

Variation in the “coefficient of variation”: rethinking the violation of the scalar property in time-duration judgments

Yue Ren¹, Fredrik Allenmark¹, Hermann J. Müller¹, Zhuanghua Shi^{1*}

1. General and Experimental Psychology, Psychology Department, LMU Munich

* correspondent author: shi@lmu.de

Abstract

The coefficient of variation (CV), also known as relative standard deviation, has been used to measure the constancy of the Weber fraction, a key signature of efficient neural coding in time perception. It has long been debated whether or not duration judgments follow Weber’s law, with arguments based on examinations of the CV. However, what has been largely ignored in this debate is that the observed CVs may be modulated by temporal context and decision uncertainty, thus questioning conclusions based on this measure. Here, we used a temporal reproduction paradigm to examine the variation of the CV with two types of temporal context: full-range mixed vs. sub-range blocked intervals, separately for intervals presented in the visual and auditory modalities. We found a strong contextual modulation of both interval-duration reproductions and the observed CVs. We then applied a two-stage Bayesian model to predict those variations. Without assuming a violation of the constancy of the Weber fraction, our model successfully predicted the central-tendency effect and the variation in the CV. Our findings and modeling results indicate that both the accuracy and precision of our timing behavior are highly dependent on the temporal context and decision uncertainty. And, critically, they advise caution with using variations of the CV to reject the constancy of the Weber fraction of duration estimation.

Keywords: Coefficient of Variation (CV), Weber fraction, Duration Reproduction, Bayesian model

1. Introduction

Maintaining both high accuracy and precision of timing performance, particularly in the range of milliseconds to seconds, is fundamental to our basic functioning and survival (Buhusi & Meck, 2005; Meck, 1983). Most timing theories explicitly incorporate Weber scaling in their models (Church et al., 1994): the standard deviation of interval estimation is proportional to the absolute time interval – a characteristic also referred to as the scalar property. For example, the information-processing model of “scalar timing theory” (STT) (Gibbon et al., 1984; Gibbon & Church, 1990) adopts the framework of the classical internal-clock model (Treisman, 1963), which assumes a pacemaker-accumulator that linearly records timed durations. In the STT, the scalar property arises mainly from the variability in the memory transformation of the accumulated ticks (Gibbon, 1991). The scalar property has been confirmed in many animal studies (Gibbon et al., 1997) and in human time perception (Wearden & Lejeune, 2008), and recently has been shown to be an emergent property in artificial perceptron neurons (Buhusi & Oprisan, 2013).

One way of testing the scalar property is to use the estimated standard deviation (SD) and mean (M) from each duration timed to construct a coefficient of variance ($CV=SD/M$). The scalar property requires the CV to be constant across the tested range of time samples. Reviewing studies of human timing, Wearden and Lejeune (2008) found that many studies, varying from time reproduction to verbal estimation, confirmed the scalar property. However, violations of the scalar property have been observed when the durations timed were ultrashort or with tasks varying in difficulty or with extensive training (Bizo et al., 2006; Grondin & Killeen, 2009; Matthews & Grondin, 2012). For example, when the range of to-be-estimated durations was rather broad (from 68 ms to 16 minutes), the observed CVs decreased as duration increased (Lewis & Miall, 2009). Also, it is known that the variability is higher for very short durations (< 100 ms) relative to long durations, due to the sensory limits and temporal summation (Scharnowski et al., 2007). Violations of the scalar property manifest mainly in a change of the CV across different time ranges, where the CV is determined by two parameters, SD and M . Since both parameters are susceptible to bias in certain contexts, errors may creep in the estimation of the CV. For example, the range of to-be-tested durations can heavily influence duration estimation, which is known as the central-tendency effect (Bausenhardt et al., 2014; Gu et al., 2016; Gu & Meck, 2011; Jazayeri & Shadlen, 2010; Lejeune & Wearden, 2009; Shi, Church, et al., 2013): short durations are often overestimated and long durations underestimated. Thus, the mean estimate for the short/long durations would be larger/smaller than expected, resulting in the variation of the CV across different durations. Similarly, sensory limits in the short-range and decision noise, in general, could affect the observed SDs. As a consequence, the variation of the CVs may well be attributable to context modulation and/or noise perturbation, rather than reflect true violations of the scalar property.

Accordingly, the present study aimed to investigate whether the variation of the CV could be explained solely by contextual modulation and sensory and decision noise, without resorting to “violations” of the scalar property. We adopted the duration-reproduction paradigm, which has been used in previous studies for demonstrating violations of the scalar property (e.g., Lewis & Miall, 2009). We hypothesized that a large range of test durations would impact the variation of the CV more than a small range, given that the central-tendency effect would be stronger for extreme durations. As the central-

tendency effect can be predicted quantitatively by the Bayesian inference framework (Jazayeri & Shadlen, 2010; Petzschner et al., 2015; Shi, Church, et al., 2013; Shi & Burr, 2016), we also modeled the reproduced duration using Bayesian models. Importantly, we incorporated the scalar property in the model, that is: we assume there is *no* violation of the scalar property. If the model with the scalar property can predict the variation of the CV, relying on the variation of the CV alone would be *insufficient* to falsify the scalar property.

In more detail, we asked participants to reproduce the same physical intervals presented under two different conditions. In the whole-range (“mixed”) condition, intervals from 300 ms to 16 seconds were randomly intermixed, while in the other, sub-range (“blocked”) conditions, intervals were separately tested within sub-second, second, and supra-second (sub-)ranges. In addition, we tested reproduction in both the auditory and visual modalities, given that subjective durations are known to differ between visual and auditory signals (Ganzenmüller et al., 2012; Matthews & Meck, 2014; Shi, Ganzenmüller, et al., 2013; Wearden, 2006; Wearden et al., 1998): temporal precision is higher for the auditory system. Finding essentially similar result patterns would strengthen the generalizability of any conclusions. We hypothesized that the CV would show greater variation in the “mixed”, whole-range versus the “blocked”, sub-range conditions, and in vision as compared to audition. And we expected the Bayesian model with the scalar property would predict those variations in CVs.

2. Method

2.1. Participants

A total of 52 volunteers (13 participants for each experimental session), aged 21–33 years (27 females), were recruited from the subject pool of LMU Munich Psychology Department. This number was based on the sample sizes in previous duration-reproduction studies (Cicchini et al., 2012; Jazayeri & Shadlen, 2010; Lewis & Miall, 2009), which ranged from 5 to 14 participants. Our participants had all normal or corrected-to-normal vision, normal hearing, and no somatosensory disorders. Participants provided written informed consent in accordance with the Declaration of Helsinki prior to the experiment and received 9 Euro/hour for their participation.

2.2. Apparatus and Stimuli

The study was conducted in a dimly lit, sound-attenuated laboratory cabin. Visual stimuli were displayed on a 21-inch CRT monitor with a refresh rate of 100 Hz and a resolution of 800 x 600 pixels. In visual sessions, targets were grey squares ($8^\circ \times 8^\circ$ of visual angle), with two levels of luminance: grey (17.5 cd/m^2) and white (42.7 cd/m^2), presented on a black screen background (1.60 cd/m^2). In Experiments 3 and 4 (auditory sessions), a natural water-flow sound (60 dB, measured at the sound source) was chosen for auditory presentation. Compared to simple sine waves, the water-flow sound potentially reduces fatigue and discomfort during longer stimulus presentations (e.g., 16 s). Auditory stimuli were delivered via two loudspeakers placed on both sides of the computer screen (with a separation of 40 cm). The

experimental presentations were generated in Matlab (version 2015a) and with the Psychtoolbox-3 toolbox.

2.3. Experimental Procedure

We adopted a between-subject design considering two experimental factors: Temporal Context (“Mix” versus “Block(ed)”) and Modality (“Vision” vs. “Audition”). Accordingly, four separate sessions were conducted with separate groups of participants, hereafter referred to as: “Vis/Mix”, “Vis/Block(ed)”, “Aud/Mix”, “Aud/Block(ed)”. We employed a temporal reproduction task (Lewis & Miall, 2009), in which participants first received a stimulus (a white square or a sound) for a given duration, and then were asked to reproduce that duration by pressing a response key for as long as they had perceived the duration.

