Modeling how contextual and structural biases shape duration perception

Laetitia Grabot[1], Anne Giersch[2], Pascal Mamassian[1]

[1] Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, 75005 Paris, France

[2] Strasbourg Translational Neuroscience & Psychiatry, Strasbourg University, Strasbourg, France

Abstract

Human time perception is flexible and shaped by both structural constraints and contextual influences. Disentangling these sources of bias is essential for understanding the predictive mechanisms underlying temporal perception, yet no unified model currently integrates them. Here, we quantified precisely the structural and contextual biases in a duration discrimination task to constrain models of duration perception. Using a duration discrimination task, healthy participants judged which of two consecutive stimuli lasted longer. Stimuli were visual, auditory, or cross-modal. Consistent with previous findings, auditory stimuli were perceived as longer than visual stimuli of equal duration, reflecting intrinsic properties of audiovisual neural processing. Contextual biases were manipulated through different duration distributions known to influence temporal judgments. Our results show that duration discrimination relies on distinct representations of each stimulus distribution and is best explained by a combination of Bayesian inference and rescaling. While Bayesian inference accounts for contextual effects such as regression to the mean, rescaling mechanisms predict opposite effects by correcting perceived durations after repeated exposure. These findings challenge existing accounts of contextual effects in time perception and suggest that the brain normalizes temporal representations to adapt to environmental statistics.

Key words: Time Perception, Vision & Auditory Perception, Individual Differences, perceptual biases, Bayesian model, rescaling

Fundings: This work was supported by the FRC and the UNAFAM association, and scientific expertise is provided by the FRC's Scientific Advisory Board, and the Anneliese Maier Award to Pascal Mamassian from the Alexander von Humboldt Foundation.

Introduction

Time does not flow the same way for everyone because individuals weigh sensory modalities and temporal context differently. For example, when watching dubbed movies for an extended duration, some people are sensitive to problems of approximate synchronization between image and sound, while others do not notice anything (Ipser et al., 2017, 2018). These discrepancies in time perception may reflect adaptive mechanisms operating across multiple timescales. Individuals can temporarily adapt to and compensate for physical asynchronies, yet they also exhibit stable, idiosyncratic differences in brain structure, such as the connectivity between timing-related brain areas. Such differences may arise from longer-term adaptive processes. Both contextual information and structural brain constraints (or structural priors) shape time perception. Because these two types of priors differ in nature and likely rely on distinct mechanisms, clearly distinguishing between them will be essential for understanding how the brain constructs the experience of time. More generally, this idea has recently been discussed as a general approach (de Lange et al., 2018; Yon et al., 2019; Teufel & Fletcher, 2020) that could bring new insights in understanding the biological implementations of the predictive brain framework. However, while contextual priors have been extensively studied, structural priors remain comparatively underexplored (Seriès & Seitz, 2013; Teufel & Fletcher, 2020). As a first step to fill this gap, we aim here to develop a computational model to quantify the impact of both structural and contextual priors on time perception in humans.

Time perception is surprisingly highly variable and depends on the context and the task in hand, and this variability hinders making progress on major questions about time perception, such as how the brain computes time (Matthews & Meck, 2014). The extensive literature shows that all aspects of time, whether duration, order or synchrony, are influenced by context-based priors that reflect environmental constraints. For instance, perceived duration strongly depends on the temporal statistics of the environment (Jazayeri & Shadlen, 2010), and perceived simultaneity can expand following an exposure to persistent stimuli asynchrony (Fujisaki et al., 2004). This lability in time perception has therefore been nuanced (Sanford, 1888; Grabot & van Wassenhove, 2017). Indeed, even if inter-individual variability is large, the characteristics of a given participant remain stable when tested several times, such as for instance their bias to order auditory and visual stimuli (Grabot & van Wassenhove, 2017). This observation led to the idea that perceived simultaneity relies in part on idiosyncratic traits (Grabot & van Wassenhove, 2017; Sanford, 1888). Importantly, these biases are not overcome by attentional manipulation (Grabot & van Wassenhove, 2017) or contextual influences (Grabot & Kayser, 2020), suggesting that structural priors may play an important role in time perception at the individual level. However, time perception is usually studied on the basis of group average data, ignoring individual biases, and thus potentially leading to an overlook of crucial mechanisms.

The present study aims at jointly quantifying the impact of both contextual and structural priors on time perception, by combining psychophysics with individual-level modeling of behavioral data. Participants performed a duration discrimination task, comparing the duration of two stimuli, both visual, both auditory, or one of each (Figure 1). Sounds are usually perceived longer than visual events (Goldstone, 1968; Goldstone & Lhamon, 1972, 1974), which may be due to the intrinsic properties of sensory timing systems (Wearden et al., 1998). Inter-individual differences in this audiovisual bias will

be used as a proxy to measure structural priors. The contextual prior was manipulated by changing the distribution of presented stimuli. Temporal estimates are known to depend on the distribution from which they are drawn: the same duration is underestimated if drawn from a pool of short durations, but overestimated if drawn from a pool of long durations (Jazayeri & Shadlen, 2010; Acerbi et al., 2012; Murai & Yotsumoto, 2016; Roach et al., 2017; Maaß et al., 2019; Damsma, Schlichting, & van Rijn, 2021; Damsma, Schlichting, van Rijn, et al., 2021). This reflects a classical central tendency effect, in which perceived durations are biased toward the mean of the overall distribution.

