t_0 -variability model showed the corresponding coupling between $\hat{\mu}_{t_0}$ and $\hat{\sigma}_{t_0}$, with biases in t_0 of 1.18 , 0.50 , and 1.09 ms in Scenarios 1 to 3. As the true level

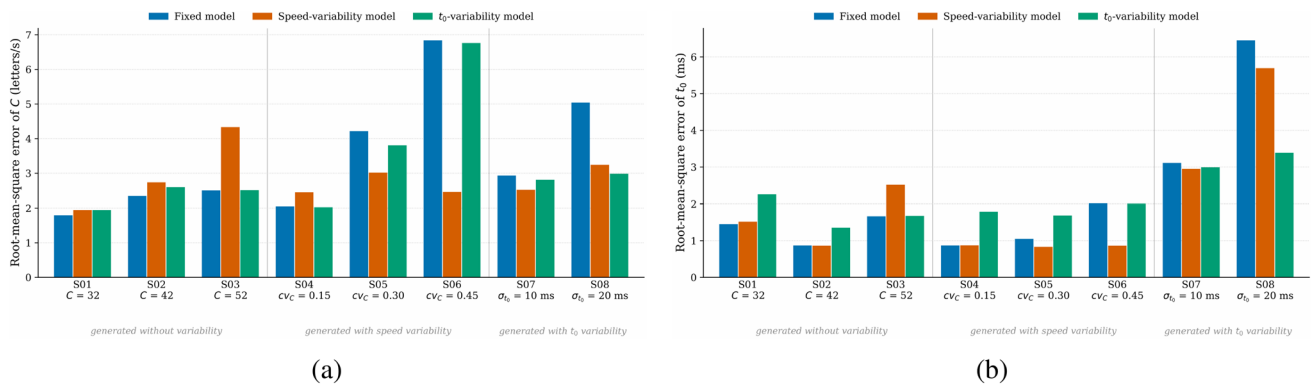


Fig. 5 Recovery RMSE for C (a) and t_0 (b) under the three fitted models across the eight generating scenarios (420 trials per dataset). For the t_0 -variability model, t_0 refers to the estimated mean threshold

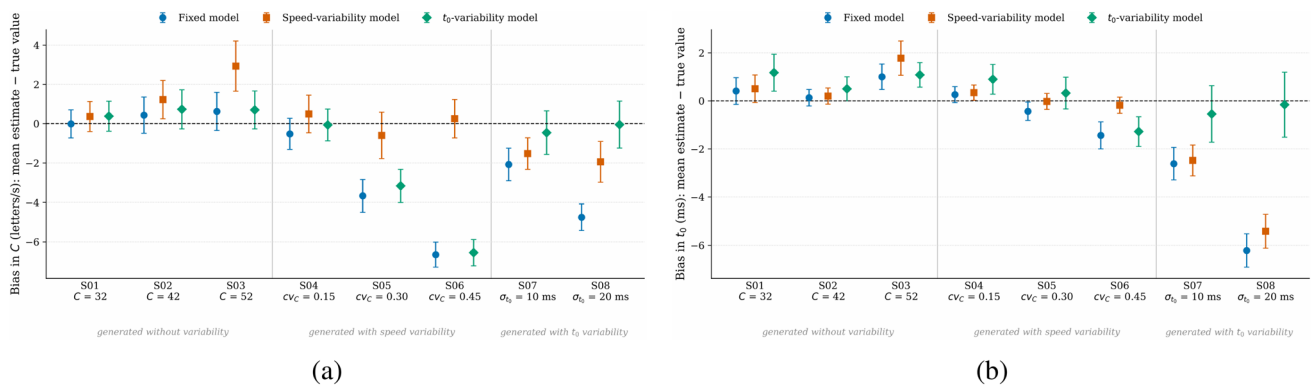


Fig. 6 Mean signed bias (estimated minus true) for C (a) and t_0 (b) under the three fitted models across the eight generating scenarios. Values near zero indicate unbiased recovery. Error bars are 95% confidence intervals across the 25 datasets of each scenario. The old model

developed a systematic negative bias for C as true trial-to-trial variability increased, and a systematic negative bias for t_0 as t_0 variability increased; only the correctly specified variability model removed the respective bias

Table 5 Mean signed bias (estimated minus true) of C (letters per second) and t_0 (ms) for the three fitted models (420 trials per dataset, 25 datasets per scenario). For the t_0 -variability model, t_0 refers to the estimated mean threshold