2.3.1. The visual sessions (“Vis./Mix” and “Vis./Block(ed)”)

Each trial started with the word prompt “Presentation” shown for 300 ms, indicating the initial presentation of the experiment. This cue was followed by a grey square presented at the center of the display, to which participants were instructed to press the left arrow key with the left index finger when they were ready to start the presentation. Pressing of the key triggered a color change of the square from grey to white. Participants had been told that the duration of the white square was the “target duration” that they had to remember and to reproduce. The duration of the white square was selected from nine intervals (separated equally on the logarithmic scale): 0.30, 0.49, 0.81, 1.33, 2.19, 3.60, 5.92, 9.73, and 16.00 s. At the end of the interval, the color of the square changed (automatically) back to the initial grey, upon which participants had to release the left arrow key. That is, participants were required to hold the key throughout the presentation of the white square and only release it at the end of its duration, signaled by its reversion to a grey square. We kept this production procedure the same as the previous study (Lewis & Miall, 2009) for the purpose of cross-study comparison. The reproduction phase was separated from the target presentation by a 250-ms blank screen, upon which the word “Reproduction” appeared at the screen center for 300 ms. Immediately after this verbal signal, a grey square appeared, prompting participants to start reproducing the given target duration by pressing the right arrow key with their right index finger; participants proceeded to the reproduction at their own pace. Their keypress, again, triggered the change of the square’s color from grey to white. Participants were instructed to keep pressing the key for as long as they had perceived the target duration. The key release triggered a color change back to grey. The next trial started following a random blank interval varying from 500 to 1000 ms.

The same procedure was applied to both visual sessions, the only difference being the exposure to different temporal contexts. In the “Vis./Mix” condition, nine intervals were randomly mixed within each block, whereas in the “Vis./Block(ed)” condition, they were divided into three sub-groups according to their temporal range: the “Short” (0.30, 0.49, 0.81 s), “Intermediate” (1.33, 2.19, 3.60 s), and “Long” groups (5.92, 9.73, 16.00 s). Each experimental session consisted of 15 blocks of 18 trials each.

2.3.2. The auditory sessions (“Aud./Mix” and “Aud./Block(ed)”)

The same paradigm was adopted for auditory sessions, in which auditory excerpts of a natural water-flow sound were used for duration presentation. On each trial, after the visual cues “Presentation”

and, respectively, “Reproduction” (indicating the initiation of the respective trial phase) for 300 ms, a fixation cross was presented in the center of the display until the end of the phase. In the presentation phase, participants were asked to press the left arrow key with the left index finger when they were ready to start the trial. Pressing of the key triggered the water-flow sound for a given duration. Again, participants were required to hold down the key throughout the presentation of the water-flow sound and only release it once the sound stopped. In the reproduction phase, participants were instructed to self-initiate reproduction by pressing the right arrow key with the right index finger, which triggered the water-flow sound. Participants were instructed to keep pressing the key for as long as they had perceived the target duration.

Participants underwent three blocks of training (9 trials per block) prior to performing the experiment proper, which took some 10–15 minutes to complete. During training, participants’ reproduction errors that exceeded 50% of the sample duration were followed by a “warning” message, “Too Short” or “Too Long”, respectively. Participants were told to perform the task avoiding any form of counting. No feedback was given during the formal experiment. To avoid contamination of different temporal contexts, each participant took part in only one experiment in this study, which lasted about 1.5 hours. Participants were free to take a break between blocks when needed to prevent loss of concentration and alertness.

2.4. Bayesian Modeling

Here, we propose a two-stage Bayesian Estimator to model performance in the temporal reproduction task. Figure 1 illustrates the computation processes of the model; the computational steps are as follows:

Stage I: duration coding. Following classical psychophysics (Fechner, 1860), we assume the internal duration coding inherits the scalar property from Weber’s law: the just noticeable difference is proportional to the absolute magnitude. Accordingly, we introduced a logarithmic transformation of a given sample interval D to the internally measured time:

$$S = \log(D) + \epsilon, \quad (1)$$

where D is the sample duration on the linear scale and S the internal representation of measured duration on the logarithmic scale. The random variable ϵ represents normally distributed internal-measurement noise.

Because sensory input is noisy, duration estimates can be improved by taking into account the prior probability of encountering a particular duration. In general, when an ideal observer follows Bayesian integration for the perceived duration, where both the prior and the likelihood are independent Gaussians, the optimal internal estimate μ_{X_i} for a given interval is essentially a weighted sum of the interval measure S_i and the mean of the (biased) prior (Cicchini et al., 2012; Jazayeri & Shadlen, 2010; Petzschner et al., 2015; Shi, Church, et al., 2013):

$$\mu_{X_i} = (1 - \omega_p)S_i + \omega_p(\mu_p + \Delta_l), \quad (2)$$

and its variance is

$$\sigma_{X_i}^2 = \frac{\sigma_s^2 \sigma_p^2}{\sigma_p^2 + \sigma_s^2}, \quad (3)$$

where the weight $\omega_p = \frac{1/\sigma_p^2}{1/\sigma_p^2 + 1/\sigma_s^2}$ is proportional to the inverse of the sum of the variances; σ_s^2 is the variance of the sensory measurement, which is constant in log-scale representation for a given modality for each participant (i.e., the scalar property on the linear scale); and σ_p^2 is the variance of the prior. Additionally, we consider a shift term (Δ_I) to incorporate a general bias on the mean of the prior distribution.

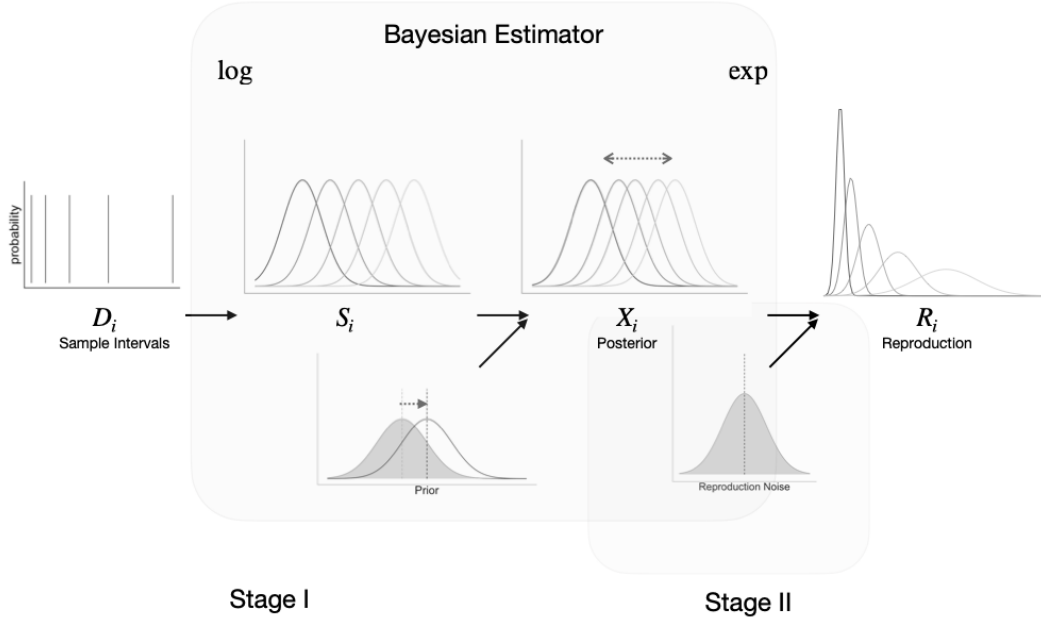


Figure 1: Two-stage Bayesian-Estimator model for temporal duration reproduction. In the first, “presentation” phase, physical time is encoded through a logarithmic transformation, according to Weber-Fechner’s law. The sensory input is then integrated with the prior, which is developed based on the past sample durations encountered. Both the prior and sensory likelihood are assumed to be independent Gaussians. A normally distributed decision noise, which is independent of the sensory magnitudes, contributes to the “reproduction” in the second stage.

Stage II: duration reproduction. Duration estimates are transformed back to a linear scale; accordingly, the variance $\sigma_{X_i}^2$ is transformed into the variance of a log-normal distribution:

$$\sigma_{\hat{X}_i}^2 = |e^{\sigma_{X_i}^2} - 1| e^{2\mu_{X_i} + \sigma_{X_i}^2}. \quad (4)$$

In modeling this phase, we also take into account an additional source of variability resulting from the response uncertainty, which has been considered in previous studies (Bizo et al., 2006; Getty, 1975). We assume this duration-independent variability σ_r remains the same across all tested durations for a given participant, which forms the variance:

$$\sigma_i = \sqrt{\sigma_{\hat{x}_i}^2 + \sigma_r^2}. \quad (5)$$

We also assume that an additional bias may be introduced at the reproduction stage. Thus, the reproduced duration is:

$$R_i = e^{\mu_{x_i} + \Delta_2 + \sigma_{\hat{x}_i}^2/2}. \quad (6)$$

where Δ_2 is the reproduction stage bias. And the observed CV_i of each interval can be calculated as:

$$CV_i = \frac{\sigma_i}{R_i}. \quad (7)$$

In total this model has five parameters: the standard deviations of the sensory noise and the prior (σ_s and σ_p), the two shift terms (Δ_1 and Δ_2) and the stimulus-independent response uncertainty (σ_r). For fitting the data from the mixed conditions the model had only these five parameters. However, for the blocked conditions, we expected there to be contextual influences based on the range of durations within a block (Teghtsoonian & Teghtsoonian, 1978). In order to explore which parameters would be influenced the most by this context, we compared models with each parameter fixed to the same value in each block to models where the parameters could differ between the “short”, “medium” and “long” duration blocks (see Appendix A for details). Thus, in total 32 models were compared. The model fitting was performed in two steps: first, the model predictions for the logarithmic scale mean μ_{x_i} as a function of the log-transformed sample intervals ($\log(D)$) was fitted, assuming a normally distributed likelihood with a constant mean (this was effectively fitting a straight line, in the mixed condition, or three line segments, in the blocked condition). This provided the starting values of the σ_s and σ_p parameters for the full model fit, which was a maximum likelihood fit to the distribution of reproduced durations, assuming that these were normally distributed with the mean given by Eq. 6 and the standard deviation given by Eq. 5 (assuming a normal distribution was an approximation, used in order to simplify the calculations, since the model actually predicted that the distribution should be the convolution of a normal and a log-normal distribution). The optimizations for the maximum likelihood fitting were performed using the “optim” function in R with the “L-BFGS_B” algorithm. If an optimization failed to converge it was attempted again after adding small random numbers to the starting values, and this was repeated until it converged.