The durations perceived by each participant, as well as the inter-individual differences in performance will be modeled using tools of Bayesian theory. According to this probabilistic framework, perception arises from a combination of noisy sensory information and prior knowledge. The main components of Bayesian models are the likelihood function (the probability distribution of measured durations given the physical duration), the prior probability distribution (knowledge of the temporal context), and the posterior distribution obtained by combining the first two (estimation of the physical duration given the measurement). Bayesian models have successfully explained the change of performance in a temporal task due to contextual priors (Jazayeri & Shadlen, 2010; Mamassian & Landy, 2010). The main novelty of our approach is to consider two sources of priors in the model (structural and contextual), and test how they are combined, by assessing their respective weight in the final duration estimate.

We considered several implementations of the model to test specific hypothesis on how durations are represented and processed during a discrimination task. Because the two stimuli to be compared were drawn from distinct distributions, our first question was whether the perceptual system maintains separate representations for each distribution or instead encodes a single, broader distribution encompassing all possible durations. The results supported the former: each stimulus distribution appeared to be distinctly represented and this impacted behavior. Next, we tested whether the behavioral data were purely driven by Bayesian processes predicting central tendency effect. Alternatively, we considered whether an additional recalibration mechanism might better account for the data (Fujisaki et al., 2004). After repeated exposure to a constant lag, the perceptual system might compensate for this lag, effectively rescaling the temporal representation and shifting the perception of subsequent durations (Miyazaki et al., 2006; Murai & Yotsumoto, 2016). Our results indicated that the best-fitting model required both Bayesian and rescaling processes. Finally, we included in the models a parameter to capture an ordering effect, based on the idea that the first stimulus may be represented more noisily than the second because it must be retained in memory for comparison during the discrimination judgment (Ellinghaus et al., 2024a; Raviv et al., 2012).

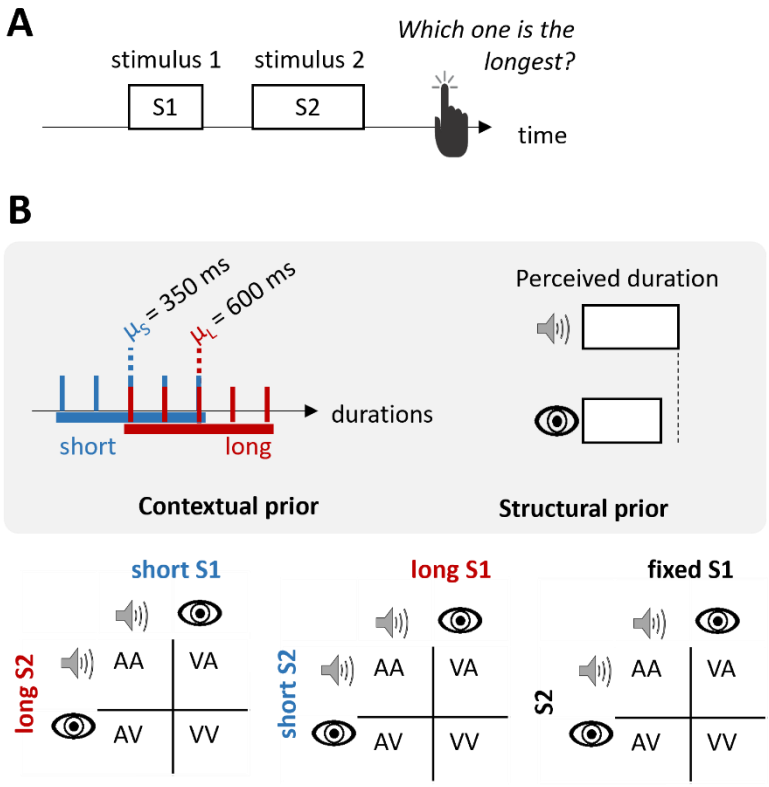

Figure 1. Duration discrimination task involving structural and contextual priors. A*. Participants judged which of the two presented durations was perceived longer.* ***B****. The contextual prior was manipulated by altering the distributions from which the first and second stimuli were drawn (short or long distributions, in blue and red respectively, in log space). The structural prior was assessed by comparing participants' perceptions of the duration of visual and auditory stimuli with identical physical durations. All combinations of conditions were tested, as well as 4 supplementary conditions where the duration of the first stimulus was fixed.*

Materials & Methods

Participants

31 participants (19 female, age M = 27 years, SD = 5) took part in the study. All had normal or corrected-to-normal vision, and no known neurological or psychiatric disorder history. All participants were naive as to the purpose of the study, provided written informed consent and were compensated for their participation. The study was performed at the Ecole Normale Supérieure (ENS, Paris, France), in accordance with the Declaration of Helsinki (2013) and approved by the French Ethics Committee on Human Research (N° 2022-A00963-40). One participant only performed one experimental session and was therefore excluded, leaving a total of 30 participants considered for the analyses.

Stimuli

The experiment was written in MATLAB (Version R2023a; The MathWorks, Inc., Natick, MA) using Psychtoolbox toolbox Version 3.0.19 (Brainard, 1997; Kleiner et al., 2007). The visual stimuli were pink noise annuli (inner diameter = 1°VA, outer diameter = 4°VA, contrast = 0.5) centered on the middle of

the screen, presented on a gray background (refresh rate = 120 Hz, resolution = 1920*1080). The auditory stimuli were pink noise (60 dB, 48 kHz, with 5 ms fade-in and 5 ms fade-out), presented with a sound bar below the screen. They were created with the Binaural toolbox (Akeroyd, 2017). The stability of stimulus timing was verified with an oscilloscope and a photodiode. Auditory and visual 200ms-duration has a standard deviation of 0.7 and 0.3 ms, respectively (across 10 repetitions), ensuring stable and reliable presented durations.