Scenario	Truth	C_{old}	C_{new}	C_{t_0}	$t_{0,old}$	$t_{0,new}$	t_{0,t_0}
Scenario 1: old, low C	old	-0.0089	0.3638	0.3824	0.4139	0.5072	1.1753
Scenario 2: old, typical C	old	0.4347	1.2283	0.7314	0.1329	0.2012	0.5045
Scenario 3: old, high C	old	0.6232	2.9332	0.7004	1.0057	1.7792	1.0859
Scenario 4: new, mild variability	new	-0.5194	0.4972	-0.0629	0.2632	0.3435	0.9008
Scenario 5: new, moderate variability	new	-3.6697	-0.5960	-3.1682	-0.4317	-0.0217	0.3274
Scenario 6: new, strong variability	new	-6.6528	0.2558	-6.5518	-1.4364	-0.1796	-1.2761
Scenario 7: t_0 var., moderate	t_0 var.	-2.0750	-1.5237	-0.4567	-2.6130	-2.4785	-0.5415
Scenario 8: t_0 var., strong	t_0 var.	-4.7503	-1.9423	-0.0455	-6.2152	-5.4119	-0.1598

of variability increased, however, the old model developed a systematic negative bias for C , reaching -6.65 letters per second in Scenario 6. This indicates that, when processing speed fluctuated across trials, the fixed-speed model consistently underestimated mean processing capacity. A corresponding negative bias emerged for t_0 under the old model in the variability scenario (-1.44 ms), suggesting that some of the unmodelled speed variability was absorbed as a compensatory shift in the estimated perceptual threshold. The new model kept the bias in C small in these scenarios

(0.50, -0.60 , and 0.26 letters per second in Scenarios 4 to 6) and the bias in t_0 below 0.35 ms in absolute value. The t_0 -variability model, fitted to the same data, reproduced the bias of the old model almost (-6.55 letters per second for C and -1.28 ms for t_0 in Scenario 6): allowing the threshold to vary offers no protection against the bias caused by speed variability. The t_0 -variability scenarios show the mirror image. There, the old model underestimated t_0 by 2.61 and 6.22 ms and C by 2.08 and 4.75 letters per second in Scenarios 7 and 8, the new model reduced the bias in C

(to -1.52 and -1.94) but left most of the bias in t_0 in place (-2.48 and -5.41 ms), and only the t_0 -variability model was close to unbiased for both parameters (-0.46 and -0.05 letters per second for C , -0.54 and -0.16 ms for t_0). This directional pattern is important because it shows that the old model's increased RMSE under variability was not merely due to increased scatter around the true value, but reflected a systematic distortion of the parameter estimates. Such distortion would be consequential in any application in which the absolute level of estimated processing speed is used to compare groups or conditions that differ in the amount of trial-to-trial variability. If two groups share the same cv_C , the bias is common to both and cancels in their difference; if a clinical group shows more variability than a control group, the fixed-speed model would attribute part of that difference to a lower mean processing speed.

Recovery of Trial-to-Trial Variability

Because the defining addition of the new model is the variability parameter cv_C , its recovery is central to the interpretation of the simulation. Figure 7a shows the replication-level distributions of the fitted values across scenarios together with the true values, and Figure 7b shows the same for σ_{t_0} under the t_0 -variability model.

In the three old-model scenarios, where the generating variability was effectively zero, the new model still estimated small positive values of cv_C . The corresponding RMSE values were 0.1146, 0.1424, and 0.1773 for Scenarios 1–3, respectively. This indicates that the extended model can absorb a modest amount of random fluctuation as apparent variability even when the generating process is fixed-speed. However, this tendency did not lead to systematic model-selection dominance, as shown above, because the information criteria still generally penalised the unnecessary parameter. Because cv_C was constrained to be non-negative with a lower bound of 0.0001, estimates in the

old-truth scenarios are necessarily bounded away from the true generating value of approximately zero, which partially contributes to the reported RMSE for this parameter in those scenarios. The distribution of these estimates is the one expected for a parameter whose true value lies on the boundary: in 14, 12, and 6 of the 25 datasets of Scenarios 1 to 3 the estimate was at the lower bound, and the remaining estimates were spread between 0.03 and 0.33 (Figure 7a). Sampling noise that happens to look like overdispersion is absorbed as positive cv_C , whereas noise in the other direction cannot be absorbed because cv_C cannot become negative, so that the mean estimate is positive even though the true value is zero. The estimates of σ_{t_0} under the t_0 -variability model showed the same pattern on the same data, with 16, 20, and 22 of 25 estimates at the lower bound and a mean bias of 3.19, 1.09, and 0.23 ms.