3. Results

To exclude duration reproductions likely reflecting lapses of attention or accidental responses, we adopted an outlier criterion based on the interquartile range (IQR): for each participant, at each given duration, reproduced durations exceeding three times the IQR were omitted from further analysis. In addition, one participant from the “Vis/Block” condition and two from the “Aud/Mix” condition were excluded, due to their outliers exceeding 20% of the total trials.

3.1. Mean reproduction errors and CV

We measured the *Relative Reproduction Error (RRE)* by calculating the difference between the actual response (R_i) and the sample duration (D_i), normalized by the sample duration ($RRE = \frac{R_i - D_i}{D_i}$). This metric provides a measure of the degree of the estimation bias, which is comparable across different durations. Figure 2 shows both the mean RRE (colored dots) and the model fitting (colored lines) for Experiments 1 to 4, separately for the auditory and visual reproductions. By visual inspection, the RREs exhibit different patterns with respect to the modality of interval presentation. The mean reproductions show larger biases for the visual presentation as compared to the auditory presentation. In addition, the mean RREs show overestimations for both the “short” and “intermediate” blocks (the first six durations ranging from 300 to 811 ms) but underestimation for the “long” blocks, evidencing a clear central-tendency effect. Moreover, the central-tendency effects likely occurred within the test range - a grand central-tendency effect for the “Mixed” conditions, and three separate effects for the “Blocked” conditions. To confirm the differences in RREs among conditions, we further run ANOVAs on the mean absolute RREs, which only revealed a main effect of Modality, $F(1,45) = 9.10, p < .01, \eta_p^2 = 0.17, BF_{incl} = 13.01$. The main effect of Condition (“Blocked” vs. “Mixed”), $F(1,45) = 1.67, p = .20, \eta_p^2 = 0.04, BF_{incl} = 1.01$, and the interaction between Modality and Condition, $F(1,45) = 1.40, p = .24, \eta_p^2 = 0.03, BF_{incl} = 0.33$, were non-significant, indicative of the main difference in RREs being that between audition and vision.

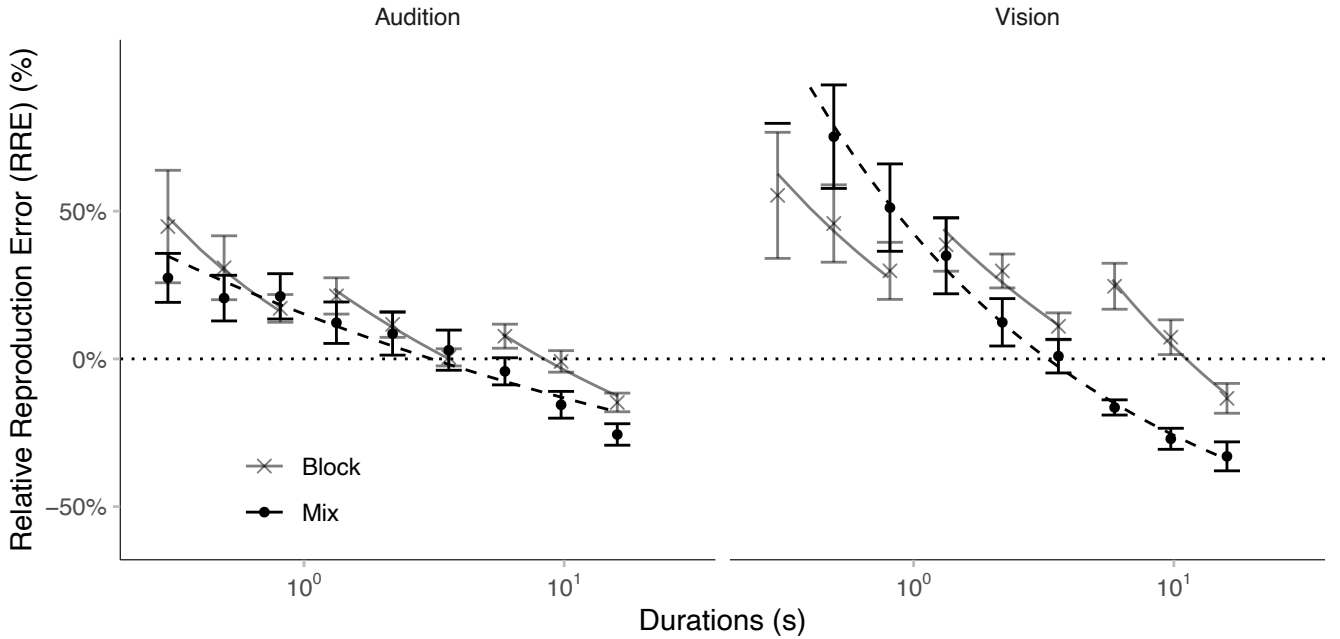


Figure 2. *Relative Reproduction Errors (RREs) (dots and crosses) and RREs predicted by the two-stage model (lines), as a function of the sample interval. The left panel depicts the auditory sessions, the right*

panel the visual sessions. The black dots represent reproductions from “Mixed” conditions and grey crosses from “Blocked” conditions. The dashed lines represent the prediction from “Mixed” condition and solid lines from the “Blocked” conditions.

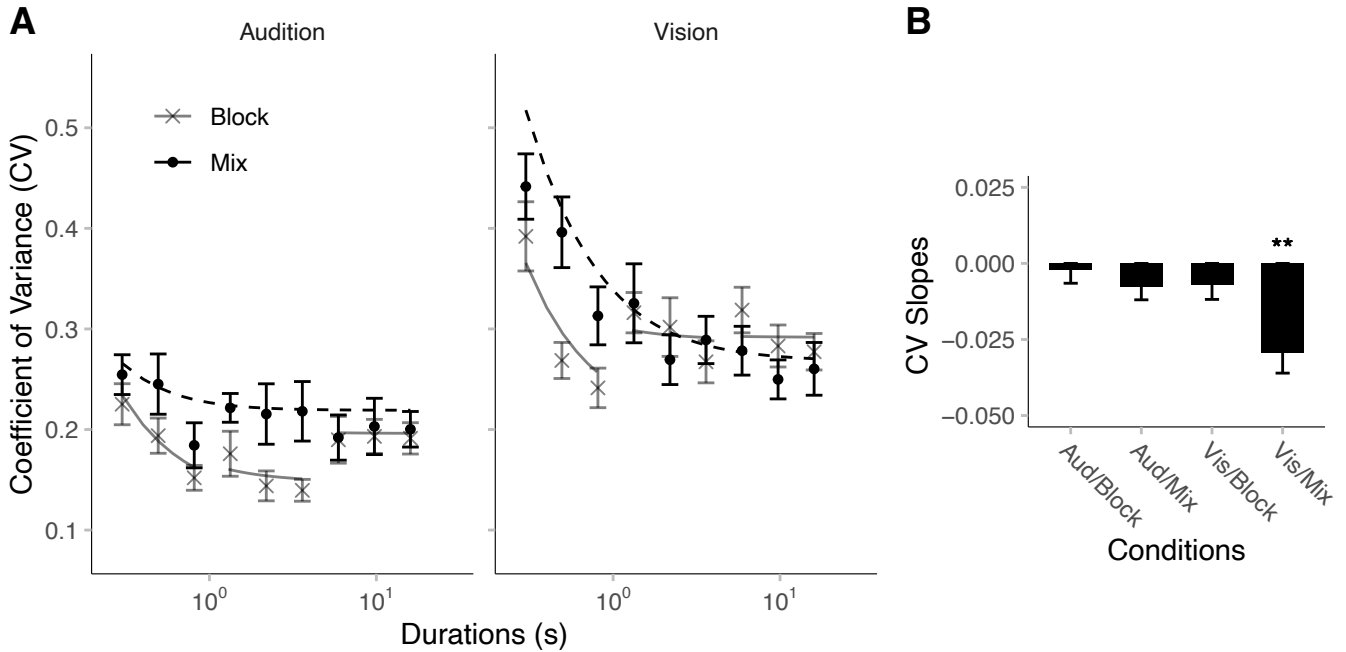


Figure 3: (A) Mean CVs (dots and crosses) as a function of the sample duration, separately for the auditory (left panel) and visual (right panel) sessions. The black dots represent reproductions from “Mixed” conditions and grey crosses from “Blocked” conditions. The dashed lines represent the model predictions for the “Mixed” condition and solid lines for the “Blocked” conditions. (B) Mean CV slopes as a linear function of logarithmic duration in four experimental conditions. Error bars indicate one standard error. The slope was obtained by estimating parameter b of the linear function $CV = a + b \log(\text{Duration})$.