Experimental procedure

Participants performed a duration discrimination task in a darkened room, with their head on a chin rest located 85 cm away from the computer screen. During a trial, two stimuli were presented sequentially and participants were asked to judge which one lasted longer (Figure 1A). They replied with a button press; the button mapping was pseudorandomized between participants. The delay between the first stimulus S1 and the second S2 was 0.2 s, and the inter-trial interval lasted between 0.9 and 1.1s (randomly distributed). Participant were asked to fixate a central fixation cross during the whole experiment.

There were 8 main different conditions varying in both the modality and the distribution from which each stimulus was sampled (Figure 1B). The pair of stimuli were either both auditory, both visual, or one auditory and the other visual (denoted AA, VV, AV and VA). Depending on the condition, the duration of a stimulus in one interval could be drawn from either a short (0.2, 0.267, 0.35, 0.45, 0.6 s) or a long distribution (0.35, 0.45, 0.6, 0.783, 1.033 s). The five items of these distributions were placed on a logarithmic scale. Their median values were 0.35 s and 0.6 s, which corresponds to a ratio of 1.7, shown to be large enough to create two distinct prior representations (Maaß et al., 2019). We tested all possible combinations for S1 coming from one distribution, and S2 coming from the other distribution (denoted SL-AA, SL-VV, SL-AV, SL-VA, LS-AA, LS-VV, SL-AV and SL-VA conditions), and each S1-S2 pair was repeated 10 times.

To assess the structural prior independently of the contextual prior, we added 4 baseline conditions designed so that participants had to register only one distribution. S1 was fixed to 0.45 s and S2 was sampled from a distribution that was larger than those used in the 8 main conditions, covering both the short and long distributions (7 values ranging from 0.2 to 1.033 s). The four combinations of stimuli and modality were tested (AA, VV, AV, VA), with 10 repetitions per duration.

The experiment was split into two identical sessions to ensure it remained manageable in duration (~1h20 per session, breaks included). At the beginning of the session, the participants were trained to the task during ~5 min with the four baseline conditions (S1-fixed) using only the most extreme durations (0.2 and 1.03, 3 repetitions per duration). They received feedback after each trial (color of the fixation cross changing to green if correct, red if incorrect). Feedback was then removed for the remaining of the experiment. Participants then performed the 4 blocks of baseline task, presented in pseudorandom order. After a break, they performed 8 blocks of the main task (one condition per block), presented in pseudorandom order.

Behavioral models

To quantify both contextual and structural priors, we used Bayesian models to estimate the behavioral response of participants to the duration discrimination task. Five models were built, each testing

different hypotheses underlying the implementation of contextual priors (Figure 2A). These five models are derived from combining three main models described in this section: “2distrib”, “1distrib” and “rescaling”. The “2distrib” model posits that the perceptual system maintains separate representations of the distributions of the first and second intervals, from which the first and second durations are drawn. In contrast, the “1distrib” model assumes that the system relies on a single unified distribution encompassing all possible durations experienced in this block of trials. Lastly, the “rescaling” model adds an additional step where both distributions are first rescaled into a common internal axis before being used in the Bayesian estimation.

In the duration discrimination task, participants are presented with two sequential stimuli and have to judge which duration is the longest. The “2distrib” model estimates durations as follows. A duration *ts* is represented in log space by a Gaussian distribution centered on the true physical duration *t* shifted by δ, a free parameter capturing the structural prior, i.e. the extent to which a participant overestimates a sound compared to a visual stimulus of the same duration. The standard deviation of the Gaussian is a free parameter, σ_A or σ_V depending on the modality of the stimulus. Therefore, the distribution for an auditory stimulus is:

$$ts_A = \mathcal{N}\big(\log\big(t\ (1+\delta/2)\big),\ \sigma_A\big) \qquad \text{(Eq. 1)}$$

and for a visual stimulus, it is:

$$ts_V = \mathcal{N}\big(\log\big(t\ (1-\delta/2)\big),\ \sigma_V\big) \qquad \text{(Eq. 2)}$$

By the time of the perceptual decision, the representation of the stimulus that was presented in the first interval might be noisier than the one in the second interval due to memory maintenance (Raviv et al., 2012; Ellinghaus et al., 2024b). To account for this effect, we introduce an additional parameter α that scales the standard deviation of the first stimulus. Therefore, the equations above correspond to the stimuli presented in the second interval, and to obtain the distributions for stimuli presented in the first interval, the standard deviations are replaced by $\alpha\ .\ \sigma_{AV}$ or $\alpha\ .\ \sigma_V$.