In Scenario 4 (new model, mild variability), the true value was $cv_C = 0.15$, and the RMSE of the fitted variability parameter was 0.0966. This recovery was limited and consistent with the weak model-selection evidence in this scenario. In Scenario 5 (new model, moderate variability), the RMSE of cv_C improved to 0.0864, indicating that the variability parameter had become more identifiable. In Scenario 6 (new model, strong variability), recovery was clearest, with an RMSE of 0.0698. As shown in Figure 7a, the estimated variability parameter tracked the increase in the true generating variability and did so with increasing precision as variability became stronger. The mean bias of $\hat{cv}_C$ in these three scenarios was small and slightly negative (-0.015 , -0.020 , and -0.010). The estimates of σ_{t_0} in the two t_0 -variability scenarios behaved in the same way (Figure 7b): in Scenario 8 they were centered on the true value of 20 ms (bias 0.67 ms, RMSE 4.21 ms, no estimate at the lower bound), whereas in Scenario 7, with a true value of 10 ms, 6 of 25 estimates fell to the lower bound and the mean estimate was biased downward by 2.96 ms (RMSE 6.01 ms), which is the boundary effect seen from the other side.

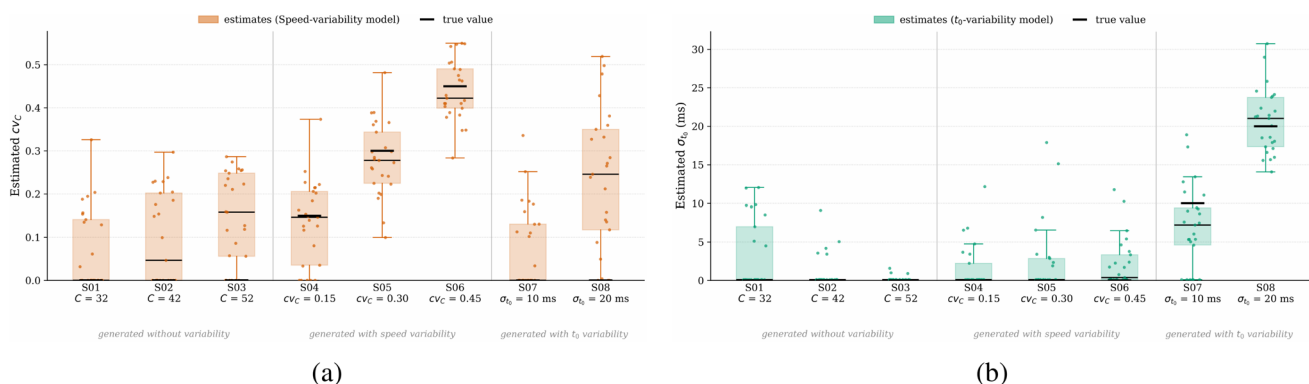


Fig. 7 Dataset-level estimates of the variability parameters (420 trials per dataset) with true values as black bars. **(a)** cv_C from the new model. **(b)** σ_{t_0} from the t_0 -variability model. Boxes show the inter-

quartile range and median; points are individual datasets. Estimates pile up at the lower bound when the corresponding type of variability is absent

Figure 7 also shows what each variability model estimates when the data contain the other form of variability. In the speed-variability scenarios, the t_0 -variability model estimated σ_{t_0} at or near the lower bound (18, 16, and 12 of 25 estimates at the bound in Scenarios 4 to 6; median 0.4 ms in Scenario 6), so that it did not register speed variability as threshold variability. In the t_0 -variability scenarios, the new model estimated substantial spurious speed variability: in Scenario 8, the median estimate of cv_C was 0.25 and only 3 of 25 estimates were at the lower bound, with a mean bias of 0.24. Threshold variability can therefore partly masquerade as speed variability, but speed variability does not masquerade as threshold variability. This asymmetry follows from Figure 1b: a heavy-tailed distribution of speeds produces some trials with very fast encoding and thereby mimics part of the raised performance at short exposures that threshold variability produces, whereas no distribution of thresholds can produce the overdispersion at long exposures that speed variability produces, because at long exposures the threshold no longer matters.