The mean CVs are plotted in Figure 3A as a function of the sample interval, separately for experimental conditions. Similar to mean reproductions, the CVs, too, were influenced by Temporal Context and Modality. Under the “Vis/Mix” condition, the CVs appeared to largely decrease as the sample duration increased, which is consistent with previous reports (Lewis & Miall, 2009). In addition, higher precision of temporal estimation in the auditory modality was evidenced by smaller CV values under both (“Mix(ed)” and “Block(ed)”) Context conditions. To quantify the variations of CVs across time, for each participant, we fitted the decreasing trend (hereafter referred to as “CV slopes”) using simple linear regression: $CV = a + b \cdot \log(\text{Duration})$. The reason for using the linear regression instead of more complex forms of curve fitting was two-fold: (i) we aimed to test whether the CV remained constant across the four experimental conditions; (ii) and the results would be comparable to previous reports (e.g., Lewis & Miall, 2009). Note, though, this analysis does not include the possible “jumps” that the Blocked conditions

introduced (see Figure 3, abrupt changes of the CV across the blocked ranges). We will return to this issue in the next subsection “Model predictions”, which provides more specific psychological grounded model predictions and explanations. The mean CV slopes for four experimental conditions are depicted in Figure 3B.

The estimated slopes were then submitted to an ANOVA with the (between-subject) factors of Modality (Vision vs. Audition) and Temporal context (“Mixed” vs. “Blocked”), which revealed both main effects to be significant: Modality, $F(1,45) = 6.84, p = .01, \eta_p^2 = 0.13, BF_{incl} = 3.72$; Context, $F(1,45) = 7.36, p < .01, \eta_p^2 = 0.14, BF_{incl} = 5.26$; the interaction was non-significant: $F(1,45) = 2.74, p = .11, \eta_p^2 = 0.06, BF_{incl} = 0.96$. Thus, the variation in the CVs is mainly influenced by the two factors: the decrease was larger with visual than with auditory intervals and larger for whole-range mixed versus sub-range blocked interval durations. Notably, however, the CV slope was largest in the “Vis/Mix” condition. Testing the CV slopes against the null-hypothesis of constant CVs across the range of sample durations (i.e., CV slope = 0) for each experiment revealed there to be a significant (decreasing) trend only in the “Vis/Mix” condition, with a slope of -0.03 from the simple linear regression with logarithmically scaled durations, $t(12) = 4.50, p < .001$. For the other three conditions, by contrast, the (numerically decreasing) slopes did not differ reliably from 0. In the next subsection, we look into these modulations of the modality and temporal context from the perspective of cognitive and computational models.

3.2. Model predictions

The proposed Bayesian-Estimator model for predicting reproduction durations and the CVs has five free parameters: the variance of the prior (σ_p^2), the deviation of the mean of the prior from the physical mean (Δ_I), the variance of the likelihood (σ_s^2), the reproduction stage bias (Δ_2), and the variance of the decision noise (σ_r^2). The Bayesian observer model described above yields a prediction that minimizes the expected loss associated with the coefficient of variation and erroneous estimates in duration reproduction. The model comparison revealed that the best model (in terms of the Bayesian Information Criterion, BIC) allowed the reproduction bias (shift term) Δ_2 and the standard deviation of the prior σ_p , but not any of the other three parameters, to differ among blocks in the “Blocked” conditions. That is, these parameters differ from the “Mixed” conditions in that separate σ_p and Δ_2 parameters were used for “short”, “medium” and “long” duration blocks, resulting in nine parameters in total. The predicted reproduction errors of the best model are shown as lines (dashed lines for Mix(ed) and solid lines for “Block(ed)” conditions) in Figure 2, and the model-predicted CVs as curves in Figure 3A. And the estimates of the model parameters for the four experimental conditions are summarized in Table 1.

Table 1. Model parameters of the Bayesian predictions

	Vision				Audition			
	Mix(ed)		Block(ed)		Mix(ed)		Block(ed)	
		Short	Intermediate	Long		Short	Intermediate	Long
ω_p	0.26 \pm 0.17	0.22 \pm 0.18	0.24 \pm 0.15	0.35 \pm 0.14	0.12 \pm 0.07	0.19 \pm 0.23	0.20 \pm 0.13	0.21 \pm 0.16
σ_p	0.73 \pm 0.70	1.42 \pm 1.27	0.95 \pm 0.98	0.58 \pm 0.70	0.68 \pm 0.23	1.38 \pm 1.19	0.80 \pm 0.80	0.94 \pm 0.97
Δ_1	0.01 \pm 0.15		0.18 \pm 3.58		-0.02 \pm 0.15		-2.20 \pm 4.76	
σ_s	0.34 \pm 0.14		0.39 \pm 0.41		0.24 \pm 0.09		0.41 \pm 0.67	
σ_r	0.28 \pm 0.23		0.13 \pm 0.11		0.05 \pm 0.06		0.09 \pm 0.09	
Δ_2	0.04 \pm 0.16	0.32 \pm 0.71	-0.06 \pm 0.83	-0.17 \pm 0.77	0.00 \pm 0.15	0.98 \pm 2.27	0.49 \pm 0.97	0.17 \pm 1.00

Table 1. Model parameters of the Bayesian predictions on CVs averaged across participants in four experimental conditions.

In our model framework, we first adopted two parameters (σ_p and Δ_I) to account for the variability and shift in the prior distribution. The best model showed that the shifts of the prior Δ_I did not differ significantly from 0 in any of the four conditions (one-sample t-test compared to 0: $p = .71$, $BF_{10} = 0.29$ in the “Vis/Mix” condition; $p = .84$, $BF_{10} = 0.33$ in the “Aud/Mix” condition; $p = .92$, $BF_{10} = 0.29$ in the “Aud/Mix” condition; $p = .10$, $BF_{10} = 0.84$ in the “Aud/Block” condition). Moreover, the variability parameter σ_p was not significantly different between the visual and auditory modalities in the “Mixed” condition (two-sample t-test: $p = .81$, $BF_{10} = 0.38$). A further 2×3 mixed ANOVA with the between-subject factor (“Modality”) and the within-subject factor (i.e., the Temporal Range) for the “Blocked” conditions also failed to reveal any significant difference in σ_p between vision and audition, $F(1,23) = 0.04$, $p = .85$, $\eta_p^2 = 0.002$, $BF_{incl} = 0.36$, or Modality \times Temporal Range interaction $F(1.53,35.26) = 0.64$, $p = .49$, $\eta_p^2 = 0.03$, $BF_{incl} = 0.28$. Only the main effect of the Temporal Range turned out significant, $F(1.53,35.26) = 4.09$, $p = .04$, $\eta_p^2 = 0.15$, $BF_{incl} = 2.80$ (degree of freedoms adjusted by Greenhouse-Geisser Sphericity correction, same for the other tests): the variability σ_p decreased as the duration range of the “Blocked” condition increased. The Bayes factors associated with the modality differences can be interpreted in terms of strong evidence of “no difference” in Δ_I and σ_p between the auditory and visual modalities, arguing in favor of an amodal/supramodal representation of the prior. Interestingly, in the “Mixed” conditions, the weight ω_p was higher with visual than with auditory interval presentation (Welch Two Sample t-test: $p = .01$, $BF_{10} = 4.28$), which is consistent with literature that audition has general higher temporal resolution than vision. However, the weights were not different in the “Blocked” conditions (all $ps \geq .1$, $BF_{incl} < 0.74$), in part likely due to the block-wise variation of the temporal ranges.