Under the assumption that the perceptual system has a noisy representation of a given physical duration, a duration *tm* is sampled from the Gaussian distribution defined above to generate the likelihood probability of *tm* knowing *ts*, as a Gaussian centered on *tm*, with a standard deviation of σ_A or σ_V ($\alpha\ .\ \sigma_{AV}$ or $\alpha\ .\ \sigma_V$ for a stimulus in the first interval). The prior probability p is the knowledge that the perceptual system has about the durations distribution. It is represented in log space by a Gaussian centered on the log mean of the N durations d_n of the distribution, denoted tp (N = 5 for the main conditions, 7 for the baseline conditions). The Gaussian has a standard deviation σp equivalent to the standard deviation of N Gaussians of standard deviation σc, following the law of total variance.

$$\mathrm{p} = \mathcal{N}(tp, \sigma\mathrm{p})) \qquad \text{(Eq. 3)}$$

$$\text{with } \sigma\mathrm{p} = \sqrt{\sigma\mathrm{c}^2 + \frac{1}{N}\sum_{n=1}^{N} log(d_n)^2 - (\frac{1}{N}\sum_{n=1}^{N} log(d_n))^2} \qquad \text{(Eq. 4)}$$

$$\text{And } tp = \frac{1}{N}\sum_{n=1}^{N}\log(d_n * (1+\ \delta/2)) \qquad \text{(Eq. 5)}$$

The structural prior also influences the prior probability, which accounts for the presence of the δ term in Equation 5. Note that this equation is formulated for a distribution of auditory stimuli; for a distribution of visual stimuli, the sign of the δ term is reversed.

The posterior probability is calculated as the product of the likelihood probability with the prior probability, according to Bayes' rule. This gives a Gaussian with a mean *te* and a standard deviation σe defined as follows:

$$te = (tm * \sigma p^2 + tp * \sigma_A^2) / (\sigma_A^2 + \sigma p^2) \quad \text{(Eq. 6)}$$

$$\text{and } \sigma e = \sqrt{\sigma_A^2 \, \sigma p^2 / (\sigma_A^2 + \sigma p^2)} \quad \text{(Eq. 7)}$$

The estimated duration is finally calculated as the duration minimizing the posterior expected loss, using Bayesian least square estimation (Jazayeri & Shadlen, 2010). The output of a given trial is given by estimating the durations of the first and second stimulus and by comparing the two estimates.

The three main models follow a similar sequence of steps, outlined below, with each model incorporating specific variations. The "2distrib" model considered as prior only the distribution from with the duration to estimate was drawn (i.e. the short or the long distribution). In contrast, the "1distrib" model consider all possible durations (coming from both first and second intervals distributions) in the calculation of the prior. The "rescaling" model has a supplementary first step where the Gaussians representing the prior and the stimulus are shifted toward the barycenter of the distributions of the first and second stimuli (composed of N durations, N = 7). The free parameter *c* indicates the extent of rescaling (0: none, 1: full realignment of the two distributions). Equation 1 is therefore rewritten as:

$$ts = \mathcal{N}\left(\log(t * (1 + \delta)) + c * (\tfrac{1}{N}\textstyle\sum_{n=1}^{N} \log(d_n) - \tfrac{1}{N}\textstyle\sum_{n=1}^{N} \log(d_n)), \sigma A\right) \quad \text{(Eq. 8)}$$

and Equation 4 becomes:

$$tp = \tfrac{1}{N}\textstyle\sum_{n=1}^{N} \log(d_n * (1 + \delta)) + c * (\tfrac{1}{N}\textstyle\sum_{n=1}^{N} \log(d_n) - \tfrac{1}{N}\textstyle\sum_{n=1}^{N} \log(d_n)) \quad \text{(Eq. 9)}$$