Effect of the Number of Trials

Table 6 shows the results of the reduced-trial run, in which four scenarios were repeated with 140 instead of 420 trials per dataset. Detection of variability weakened as expected. In Scenario 6, the new model was preferred over the old model by AIC in 18 of 25 datasets (95% CI 0.52–0.86) instead of 25 of 25, and by BIC in 9 of 25 instead of 22 of 25; in Scenario 5 the counts fell from 14 and 7 to 9 and 3 of 25; and in Scenario 8 the t_0 -variability model was preferred in 21 instead of 25 datasets by AIC and in 17 instead of 23 by BIC. Misattribution between the two forms of variability remained rare (Table 3). The false-alarm rate of the new model on the fixed-speed data of Scenario 2 stayed within the boundary-null expectation (1 of 25 by AIC and 1 of 25 by BIC), whereas the t_0 -variability model was preferred by

AIC in 5 of 25 of these datasets (95% CI 0.09–0.39), above the 2 expected, although not by BIC in any.

Parameter recovery with 140 trials makes the cost of the additional parameter visible. On the fixed-speed data of Scenario 2, the RMSE of C was 3.16 letters per second for the old model but 5.70 for the new model and 4.82 for the t_0 -variability model, and the upward bias of the new model in C grew to 2.85 letters per second. In Scenario 5, the new model removed most of the bias in C (1.32 against –2.48 for the old model) but had the larger RMSE (5.10 against 4.13), because the variance added by the poorly constrained variability parameter outweighed the bias it removed. Only in Scenario 6 did the new model still have the lower RMSE for C (6.83 against 8.14), with a bias of –0.56 against –7.00 for the old model, and even there its advantage was much smaller than with 420 trials (2.46 against 6.84). In Scenario 8, all three models had a similar RMSE for C with 140 trials (5.72, 5.77, and 6.93), and the t_0 -variability model retained its advantage only for t_0 (6.55 ms against 8.25 ms for the old model). With few trials, the variability models therefore pay for their extra parameter in precision, and this cost is recovered only when variability is strong.

Discussion

The present study introduced and evaluated a variability-augmented extension of a behavioural TVA count model in which trial-to-trial fluctuations in processing speed are modelled explicitly. The main result is that this extension is useful under specific conditions rather than universally. As shown in Table 1 and Figures 3 and 4, the new model was not systematically favoured when the generating process was effectively fixed-speed, was not clearly supported when variability was only mild, and became increasingly advantageous as the true level of trial-to-trial variability increased. This pattern is important because it shows both

Table 6 Effect of the number of trials per dataset (140 vs 420; 25 datasets per scenario and trial count). Columns 3 and 4 give the number of datasets in which the variability model under evaluation, the new model in Scenarios 2, 5, and 6 and the t_0 -variability model in Scenario 8, was preferred over the old model. In Scenarios 5, 6, and 8 this model matches the generating truth; Scenario 2 contains no variability, so its counts are the false-alarm rate of the new model. The remaining columns give the RMSE and the bias of C (letters per second) for the three fitted models

Scenario	Trials	Var. model wins AIC	Var. model wins BIC	RMSE C_{old}	RMSE C_{new}	RMSE C_{t_0}	Bias C_{old}	Bias C_{new}	Bias C_{t_0}
Scenario 2: old, typical C	140	1/25	1/25	3.1607	5.7046	4.8249	1.0303	2.8549	2.7865
	420	1/25	0/25	2.3461	2.7354	2.5969	0.4347	1.2283	0.7314
Scenario 5: new, moderate variability	140	9/25	3/25	4.1262	5.1009	4.2727	-2.4780	1.3213	-1.3073
	420	14/25	7/25	4.2137	3.0167	3.8052	-3.6697	-0.5960	-3.1682
Scenario 6: new, strong variability	140	18/25	9/25	8.1365	6.8291	7.6942	-6.9962	-0.5646	-6.1497
	420	25/25	22/25	6.8408	2.4605	6.7629	-6.6528	0.2558	-6.5518
Scenario 8: t_0 var., strong	140	21/25	17/25	5.7199	5.7738	6.9254	-4.3792	-2.4623	2.1557
	420	25/25	23/25	5.0400	3.2421	2.9868	-4.7503	-1.9423	-0.0455

specificity and sensitivity. The new model did not simply dominate because it had an additional parameter; it became beneficial once the data were generated under conditions in which the fixed-speed assumption became too restrictive. A second result concerns the relation between the new model and the trial-to-trial t_0 -variability model of Dyrholm et al. (2011), which is the formulation most commonly fitted to whole-report data today. The two models proved to be distinguishable from whole-report data and, more than that, asymmetric. The t_0 -variability model did not register speed variability at all: on data with strong speed variability it was preferred over the old model in none of the 25 datasets, and its estimates of C carried the same bias as those of the old model. The new model, in turn, registered part of the threshold variability in the t_0 -variability scenarios as spurious speed variability, but in the direct comparison the correctly specified model was preferred in 24 or 25 of 25 datasets whenever variability was strong (Table 2, Table 3).