Second, we assumed no violation of the scalar property (i.e., Weber’s law) during the sensory measurement. In line with this, a 2×2 mixed ANOVA on σ_s with the factors Modality (Vision, Auditory) and Condition (“Mixed”, “Block”) revealed no significant effects (Modality: $F(1,45) = 0.10$, $p =$

.76, $\eta_p^2 = 0.002$, $BF_{incl} = 0.29$; Condition: $F(1,45) = 0.94$, $p = .04$, $\eta_p^2 = 0.02$, $BF_{incl} = 0.42$; interaction, $F(1,45) = 0.28$, $p = .60$, $\eta_p^2 = 0.01$, $BF_{incl} = 0.40$. One key ingredient for capturing the variation of the CV is the noise parameter σ_r , which plays a bigger role with the short as compared to the long durations (see Equation (5) in the Bayesian Modeling section). As shown in Table 1, the mean σ_r is larger for the visual modality relative to the auditory modality, and it is the largest in the “Vis/Mix” condition. A further 2×2 ANOVA with Modality (Vision, Auditory) and Condition (“Mixed”, “Block”) as between-subject factors revealed the main effect of Modality to be significant, $F(1,45) = 11.39$, $p < .01$, $\eta_p^2 = 0.20$, $BF_{incl} = 16.04$, while there was no overall difference between the “Mixed” and “Blocked” conditions, $F(1,45) = 1.72$, $p = .20$, $\eta_p^2 = 0.04$, $BF_{incl} = 0.58$. The Modality \times Condition interaction was significant, $F(1,45) = 5.20$, $p = .03$, $\eta_p^2 = 0.10$, $BF_{incl} = 1.43$. The interaction was attributable mainly to the “Mix(ed)” conditions, with σ_r being significantly larger for the visual (“Vis/Mix”) than for the auditory modality (“Aud/Mix”), $t(22) = 3$, $p < 0.01$. This pattern suggests that a relatively high uncertainty (i.e., larger σ_r) is associated with the visual modality in general, independently of the presented durations: high uncertainty (e.g., in the “Vis/Mix” condition) introduces a large duration-independent noise term in the reproduction.

More interestingly, in the “Blocked” conditions, the shift in the reproduction Δ_2 (i.e., reproduction bias) appears to decrease as the range of durations increases. A 3×2 mixed ANOVA on Δ_2 with the factors Modality (Vision, Auditory) and Temporal Range (“Small”, “Intermediate”, “Large”) revealed a main effect on the within-subject factor Temporal Range, $F(1.07, 24.57) = 6.98$, $p = .01$, $\eta_p^2 = 0.23$, $BF_{incl} = 19.53$, but no interaction, $F(1.07, 24.57) = 0.43$, $p = .53$, $\eta_p^2 = 0.02$, $BF_{incl} = 0.24$, or main effect of Modality, $F(1,23) = 1.32$, $p = .26$, $\eta_p^2 = 0.05$, $BF_{incl} = 0.72$. This captures the “jumps” in the reproduction errors and the CVs in the “Blocked” conditions. The jumps of the CVs, captured by Δ_2 , is likely attributable to the sequential (Cicchini et al., 2018; Fischer & Whitney, 2014), block-by-block variation of the range of durations. A short-range block was most likely preceded by the long- and intermediate-range block, and vice versa. Thus, the tendency of the reproduction was partially carried over across blocks. Interestingly, the best model suggests that the across-block carry-over effect arises in the second stage of the reproduction, rather than the first stage of the duration encoding.

To visualize the goodness of the fit of the best model for individual data, we plotted the predicted RREs versus the observed RREs in Figure 4. As can be seen, the predicted individual RREs lie mostly on the diagonal line. A correlation analysis revealed high correlations across all four experimental conditions ($r^2 \geq .77$). This was also true for the reproduction variability, measured by the standard deviation of the reproduction ($r^2 \geq .88$, see Appendix B for the plot). In summary, the prediction of the best model is in good agreement with the empirical data we observed.

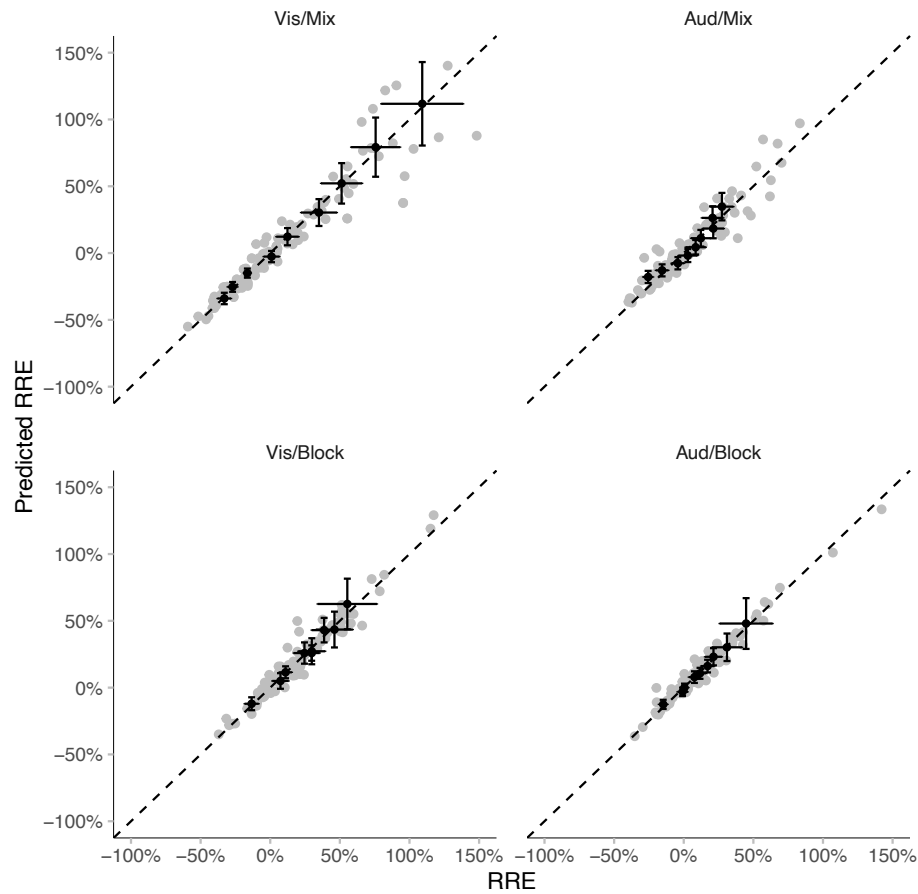


Figure 4. Predicted Relative Reproduction Errors (RREs) vs. the observed RREs for each observer in all conditions (gray dots), with the mean (\pm SE) model prediction for each combination of the test durations and the experimental conditions against the corresponding, observed mean (\pm SE) RREs. The diagonal line shows the ideal model fit (where the predicted RREs match the observed RREs).

4. Discussion

It has been long debated in the literature whether or not the scalar property is violated in duration perception. The “golden” method for testing for violations of the scalar property has been to use the coefficient of variation (CV), which is defined by the observed standard deviation over the observed mean. Yet, little consideration has been given to the fact that the two parameters may be influenced by the experimental context. The present study was designed to examine for such influences by using a duration-reproduction task, asking participants to reproduce a duration selected from a wide range of intervals (from 300 ms to 16 seconds) in two types of temporal context: a full-range mixed and sub-range blocked

condition, and from both the visual and auditory modalities. We observed the classical central-tendency effect in all four conditions, as well as the variation of the CVs. Importantly, without assuming violation of the scalar property, our two-stage Bayesian model successfully predicted the central-tendency effect and the decrease of the CVs along the timeline. Thus, our findings suggest that the observed CV is not strictly tied to the internal scalar property; and using the CV alone to reject the scalar property is not warranted by the behavioral findings, given that both contextual modulation and decision uncertainty play critical roles in time estimation.

4.1. Variation of CVs in behavioral studies

To fully evaluate the goodness of human timing performance, one has to consider both aspects of temporal judgments: the mean estimate and its precision. While the relation between these two should exhibit the scalar property characteristic, there are findings at variance with this (Meck et al., 1984; Wearden, 1992; Wearden et al., 1997). Some of the reported violations – such as a larger estimation bias with ultrashort intervals or lower precision with more difficult tasks (Lejeune & Wearden, 2006; Wearden & Lejeune, 2008) – can be easily accommodated by adding additional constraints (e.g., sensory limits). However, when it comes to systematic violations of scalar timing, such as the continuous decrease of the CVs across a very wide duration range from 68 ms to 16 minutes (Lewis & Miall, 2009), the roles played by context factors (in particular, temporal context and presentation modality) must be taken into consideration. Unfortunately, though, context modulation has been largely neglected in the debate of the variation of CVs; and instead, researchers attempted to tweak parameters of their respective timing models. For instance, Gibbon (1991) argued that the variation of the CV arises when the Poisson variance of the pacemaker and the ratio comparison between the current and remembered durations are allowed to change across durations. Memory-trace models, such as the “multiple time scales” (MTS) model (Staddon & Higa, 1999), argue that the scalar property is closely related to Jost’s logarithmic law of memory forgetting (Ebbinghaus, 1885; Jost, 1897; White, 2001). Slightly tweaking the forgetting slope would predict the observed decreasing CV (Staddon & Higa, 1999). However, even with the adjusted parameters, these models cannot explain why different CVs may be observed for the same physical duration depending on different contexts (in the present study: e.g., for the same visual duration in the full-range vs. the short sub-range condition). In the current study, we first replicated previous findings (Lewis & Miall, 2009), demonstrating a similar decreasing trend of the CVs as duration increased in the visual full-range condition; we then expanded the investigation to two different types of temporal context and two modalities. We observed a reduction of the systematic violation when the sample durations were presented in blocked sub-ranges and when the temporal intervals were delivered via the (higher-precision) auditory modality. These results confirm that both the accuracy and precision of timing performance are dependent on the context factors.