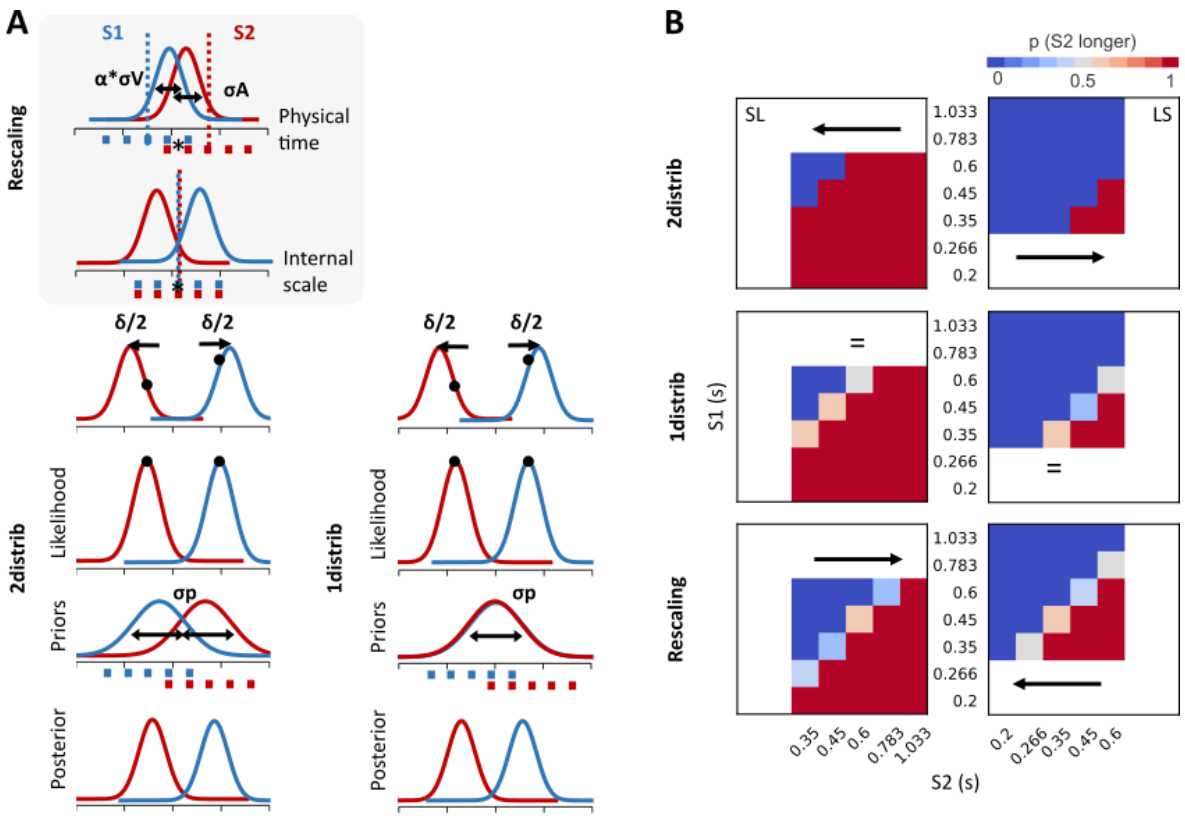

Figure 2. Behavioral models and predictions. A*. Three distinct implementations of contextual biases were developed as separate models. Within a trial, participants have to compare two durations (S1 and S2) presented in two consecutive intervals. They generate some internal representations whose uncertainty is modelled as Gaussians centered on the physical duration. If a visual stimulus is presented in the first interval and an auditory stimulus in the second, the standard deviations of these distributions are* $\alpha \cdot \sigma_V$ *and* σ_A*, respectively (the parameter* α *characterizes the fact that the first interval might be more uncertain by the time the perceptual decision is taken). The "rescaling" model realigns these Gaussians based on the barycenter of S1 and S2 distributions. Then, they are shifted by δ to account for the structural prior (difference of perceived durations between modalities). The posterior probability is calculated using the likelihood and prior probability. In the "2distrib" model, the distributions from which S1 and S2 are drawn are represented separately, while a wider single distribution is used in the "1distrib" model.* ***B****. The three models predict distinct patterns of expected duration discriminations, especially when one compares the blocks short-long (SL, left column) and long-short (LS, right column).*

Models fitting

There were six free parameters: the structural prior (δ), the standard deviation associated to each modality (σA or σV), the standard deviation used to calculate the prior (σc), the order effect parameter (α) and the rescaling parameter (c). We use the negative log likelihood (*NLL*) as cost function to fit these parameters to the data. The optimization of the models was done using Bayesian adaptive direct search (Acerbi & Ma, 2017) implemented in the PyBADS toolbox (Singh & Acerbi, 2024). This approach alternates between fast and local Bayesian optimization steps and slower exploration of the parameters space, and is particularly adapted when the objective function is stochastic. The maximal number of

function evaluations was set to 350, with the final 30 evaluations serving as repetitions of the final solution, used to compute the average parameter values. For each parameter, we specified a maximal and minimal bounds (δ = [-0.9, 0.9], σA = [0.06, 1], σV = [0.06, 1], σc = [0.06, 1], c = [0, 1], α = [0.5, 3]) as well as plausible lower and upper bounds (δ = [0, 0.4], σA = [0.07, 0.2], σV = [0.07, 0.2], σc = [0.07, 0.2], c = [0.05, 0.6] α = [1, 1.5]).

The initial parameters of each model were set after fitting data averaged across all participants, using initial parameters estimated from a grid search. The grid search explored the parameter space by evaluating all possible combinations of four values for each parameter (δ = [0.05, 0.1, 0.2, 0.3], σA = [0.09, 0.15, 0.2, 0.25], σV = [0.09, 0.15, 0.2, 0.25], σc = [0.09, 0.15, 0.2, 0.25], c = [0.1, 0.25, 0.40, 0.55], α = [1, 1.1, 1.2, 1.3]). The combination minimizing the negative log likelihood between model and data was chosen as initial parameters in the fit of average data. The resulting estimated parameters were used as initial parameters in the individual fits ("2distrib": [δ = 0.21, σA = 0.35, σV = 0.52, σc = 0.97, α = 1.17], "1distrib": [δ = 0.07, σA = 0.38, σV = 0.49, σc = 0.37, α = 1.52], "2distrib+rescaling": [δ = 0.11, σA = 0.34, σV = 0.44, σc = 0.54, c = 0.52, α = 1.32], "rescaling": [δ = 0.18, σA = 0.33, σV = 0.42, c = 0.31, α = 0.89], "2distrib + rescalingAV": [δ = 0.19, σA = 0.33, σV = 0.48, σc = 0.88, c = 0.55, α = 1.02]).

Models' comparison

In total, we tested 5 models that are combinations of the three main model presented above (2distrib, 1distrib, 2distrib+rescaling, rescaling, 2distrib+rescalingAV). The "2distrib+rescaling" is a model combining the representation of two distinct distributions for each stimulus to compare, with a first rescaling of these distributions. The "rescaling" model do not use the prior probability to compute the posterior, which is therefore equal to the likelihood. In order to test whether rescaling would be an adaptive mechanism to adjust specifically for cross-modal discrepancies (Fujisaki et al., 2004), the "2distrib+rescalingAV" model only applies the rescaling step to bimodal pairs of stimuli (auditory-visual or visual-auditory), but not to unimodal pairs. These models were independently fitted to the data and compared using Akaike Information Criterion (AIC), calculated as

$$AIC = 2*k+2*NLL, \qquad \text{(Eq. 10)}$$

with k the number of free parameters.