This pattern is consistent with the broader logic of TVA as a computational framework. Classical TVA has been valuable because it provides interpretable parameters for processing speed, storage capacity, and threshold, and because these parameters can be estimated from behavioural report data within an explicit likelihood-based structure (Bundesen, 1990, 1998; Bundesen et al., 2005; Kyllingsbæk, 2006). At the same time, the usefulness of such parameters depends on the adequacy of the assumptions under which they are estimated. In the present case, the relevant assumption is not the general TVA architecture itself, but the more specific behavioural implementation in which processing speed is treated as fixed within the modelled condition structure. Our results suggest that this implementation remains appropriate when the data are generated under effectively stable processing speed, but may become too rigid when appreciable trial-to-trial variability is present.

The recovery results strengthen this interpretation. As shown in Table 4 and Figure 5, the old model recovered the generating parameters as well as or better than the new model in the old-truth scenarios, especially for C and t_0 . This is the expected and desirable result if the extension is not merely overfitting. By contrast, in the moderate- and strong-variability scenarios, the new model improved recovery of C and t_0 , and under strong variability it also improved recovery of K . Inspection of the replication-level correlation between estimated C and $\hat{C}_C$ under the new model revealed a positive association that strengthened with true variability, ranging from 0.27 to 0.55 in the old-truth scenarios to 0.60 to 0.72 in the mild-, moderate-, and strong-variability scenarios. This positive correlation indicates that replications yielding higher estimated processing speed also tended to yield higher estimated variability, suggesting that the two parameters capture partially overlapping aspects of the

score distribution rather than compensating for each other. The same coupling explains the small upward bias of the new model in C on data without variability (0.36 to 2.93 letters per second, Table 5): when sampling noise is absorbed as spurious variability, the estimate of mean speed rises with it, because a higher mean is needed to produce the same average performance once part of the trials are assumed to be slow. While this does not invalidate the recovery results, it implies that joint uncertainty in C and $\hat{C}_C$ is larger than their marginal RMSEs alone suggest, and that individual-level estimates of these two parameters should be interpreted with this dependence in mind. The strong-variability scenario is particularly informative because the advantage was visible in both model selection and parameter recovery. Under those conditions, the old model no longer provided only a simpler representation of the same process; instead, it became measurably less accurate as a model of the generating mechanism.

The cross-fitting results add a practical point to this picture. When speed variability is present, the bias of the fixed-speed model in C (−6.65 letters per second in Scenario 6, about 16% of the true value) is not removed by switching to the t_0 -variability model, which produced a bias of −6.55 on the same data. A researcher who fits the standard t_0 -variability model and finds it adequate has therefore not ruled out speed variability, and the estimate of C may still be biased downward. When threshold variability is present, the situation is reversed: the fixed-speed model then underestimates t_0 (by 6.22 ms in Scenario 8, about a third of the true value) as well as C . The new model removes part of the bias in C but leaves most of the bias in t_0 , and only the t_0 -variability model is unbiased for both. The two extensions are thus complementary rather than interchangeable. The reason lies in the different signatures shown in Figure 1b. Speed variability enters the race multiplicatively and is expressed at long exposures, where it depresses the approach to the asymptote and increases the dispersion of scores; threshold variability enters additively and is expressed at the shortest exposures, where it smooths the onset of encoding. Because the present design covers both regions, the two forms can be told apart, and the asymmetry in absorption follows from the same picture: a heavy-tailed speed distribution can imitate part of the smoothed onset, whereas no distribution of thresholds can produce dispersion at exposures long enough for the threshold not to matter.