4.2. Bayesian integration with contexts

According to Bayesian theory, perception emerges from the probabilistic inference. The fundamental problem encountered by the brain is to cope with uncertainty from the environment. To minimize uncertainty, the brain needs to make maximal use of the available information, such as

knowledge about previously experienced events and the present sensory inputs. The uncertainty can be optimally reduced (to a minimum) when this information is integrated according to its reliability (Ernst & Di Luca, 2011; Taubert et al., 2016). The Bayesian-Estimator model proposed in the current study makes two adjustments in evaluating the sources of uncertainty arising from both stages of the task (duration production and reproduction), according to different temporal contexts: First, based on the fact that subjective duration can differ between different modalities (e.g., Wearden et al., 1998) and temporal context, we assume that the prior itself can be biased. Thus, we implemented a parameter, Δ_1 , to capture this feature. Second, we consider additional biases (Δ_2) that might occur during the reproduction in Stage II, in particular, carry-over of the response tendency from previous into the current trial blocks. Third, we assume time reproduction is corrupted by some duration-independence uncertainty factor, which is captured by the parameter σ_r . This uncertainty is relatively small and can be neglected when durations are at the super-second level. However, with durations in the sub-second range, this factor has to be taken into account in the model.

Among 32 models we compared, the best model assumes the variability of the prior (σ_p) and the reproduction bias (Δ_2) is influenced by the temporal context (i.e., three separate, narrow sub-ranges in the blocked conditions, as compared to one broader range in the full-range mixed condition). Varying the σ_p allows the model to capture the uncertainty of the “Blocked” temporal context, while the reproduction bias (Δ_2) captures the jumps between blocks. For example, the mean reproduction from both the “Short” and “Intermediate” duration blocks exhibited overall overestimations under “Block(ed)” conditions (see Figure 2). It is possible that these overall overestimations were introduced by the preceding blocks, most likely the “longer” block. This kind of carry-over effect has been found in perceptual judgments, such as serial dependence (Fischer & Whitney, 2014; S. Glasauer & Shi, 2018; Stefan Glasauer, 2019), as well as in temporal reproduction (Wiener et al., 2014). For example, examining the influence of observers’ previous “long” responses on current bisection performance (regardless of later duration presented), Wiener and colleagues (2014) found a strong tendency for observers to carry over responding “long”, thus dissociating response carry-over from perceptual bias in the bisection task. Also, this response carry-over was more marked for the visual as compared to the auditory modality. In our “Blocked” conditions, we found a similar reproduction carry-over effect captured by the reproduction bias (Δ_2), which was largest for the “short” range, and smallest for the “long” range. Interestingly, the estimated values of the Δ_1 parameter in the “Mixed” conditions (“Vix/Mix” and “Aud/Mix”) were close to unbiased (i.e., 0), with little variability across participants, whereas the estimates varied greatly across individuals in the “Blocked” conditions (the standard error was almost tenfold for the “Blocked” as compared to the “Mixed” conditions). This suggests that participants could establish a relatively unbiased prior given a stable temporal context (the durations were randomly sampled from one distribution), while the priors may change dramatically when the environment changes (witness the large variation with the blockwise presentation). Our model thus incorporates this volatility change in determining the perceptual bias and the response carry-over effect in temporal reproduction.

The duration-independent variability that we introduced in the model – parameter σ_r – was assumed to be independent of duration magnitude. Without assuming any violation of the scalar property, this key parameter captures the variation of the CV, which is consistent with the variation of CVs reported

in the literature (Lewis & Miall, 2009). In our model, we considered two sources of variability: the scalar property (the square root of which should linearly increase with the mean) and σ_r . The estimates of σ_r from our model agree with the behavioral observations of the CVs: across the four conditions, the estimated value of σ_r was largest in the “Vis/Mix” condition (see Table 1), indicative of the largest contribution of decision uncertainty to the variance of reproductions in this condition. Importantly, while we have referred to σ_r as decision uncertainty, our key prediction of larger CVs for the shortest duration is not dependent on whether σ_r solely reflects the variance introduced during the decision of when to stop the reproduction or whether there are also contributions from earlier processes, such as “sensory-onset” variance. Rather, our key prediction only requires that σ_r is independent of duration magnitude.

4.3. Modality difference and the central tendency

Cicchini and colleagues (2012) demonstrated a robust audiovisual difference in duration reproduction: participants’ responses to visual stimuli tended to gravitate toward the mean of the sample durations (the central-tendency effect) but remained veridical with auditory stimuli (the latter was the case even though participants differed widely in the level of musical expertise). In the current study, we showed that a central-tendency effect in duration reproduction exists in both presentation modalities, though it is more pronounced in vision than in audition, indicated by the relative RREs - thus confirming the audition-vision difference in time judgments (Cicchini et al., 2012).

Interestingly, though, according to our best model, the prior representation of the range of the test durations does not differ much between the audition and vision: there was no significant difference in the two parameters representing the prior knowledge – that is, σ_p (variance of the prior distribution) and Δ_1 (the shift term of the prior) – between the visual and auditory modalities. This suggests that the internal presentation of the temporal context is amodal, consistent with previous findings (Zhang & Zhou, 2017). Our Bayesian model provides a framework of where the modality difference may arise in each phase of the production-reproduction task. In the production phase, when the duration is transferred to short-term memory, the representation of the temporal context (prior knowledge) appears to be little influenced by the presentation modality. Accordingly, the prior knowledge representation appears largely modality-independent. At the same time, the reproduction carry-over bias Δ_2 did show a decreasing trend across the range of the “Blocked” conditions, reflecting block-wise bias.

4.4. Variation of the CV and sub-second timing

In a meta-analysis of animal studies Gibbon and colleagues (1997) discerned two “jumps” in the CVs at approximately 0.1 and 1.5 s, which they took to suggest that there might be different timing mechanisms for different time ranges. Moreover, the CV has been reported to be particularly large for durations below 100 ms (see a review, Wearden & Lejeune, 2008), for which duration judgments also exhibit a large overestimation (Chen et al., 2016). The large CV for ultrashort visual durations has been argued to be influenced by temporal summation (Gorea, 2015; Scharnowski et al., 2007), that is: below 100 ms, perceived duration depends heavily on light intensity. This would be associated with rather high uncertainty in the estimation of short visual time intervals. In our study, the overall reproduction error for

the 300-ms interval reached 102%. However, the “jumps” in our study were observed only in the “Blocked” (and not the “Mixed”) conditions, reflecting a carry-over effect induced by the block-wise variation of the duration ranges.

Of note, we replicated the previous report of the CV changing continuously along the timeline, rather than exhibiting abrupt jumps (Lewis & Miall, 2009). The continuous change of the CV is likely driven by a “hidden” (not directly observable) factor that had not been revealed in early studies. According to our model, which does capture the continuous decrease of the CV, this pattern arises because, while the sensory uncertainty scales with sample duration according to the scalar property, the decision uncertainty itself is independent of the length of the to-be-judged duration. As a result, the contribution of the decision uncertainty to the total uncertainty increases the CV more for short durations compared to longer durations.

It should be noted that the variation of CV does not solely depend on the temporal context and non-decision noise. It has been reported that extensive learning can enhance temporal discrimination, indicated by a decrease of the Weber fraction (i.e., CV) over the course of the training (Karmarkar & Buonomano, 2003; Matthews & Grondin, 2012). In fact, perceptual learning can boost perceptual discrimination, generally, and not just in the time domain (e.g., Schwartz et al., 2002; Shams & Seitz, 2008). There is evidence indicating that perceptual learning involves a re-tuning of decision templates over the course of training (Li et al., 2004). Interestingly, though, there have been several reports that the enhanced discrimination brought about by learning is rather target-specific. For instance, Karmarkar and Buonomano (2003) showed that training on a 100- or 200-msec interval did enhance the temporal discrimination of the respective (target) interval, without generalizing to untrained intervals. Similarly, in animal temporal-reproduction study, Bizo et al. (2006) found a U-shaped Weber fraction which they attributed to different rates of reinforcement training: a high rate of reinforcement for intermediate durations, relative to extremely short or long durations, led to enhance temporal discrimination (evidenced by a reduced Weber fraction) for the intermediate range of durations. From the perspective of the optimal model framework, perceptual or reinforcement learning of particular durations would reduce the variability of sensory measure σ_s for the learned durations, which may lead to unequal measurement uncertainty σ_s in the logarithmic timeline. This kind of ‘real’ violation in the scalar property can also be reflected in other models we compared (see *Appendix A: Model comparison*) without assuming the constancy of σ_s . However, in the present study, the best model doesn’t need this additional variation assumption.