Parameter recovery

To assess the robustness of the model, we performed a parameter recovery analysis that consisted in simulating 1023 datasets from known parameters. The parameters were randomly picked within a 40% range of the final group-average parameters found for the best fitting model (2distrib+rescaling: δ = 0.07, σA = 0.29, σV = 0.37, σc = 0.37, c = 0.65, α = 1.47). Each simulated dataset was then fitted to the "2distrib+rescaling" model (best model found in the model's comparison analysis)and the recovered parameters were correlated to the known parameters. To investigate potential covariations between parameters, we also correlated between-parameter the deviation of recovered parameter from the known value.

Results

Contextual prior: two distributions are represented and rescaled onto an internal axis

We manipulated both contextual and structural priors in a duration discrimination task (Figure 1 and see Figure 3A and 3B for a visualization of the average data). We compared five models based on Bayesian integration, but specifying different implementations of the contextual prior. The "2distrib" model posits that the perceptual system represents the two durations distributions from which each stimulus is drawn. In contrast, the "1distrib" model proposes that only one single distribution is represented that encompasses all possible durations for S1 and S2. The "rescaling" model provides an additional step in which both distributions are first mapped onto an internal axis. Figure 2B shows how these different models lead to different predicted behavioral data. Using AIC, we compared five combinations of the three main models (Figure 3C). One combination specifically tested the hypothesis that the perceptual system performs rescaling only when the distributions are cross-modal ("2distrib+rescalingAV"). The best model across participants was the model combining the representation of two distinct distributions with a rescaling component for all stimuli pairs (ΔAIC summed across participants, relative to "2distrib+rescaling" for "2distrib": 6750, "1distrib": 1547, "rescaling": 468, "2distrib+rescalingAV": 1487). On an individual basis, this model minimized the AIC for 14 participants, the "rescaling" for 9 participants, the "2distrib+rescalingAV" for 4 participants, the "1distrib" for 2 participants and the "2distrib" for 1 participant.

The parameters found after fitting the best model to the data are shown in Figure 3D. The uncertainty on the auditory stimuli is significantly smaller than the one on to the visual stimuli (mean ± standard deviation: $\sigma A = 0.29 \pm 0.12$, $\sigma V = 0.37 \pm 0.15$, paired t-test: $t = -6.06$, $p = 1.4e-6$), which is consistent with previous findings (e.g. Burr et al., 2009; Wearden & Jones, 2021). The parameter measuring the contextual prior, σc, or the standard deviation of the prior probability is equal to 0.37 ± 0.08. The rescaling parameter c is equal to 0.65 ± 0.08 indicating a partial rescaling (a full rescaling corresponds to $c = 1$, with no rescaling when $c = 0$). Lastly, there is a stimulus order effect, since the parameter α is significantly above 1 ($\alpha = 1.47 \pm 0.43$, $t = 5.92$, $p = 2.0\ e-6$), indicating that the uncertainty of the first stimulus is ~47% larger than that of the second stimulus.

Structural prior: sounds duration is overestimated compared to visual durations

The difference of perceived durations between modalities was taken into account in behavioral models as a structural prior. The structural prior, or δ parameter is on average equal to +0.07 ± 0.04, which corresponds to an average overestimation of the sound of 7% or 31 ± 18 ms for an average duration of 450 ms. The parametric structural prior is positive for all participants, showing that they all overestimated the duration of sounds compare to visual stimuli. We correlated this parameter with a non-parametric estimation of the structural prior based on the difference of PSE between the audio-visual and visuo-auditory conditions in the baseline blocks (fixed-S1; Figure 3E, mean ± SD = 152 ± 92 ms, equivalent to ~34%). As expected, we found a significant correlation ($R = 0.41$, $p = 0.023$). Note, however, that the correlation did not align with the identity line, with a parametric structural prior lower than the non-parametric one, suggesting that part of the difference in the PSE may have been accounted for by other parameters.