The recovery of the variability parameter itself also deserves comment. Figure 7 shows that the new model tracked the increase in true variability across scenarios, but that this recovery was weak in the mild-variability condition and much clearer in the moderate- and strong-variability conditions. This implies that, under the present design, mild trial-to-trial variability in processing speed is not

easily distinguishable from ordinary sampling noise or from small distortions in the core parameters. In methodological terms, the detectability of variability appears to depend not only on whether variability exists, but also on whether it is large enough to leave a stable signature in the behavioural score distributions. That finding helps explain why the new model was not selected consistently in the mild-variability scenario even though it was the true generating model there. The behaviour of the variability parameters on data without variability also deserves comment, because it could be mistaken for a defect of the estimator. In the old-truth scenarios, about half of the estimates of cv_C were at the lower bound, and the rest were positive, so that the mean estimate exceeded the true value of zero. This is the expected behaviour of a maximum-likelihood estimate whose true value lies on the boundary of the parameter space: noise that resembles overdispersion is absorbed as positive variability; noise in the opposite direction cannot be absorbed. The same argument predicts the rate at which the information criteria should prefer a variability model when none is present, 7.9% for AIC and 0.7% for BIC with 420 trials, and the observed rates (9.3% and 0% for the new model, 4.0% and 0% for the t_0 -variability model) agree with these predictions. The upper bound of the parameter space, by contrast, played no role: no estimate of cv_C or σ_{t_0} came close to it in any dataset.

These results are also relevant to previous methodological discussions within the TVA literature. Earlier work has already shown that introducing trial-by-trial parameter variability can matter for estimation and can alter the interpretation of fitted TVA parameters (Dyrholm et al., 2011). However, prior work did not establish that variability in processing speed should always be modelled in every behavioural application, nor did it imply that a fixed-parameter model is generally inadequate. The present findings fit that more cautious view. They suggest that explicit modelling of trial-to-trial variability in processing speed can be useful, but only in regimes where the variability is sufficiently pronounced to justify the added complexity. In this respect, the current model is better understood as a conditional extension of behavioural TVA rather than as a replacement for the classical framework. The present results also place the extension next to the existing generalisation of Dyrholm et al. (2011) rather than in competition with it. Trial-to-trial variation in processing rates has been modelled before with a gamma distribution in a TVA-based account of single-item recognition (Blurton et al., 2020), and single-trial analyses have shown that processing speed is not constant within a session (Dyrholm et al., 2015). The present work adds a closed-form count model for the whole report in which such variability is carried by a single additional parameter, together with evidence that this parameter captures

something the t_0 -variability model does not. Whether a model with variability in both t_0 and processing speed can be estimated from whole-report data of ordinary size is a separate question that the present design was not built to answer.

The reduced-trial run bears on the question of when the extension should be used. Adding a variability parameter is not free, a point made forcefully in the diffusion-model literature, where across-trial variability parameters introduced on good theoretical grounds turned out to be poorly constrained by data sets of the size typically collected (Boehm et al., 2018). The same trade-off appears here. With 420 trials, the new model reduced the root-mean-square error of C from 4.21 to 3.02 under moderate variability and from 6.84 to 2.46 under strong variability. With 140 trials, it still removed the bias, but its error in C was larger than that of the old model under moderate variability (5.10 against 4.13) and only modestly smaller under strong variability (6.83 against 8.14), and on data without variability its error was almost twice that of the old model (5.70 against 3.16). The practical recommendation that follows is conditional on the data. With several hundred trials per participant and a design that includes both short and long exposures, the variability parameter is worth estimating, and AIC and BIC will indicate whether it is needed. With trial numbers on the order of one hundred per participant, as is common in clinical work, the fixed-speed model is the better default unless there is an independent reason to expect large variability, because the precision lost to the extra parameter is likely to outweigh the bias it removes.

Two further points concern the size and the consequence of the variability that was simulated. The three values of cv_C imply that 95% of trial-level speeds lie within 0.73 to 1.31, 0.50 to 1.67, and 0.32 to 2.06 times the mean (Figure 1a). Whether variability of the upper end of this range occurs in practice is an empirical question that whole-report data have not yet answered; the single-trial evidence available so far comes from single-item recognition (Dyrholm et al., 2015). The values were chosen to bracket the range over which the parameter passes from undetectable to clearly detectable with 420 trials, not as claims about typical observers, and the labels mild, moderate, and strong should be read in that sense. As to the consequence, the bias of the fixed-speed model in C matters for comparisons between groups or conditions only if these differ in the amount of variability. If two groups share the same cv_C , the bias is common to both and cancels in their difference. The case of concern is the one in which a clinical or ageing group shows more trial-to-trial fluctuation than a control group; the fixed-speed model would then report part of that difference as a difference in mean processing speed.