In summary, the subjective time is susceptible to various contextual modulations, as is the coefficient of variation (CV). In light of the present results, the “golden” method to test for violation of the scalar property has to be reconsidered: we successfully modeled the observed variation of the CV without assuming any violations of the scalar property. Both the temporal context and decision uncertainty are contributing to the variation of the CV, with the latter (decision uncertainty) playing a critical role in accounting for the large CV for sub-second durations.

Acknowledgment

This work was supported by the German Research Foundation (DFG) grants SH166/3-2, awarded to ZS.

The study was conducted with the understanding and written, informed consent from each participant, with the approval of the Ethics Board of the LMU Faculty of Pedagogics and Psychology (23.04.2018).

Data accessibility

Data can be accessed on the Github platform: https://github.com/msenselab/variation_in_cv.

Authors' contributions

Y.R. and Z.S. conceived the experimental design. Y.R. collected the data. Y.R., F.A., and Z.S. carried out statistical analysis and Bayesian modeling. Y.R., F.A., Z.S., H.J.M. drafted the manuscript. All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

References

- Bausenhart, K. M., Dyjas, O., & Ulrich, R. (2014). Temporal reproductions are influenced by an internal reference: explaining the Vierordt effect. *Acta Psychologica*, 147, 60–67.
- Bizo, L. A., Chu, J. Y. M., Sanabria, F., & Killeen, P. R. (2006). The failure of Weber’s law in time perception and production. *Behavioural Processes*, 71(2-3), 201–210.
- Buhusi, C. V., & Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nature Reviews. Neuroscience*, 6(10), 755–765.
- Buhusi, C. V., & Oprisan, S. A. (2013). Time-scale invariance as an emergent property in a perceptron with realistic, noisy neurons. *Behavioural Processes*. <https://doi.org/10.1016/j.beproc.2013.02.015>
- Chen, L., Bao, Y., & Wittmann, M. (2016). *Sub-and Supra-Second Timing: Brain, Learning and Development*. Frontiers Media SA.
- Church, R. M., Meck, W. H., & Gibbon, J. (1994). Application of scalar timing theory to individual trials. *Journal of Experimental Psychology. Animal Behavior Processes*, 20(2), 135–155.
- Cicchini, G. M., Arrighi, R., Cecchetti, L., Giusti, M., & Burr, D. C. (2012). Optimal encoding of interval timing in expert percussionists. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 32(3), 1056–1060.

Cicchini, G. M., Mikellidou, K., & Burr, D. C. (2018). The functional role of serial dependence.

Proceedings. Biological Sciences / The Royal Society, 285(1890).

<https://doi.org/10.1098/rspb.2018.1722>

Ebbinghaus, H. (1885). *Über Das Gedächtnis: Untersuchungen Zur Experimentellen Psychologie*.

Creative Media Partners, LLC.

Ernst, M. O., & Di Luca, M. (2011). Multisensory Perception: From Integration to Remapping. In

Sensory Cue Integration (pp. 224–250).

<https://doi.org/10.1093/acprof:oso/9780195387247.003.0012>

Fechner, G. T. (1860). *Elemente der Psychophysik. Vol I and II*. Breitkopf und Härtel.

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience*, 17(5),

738–743.

Ganzenmüller, S., Shi, Z., & Müller, H. J. (2012). Duration reproduction with sensory feedback delay:

differential involvement of perception and action time. *Frontiers in Integrative Neuroscience*,

6(October), 1–11.

Getty, D. J. (1975). Discrimination of short temporal intervals: A comparison of two models. In

Perception & Psychophysics (Vol. 18, Issue 1, pp. 1–8). <https://doi.org/10.3758/bf03199358>

Gibbon, J. (1991). Origins of scalar timing. *Learning and Motivation*, 22(1-2), 3–38.

Gibbon, J., & Church, R. M. (1990). Representation of time. *Cognition*, 37(1-2), 23–54.

Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar Timing in Memory. *Annals of the New York*

Academy of Sciences, 423, 52–77.

Gibbon, J., Malapani, C., Dale, C. L., & Gallistel, C. (1997). Toward a neurobiology of temporal

cognition: advances and challenges. *Current Opinion in Neurobiology*, 7(2), 170–184.

Glasauer, S. (2019). Chapter 1 - Sequential Bayesian updating as a model for human perception. In S.

- Ramat & A. G. Shaikh (Eds.), *Progress in Brain Research* (Vol. 249, pp. 3–18). Elsevier.
- Glasauer, S., & Shi, Z. (2018). 150 years of research on Vierordt’s law-Fechner’s fault? *bioRxiv*.
<https://www.biorxiv.org/content/10.1101/450726v1.abstract>
- Gorea, A. (2015). A Refresher of the Original Bloch’s Law Paper (Bloch, July 1885). *I-Perception*, 6(4), 2041669515593043.
- Grondin, S., & Killeen, P. R. (2009). Tracking time with song and count: different Weber functions for musicians and nonmusicians. *Attention, Perception & Psychophysics*, 71(7), 1649–1654.
- Gu, B. M., Jurkowski, A. J., Shi, Z., & Meck, W. H. (2016). Bayesian optimization of interval timing and biases in temporal memory as a function of temporal context, feedback, and dopamine levels in young, aged and *Timing & Time Perception*. https://brill.com/view/journals/time/4/4/article-p315_1.xml
- Gu, B. M., & Meck, W. H. (2011). New Perspectives on Vierordt’s Law: Memory-Mixing in Ordinal Temporal Comparison Tasks. In *Multidisciplinary Aspects of Time and Time Perception* (pp. 67–78). https://doi.org/10.1007/978-3-642-21478-3_6
- Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nature Neuroscience*, 13(8), 1020–1026.
- Jost, A. (1897). *Die Assoziationsfestigkeit in ihrer Abhängigkeit von der Verteilung der Wiederholungen*. L. Voss.
- Karmarkar, U. R., & Buonomano, D. V. (2003). Temporal specificity of perceptual learning in an auditory discrimination task. *Learning & Memory*, 10(2), 141–147.
- Lejeune, H., & Wearden, J. H. (2006). Scalar properties in animal timing: conformity and violations. *Quarterly Journal of Experimental Psychology*, 59(11), 1875–1908.
- Lejeune, H., & Wearden, J. H. (2009). Vierordt’s The Experimental Study of the Time Sense (1868) and

its legacy. In *European Journal of Cognitive Psychology* (Vol. 21, Issue 6, pp. 941–960).

<https://doi.org/10.1080/09541440802453006>

Lewis, P. A., & Miall, R. C. (2009). The precision of temporal judgement: milliseconds, many minutes, and beyond. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 364(1525), 1897–1905.

Li, R. W., Levi, D. M., & Klein, S. A. (2004). Perceptual learning improves efficiency by re-tuning the decision template for position discrimination. *Nature Neuroscience*, 7(2), 178–183.

Matthews, W. J., & Grondin, S. (2012). On the replication of Kristofferson’s (1980) quantal timing for duration discrimination: some learning but no quanta and not much of a Weber constant. *Attention, Perception & Psychophysics*, 74(5), 1056–1072.

Matthews, W. J., & Meck, W. H. (2014). Time perception: the bad news and the good. *Wiley Interdisciplinary Reviews: ...*, 5(4), 429–446.

Meck, W. H. (1983). Selective adjustment of the speed of internal clock and memory processes. *Journal of Experimental Psychology. Animal Behavior Processes*, 9(2), 171–201.

Meck, W. H., Komeily-Zadeh, F. N., & Church, R. M. (1984). Two-step acquisition: modification of an internal clock’s criterion. *Journal of Experimental Psychology. Animal Behavior Processes*, 10(3), 297–306.

Petzschnner, F. H., Glasauer, S., & Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends in Cognitive Sciences*, 1–9.

Scharnowski, F., Hermens, F., & Herzog, M. H. (2007). Bloch’s law and the dynamics of feature fusion. *Vision Research*, 47(18), 2444–2452.

Schwartz, S., Maquet, P., & Frith, C. (2002). Neural correlates of perceptual learning: a functional MRI study of visual texture discrimination. *Proceedings of the National*.