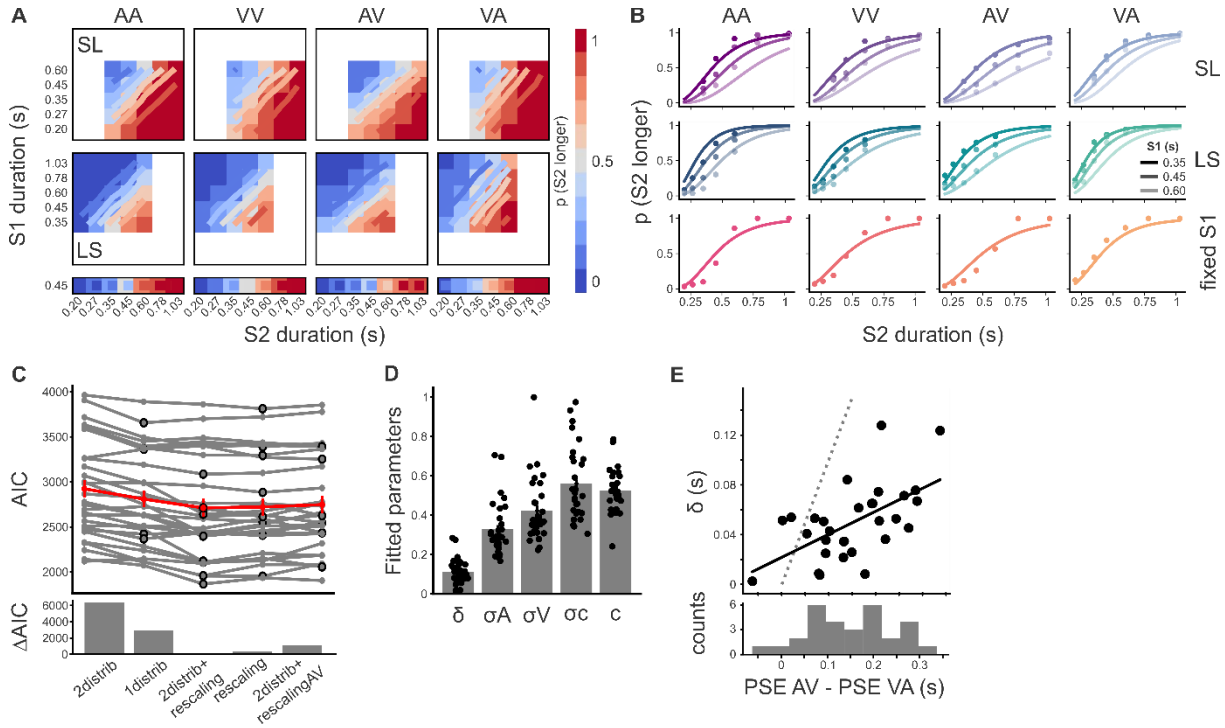

Figure 3. Best model and parameter estimates. *The behavioral data are best explained by supposing that the perceptual system represents two distinct durations distributions and rescale them onto a common internal axis.* ***A****. The behavioral data averaged across participants are plotted in matrices using the color scale that represent the probability of perceiving the second stimulus S2 as longer. Predictions from the best model (2distrib+rescaling) are plotted with contour lines superimposed on the matrices. The panels show all pairs of stimuli for each modality comparison condition (columns) and each duration distribution across the two intervals (rows, SL= short-long, LS = long-short).* ***B.*** *The same data are now plotted using psychometric curves. For the SL and LS conditions, only three curves are shown, corresponding to the three S1 durations that were common to the two intervals (0.6, 0.45 and 0.35 s). The dots correspond to the data averaged across participants, while the curves correspond to the best model's predictions.* ***C.*** *Five behavioral models were compared using AIC (Akaike Information Criterion). The best model minimizing AIC is "2distrib+rescaling". Each line corresponds to one participant, with the minimal AIC highlighted by a thick dot. The average is shown in red. In the bottom panel, AICs were first summed across participants and then displayed relative to the best model.* ***D.*** *The fitted parameters of the "2distrib+rescaling" model are shown, each dot representing one participant, and the bar shows the average.* ***E.*** *The parameter for the structural prior (δ) correlates significantly with the non-parametric structural prior, defined as the difference between the point of subjective equality (PSE) in the baseline audio-visual block and the PSE in the baseline visuo-audio block (bottom histogram).*

Parameter recovery

To test the robustness of the model, we run a parameter recovery analysis by simulating datasets (N = 1023) with parameters sampled from defined ranges. We fitted the best model ("2distrib+rescaling") to the simulated datasets and correlated fitted parameters to true known parameters. Each correlation was significant, with a slope close to 1, indicating a good recovery of known parameters (Figure 4A, Table 1). Note that the least reliable parameter is σc, which is overestimated for small true values of σc. The structural prior (δ) is slightly overestimated while the rescaling parameter c is slightly

underestimated. To investigate whether these errors might be related, we correlated the deviation from true parameters in between-parameters comparisons (Figure 4B). There is indeed a high negative correlation between δ and c (R = -0.66), which might explain the slight inaccuracies in estimating those parameters. There are also high correlations between the deviations of δ and σc (R = 0.70), σc and c (R = -0.82) and σA and α (R = -0.51). All the other correlations were also significant but with a smaller effect size (R<0.4).

	δ	**σA**	**σV**	**σc**	**c**	**α**
R	0.41	0.87	0.84	0.43	0.86	0.88
p	5.6e-42	7.9e-312	4.6e-277	1.2e-46	4.42e-298	< 10e-100
slope	0.97	0.91	0.87	0.45	0.96	0.91
intercept	0.02	0.03	0.05	0.25	-0.03	0.14

Table 1: *Correlation coefficient, p-value and regression line (slope and intercept) between fitted and true parameters from simulated datasets, for each parameter.*