At the same time, the present findings should be interpreted within the limits of the current design. First, the

simulation study was deliberately targeted at a count-based whole-report setting without distractors, at two display sizes. Accordingly, the present model does not address selectivity parameters such as α , which in standard TVA applications require partial report with distractors, nor does it address persistence parameters such as μ , which require unmasked or iconic-memory manipulations (Bundesen et al., 2005; Kyllingsbæk, 2006). The contribution of the present work is therefore narrower: it concerns the behaviour of C , K , and t_0 in a count-based report model when processing speed may vary across trials. An extension to partial report would require the per-letter rates to differ between targets and distractors, with the trial-level fluctuation acting on all rates at once; the marginalisation of Eq. 15 carries over, but the completed-count distribution is no longer binomial, and we leave this to future work.

Second, the present study is a simulation-based proof of principle rather than an empirical validation on human data. This means that the new model was evaluated under controlled generating conditions, including conditions that matched its own assumptions. Such an approach is appropriate for asking whether the model can, in principle, recover the intended structure and whether it avoids spurious dominance when that structure is absent. However, it does not by itself establish that the same pattern will necessarily hold in empirical datasets, where additional sources of misspecification may be present. For that reason, the current findings support the new model as a methodological extension that is worth testing on real data, not as a definitive statement that behavioural TVA data in general are governed by trial-to-trial variability in processing speed.

Third, several choices of the simulation set limits on what can be concluded. The speed-variability data were generated from the same gamma distribution that the new model assumes, so the recovery results for that model describe the best case; how the gamma model behaves when the true distribution of speeds has a different shape (for example, log-normal) was not examined. The t_0 -variability data were generated from, and fitted with, a normal distribution of thresholds, following (Dyrholm et al., 2011); with a mean of 18 ms and a standard deviation of 20 ms, about 18% of the trial-level thresholds were negative, which is a property of that distributional choice rather than a claim about the perceptual process, and a truncated or skewed distribution would remove it at the cost of a closed-form mean. The identifiability of cv_C and σ_{t_0} shown here is empirical, for one exposure range and one pair of display sizes, and not a formal result. With 25 datasets per scenario, proportions carry confidence intervals of roughly ± 0.15 to 0.20, so that the transitional scenarios in particular should be read as indicating a region rather than a point. Finally, the estimate of σ_{t_0} under its own model fell to the lower bound in 6 of

25 datasets of the moderate t_0 -variability scenario and was biased downward by about 3 ms there, which shows that the boundary problem discussed above affects both variability parameters and is not specific to the new model.

The results support a restrained interpretation. The old model remains a reasonable and parsimonious description when the generating process is effectively fixed-speed. The new model becomes useful when trial-to-trial variability in processing speed is appreciable, and its strongest advantage emerges when such variability is large enough to distort recovery under the fixed-speed formulation. The t_0 -variability model, which is the current standard, does not cover this case, and the new model does not cover the case of threshold variability; each detects and corrects the form of variability it was built for. Thus, the contribution of the present work is not to show that standard behavioural TVA is broadly deficient, but rather to identify a specific regime in which a fixed-speed count model may become too restrictive and to provide a mathematically explicit extension for that regime.

Conclusion

We proposed a variability-augmented extension of a behavioural TVA count model in which processing speed varies across trials. The simulation results indicate that this extension is not uniformly superior to the standard fixed-speed model. Instead, its value is conditional. When the generating process is effectively fixed-speed, the old model remains appropriate and is generally preferred once model complexity is penalised. When trial-to-trial variability in processing speed is strong, however, the new model improves model selection and parameter recovery, particularly for C and t_0 ; at moderate variability the evidence with 420 trials is transitional, and with fewer trials the precision lost to the additional parameter can outweigh the bias it removes. The extension is complementary to the established t_0 -variability model: the two forms of variability leave different traces in whole-report data, the two models are distinguishable from such data, and neither can stand in for the other. The choice between them, and the decision to fit either, should therefore be made on the data at hand, using the information criteria and the number of trials as the guide.

Acknowledgements This study has received partial funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101034319 and from the European Union-NextGenerationEU. The author is currently employed as a postdoctoral researcher at the University of Jyväskylä, where the revision was conducted. The author would like to extend their gratitude to the reviewers and editor who suggested improvements during the revision process and helped improve the current version of the manuscript.

Author Contributions All contribution goes to AK

Data Availability All data generated in this research is available at <https://github.com/Aaronkhodami/Extending-TVA-With-Trial-to-Trial-Variability-in-Processing-Speed-A-Monte-Carlo-Simulation-Study>

Declarations

Ethical Approval and Participant Consent Not applicable.