<https://www.pnas.org/content/99/26/17137.short>

- Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences*, 12(11), 411–417.
- Shi, Z., & Burr, D. (2016). Predictive coding of multisensory timing. *Current Opinion in Behavioral Sciences*, 8, 200–206.
- Shi, Z., Church, R. M., & Meck, W. H. (2013). Bayesian optimization of time perception. *Trends in Cognitive Sciences*, 17(11), 556–564.
- Shi, Z., Ganzenmüller, S., & Müller, H. J. (2013). Reducing Bias in Auditory Duration Reproduction by Integrating the Reproduced Signal. *PloS One*, 8(4), e62065.
- Staddon, J. E., & Higa, J. J. (1999). Time and memory: towards a pacemaker-free theory of interval timing. *Journal of the Experimental Analysis of Behavior*, 71(2), 215–251.
- Taubert, J., Alais, D., & Burr, D. (2016). Different coding strategies for the perception of stable and changeable facial attributes. *Scientific Reports*, 6, 32239.
- Teghtsoonian, R., & Teghtsoonian, M. (1978). Range and regression effects in magnitude scaling. *Perception & Psychophysics*, 24(4), 305–314.
- Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the “internal clock.” *Psychological Monographs: General & Applied*, 77(13), 1–31.
- Wearden, J. H. (1992). Temporal generalization in humans. In *Journal of Experimental Psychology: Animal Behavior Processes* (Vol. 18, Issue 2, pp. 134–144). <https://doi.org/10.1037//0097-7403.18.2.134>
- Wearden, J. H. (2006). When do auditory/visual differences in duration judgments occur? *The Quarterly Journal of Experimental Psychology*, 59(10), 1709–1724.
- Wearden, J. H., Denovan, L., Fakhri, M., & Haworth, R. (1997). Scalar timing in temporal

generalization in humans with longer stimulus durations. *Journal of Experimental Psychology*.

Animal Behavior Processes, 23(4), 502–511.

Wearden, J. H., Edwards, H., Fakhri, M., & Percival, A. (1998). Why “Sounds Are Judged Longer Than Lights”: Application of a Model of the Internal Clock in Humans. *The Quarterly Journal of Experimental Psychology: Section B*, 51(2), 97–120.

Wearden, J. H., & Lejeune, H. (2008). Scalar properties in human timing: conformity and violations. *Quarterly Journal of Experimental Psychology*, 61(4), 569–587.

White, K. G. (2001). Forgetting functions. *Animal Learning & Behavior*, 29(3), 193–207.

Wiener, M., Thompson, J. C., & Coslett, H. B. (2014). Continuous carryover of temporal context dissociates response bias from perceptual influence for duration. *PloS One*, 9(6), e100803.

Zhang, H., & Zhou, X. (2017). Supramodal Representation of Temporal Priors Calibrates Interval Timing. *Journal of Neurophysiology*, jn.01061.2015.

Appendix A: Model comparison

In the blocked conditions, we expected that there might be some influence of the different temporal contexts in the “short”, “medium”, and “long” duration blocks. In order to investigate what form of influence temporal context would take, we performed a model comparison in which, for each of the five model parameters, we used either a single value of the parameter for all blocks or different values for the “short”, “medium”, and “long” blocks. The five parameters we compared in this way were: the standard deviation of the sensory measurement σ_s , the standard deviation of the prior σ_p , the two shift parameters Δ_1 and Δ_2 , as well as the standard deviation related to response uncertainty. Combinatorially, these five factors yielded in a total of $2^5 = 32$ different possible models. We fitted each of these models to the data of individual participants and evaluated them in terms of the average Bayesian Information Criterion (BIC) across participants. Table A1 lists the parameters that took on different values for the “short”, “medium”, and “long” duration blocks for the five models with the lowest BIC values (averaged across participants). The best model allowed the reproduction bias Δ_2 and the standard deviation of the prior σ_p , but not the other three parameters, to differ among blocks.

Table A1: The five models with the lowest BIC

Separate σ_s	Separate σ_p	Separate Δ_1	Separate σ_r	Separate Δ_2	BIC
No	Yes	No	No	Yes	490
Yes	No	No	No	Yes	492
No	Yes	No	No	No	494
Yes	Yes	No	No	No	495
Yes	No	Yes	No	No	497

In addition to finding the best model (with the lowest BIC), we further evaluated for each parameter whether the model performed better with or without the property of allowing that parameter to vary among blocks, by comparing the BIC of the best models with that property to the BIC of the best model

without it (ΔBIC_{\min} in Table A2) as well as the difference in average BIC across all models with the property and those without it ($\Delta\text{BIC}_{\text{mean}}$ in Table A2).

Table A2: For each factor, this table presents the difference in average BIC across all models with each property and those without it ($\Delta\text{BIC}_{\text{mean}}$), as well as the difference in BIC of the best model (with the lowest BIC) with a property and the best model without it (ΔBIC_{\min}). Negative values support models with the property over models without it.

	$\Delta\text{BIC}_{\text{mean}}$	ΔBIC_{\min}
Separate σ_s	-4.8	2.0
Separate σ_p	-6.2	-2.0
Separate Δ_I	-2.8	6.3
Separate σ_r	5.5	6.4
Separate Δ_2	-5.4	-3.7

This analysis supported models with a different reproduction bias parameter in different blocks and a separate σ_p parameter in each block. For the σ_r parameter, the analysis supported models with a single parameter across blocks. For σ_s and Δ_I , the results were ambiguous: models with a single σ_s parameter and a single Δ_I parameter were supported in terms of ΔBIC_{\min} (since these were properties of the best model); in terms of $\Delta\text{BIC}_{\text{mean}}$, by contrast, models in which these parameters could differ between blocks performed better.

Appendix B:

To identify how well the model predicts the variability of the duration reproduction at the individual-participant level, we plotted the predicted standard deviations (SDs) of the duration reproduction from the best model vs. the observed reproduction variability (SDs) in Figure S1, separately for the four experimental conditions. A correlation analysis revealed the predicted and observed SDs to be highly correlated ($r^2 = .924, .879, .956, .95$ for the “Vis/Mix”, “Aud/Mix”, “Vis/Block”, “Aud/Block” conditions respectively), indicating that the model prediction is in good agreement with the observed data.

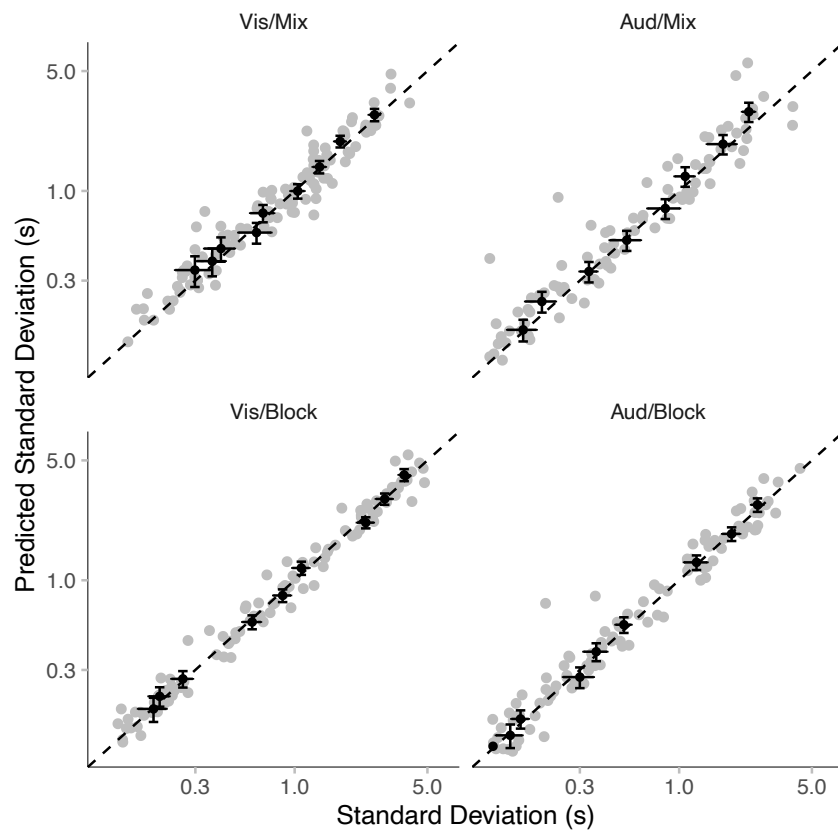


Figure S1. Predicted standard deviations (SDs) vs. observed SDs of the reproduction durations. Gray dots denote predicted vs. observed SDs for each participant, for each duration; the black dots and error bars denote the means and associated standard errors ($\pm SE$) of the predicted vs. observed SDs, for each duration. The diagonal line represents the ideal model for which the predicted SDs match the observed SDs.