Conflicts of Interest The author declares no conflict of interest.

Competing Interests The authors declare no competing interests.

The Use of Generative Artificial Intelligence (AI) The author confirms that Generative AI tools were not used in the drafting, compilation, or analysis of this manuscript. The LaTeX version was prepared using the Prism application provided by OpenAI [<https://prism.openai.com>] to create the LaTeX file. This application was solely utilised for formatting purposes and did not contribute to the analysis, writing, or conclusions presented in this work. Language Editing was done using Writefull and Grammarly.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Akaike, H. (2025). Akaike's information criterion. In: M. Lovric (Eds.) *International encyclopedia of statistical science*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-69359-9_14
- Blurton, S. P., Kyllingsbæk, S., Nielsen, C. S., & Bundesen, C. (2020). A Poisson random walk model of response times. *Psychological Review*, 127(3), 362–411. <https://doi.org/10.1037/rev0000179>
- Boehm, U., Annis, J., Frank, M. J., Hawkins, G. E., Heathcote, A., Kellen, D., & Wagenmakers, E.-J. (2018). Estimating across-trial variability parameters of the Diffusion Decision Model: Expert advice and recommendations. *Journal of Mathematical Psychology*, 87, 46–75. <https://doi.org/10.1016/j.jmp.2018.09.004>
- Bundesen, C., & Habekost, T. (2014). *Theory of visual attention (tva)*. <https://doi.org/10.1093/oxfordhb/9780199675111.013.024>
- Bundesen, C. (1990). A theory of visual attention. *Psychological Review*. <https://doi.org/10.1037/0033-295X.97.4.523>
- Bundesen, C. (1998). Visual selective attention: Outlines of a choice model, a race model and a computational theory. *Visual Cognition*. <https://doi.org/10.1080/713756774>
- Bundesen, C., Habekost, T., & Kyllingsbæk, S. (2005). A neural theory of visual attention: Bridging cognition and neurophysiology. *Psychological Review*, 112(2), 291–328. <https://doi.org/10.1037/0033-295X.112.2.291>
- Bundesen, C., Vangkilde, S., & Petersen, A. (2015). Recent developments in a computational theory of visual attention (TVA). *Vision Research*, 116, 210–218. <https://doi.org/10.1016/j.visres.2014.11.005>
- Chechacz, M., Gillebert, C., Vangkilde, S., Petersen, A., & Humphreys, G. (2015). Structural variability within frontoparietal networks and individual differences in attentional functions: An approach using the theory of visual attention. *The Journal of Neuroscience*, 35, 10647–10658. <https://doi.org/10.1523/jneurosci.0210-15.2015>
- Dyrholm, M., Kyllingsbæk, S., Espeseth, T., & Bundesen, C. (2011). Generalizing parametric models by introducing trial-by-trial parameter variability: The case of TVA. *Journal of Mathematical Psychology*. <https://doi.org/10.1016/j.jmp.2011.08.005>
- Dyrholm, M., Vangkilde, S., & Bundesen, C. (2015). Beyond trial types. *Psychological Research*, 79(3), 425–431. <https://doi.org/10.1007/s00426-014-0570-8>
- Fei, Y., Rong, G., Wang, B., & Wang, W. (2014). Parallel L-BFGS-B algorithm on GPU. *Computers & Graphics*. <https://doi.org/10.1016/j.cag.2014.01.002>
- Habekost, T. (2015). Clinical TVA-based studies: A general review. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2015.00290>
- Habekost, T., & Starrfelt, R. (2009). Visual attention capacity: A review of TVA-based patient studies. *Scandinavian Journal of Psychology*, 50(1), 23–32. <https://doi.org/10.1111/j.1467-9450.2008.00681.x>
- Kyllingsbæk, S. (2006). Modeling visual attention. *Behavior Research Methods*. <https://doi.org/10.3758/BF03192757>
- Neath, A. A., & Cavanaugh, J. E. (2011). The Bayesian information criterion: Background, derivation, and applications. *WIREs Computational Statistics*. <https://doi.org/10.1002/wics.199>
- Self, S. G., & Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398), 605–610. <https://doi.org/10.1080/01621459.1987.10478472>
- Vangkilde, S., Petersen, A., & Bundesen, C. (2013). Temporal expectancy in the context of a theory of visual attention. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368. <https://doi.org/10.1098/rstb.2013.0054>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.