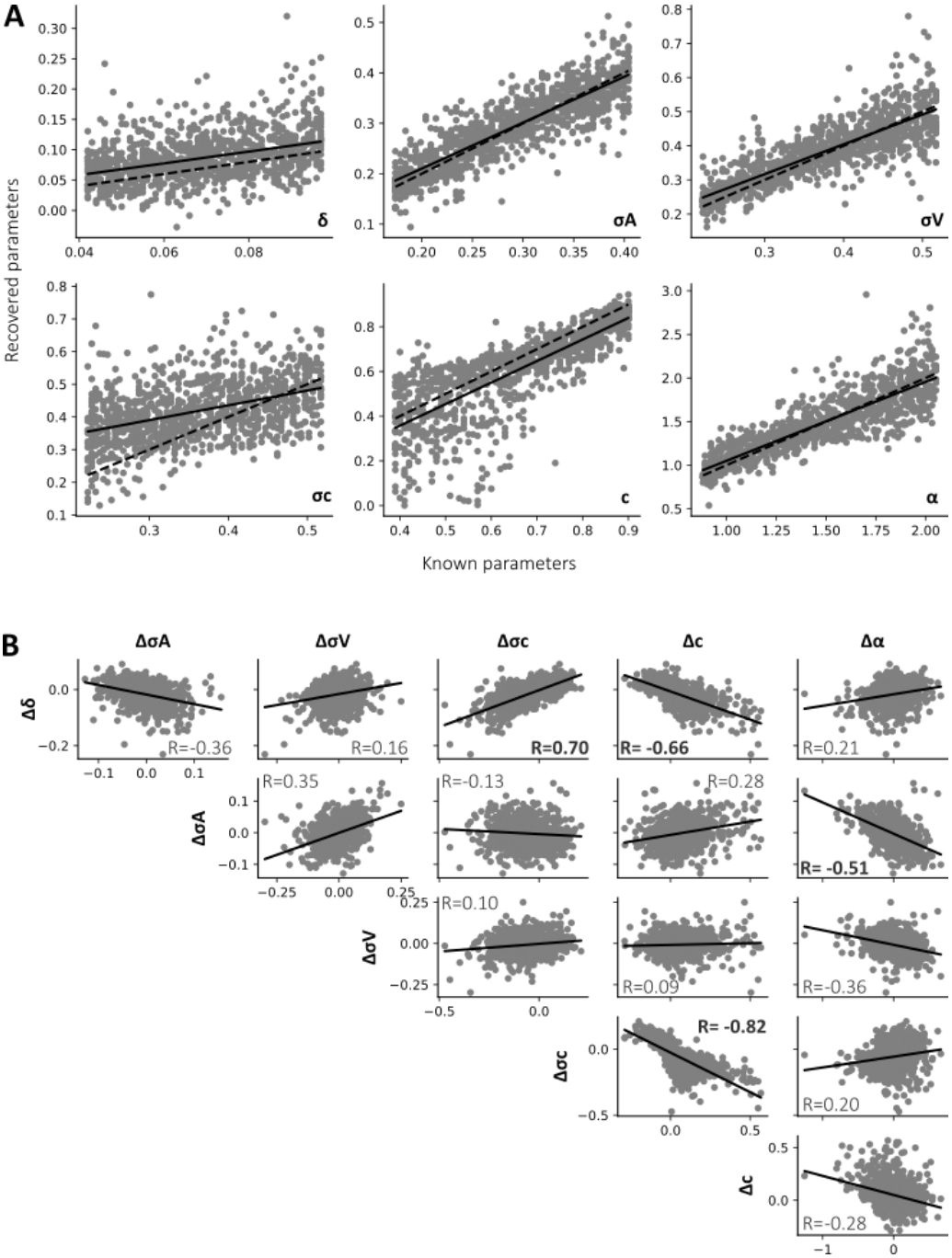

Figure 4. The parameter recovery analysis shows good robustness of the model. A*. Datasets were simulated by randomly selecting parameters, and fitted to the "2distrib+rescaling" model to recover the parameters. The dashed line is the identity line, while the bold line is the correlation between known and recovered parameter.* ***B.*** *To investigate the potential correlation between parameters, we took the differences between simulated and recovered parameters, and correlated these differences between parameters. All correlations (solid line) were significant but the size effects differ: the most significant correlations (R>0.4) are written in bold.*

Discussion

We proposed a unifying model taking into account how both contextual information and structural constraints shape duration perception. When estimating the duration of a stimulus, observers tend to perceive it as closer to the mean duration of the stimulus distribution than its actual duration. This regression-to-the-mean effect is classically well-described by Bayesian models (Jazayeri & Shadlen, 2010; Acerbi et al., 2012; Murai & Yotsumoto, 2016; Roach et al., 2017; Maaß et al., 2019). We adapted these models to a discrimination task, where participants compare two auditory or visual durations presented in succession. By comparing different models, we found that the two distributions from which each presented duration comes from are distinctly represented in the brain. Interestingly, the behavioral data were better explained by a Bayesian model that included a rescaling component. Rescaling means that before comparing the two durations, the perceptual systems realigned the two representational spaces of each stimulus within an internal scale, and this scale is used to perform the comparison. This rescaling effect biases the perceived durations in a direction opposite to that of the central tendency phenomenon accounted for by Bayesian theory. Two opposing mechanisms have also been identified in audiovisual temporal order judgments: lag adaptation, which compensates for a consistent delay between visual and auditory stimuli, and Bayesian integration, which predicts the opposite effect (Miyazaki et al., 2006; Yamamoto et al., 2012). For both lag adaptation in temporal order judgment and the rescaling process described here in duration discrimination, the perceptual system strives to establish a common perceptual space for integrating diverse sensory inputs.

In our experimental data, the rescaling of both stimulus representational spaces was substantial–around 65%–but not complete. This mechanism accounted for much of the behavioral data, as the Bayesian model alone (the “2distrib” model) provided the poorest fit among those tested. This finding may seem surprising given previous reports of successful Bayesian accounts of duration estimation (Jazayeri & Shadlen, 2010; Acerbi et al., 2012; Murai & Yotsumoto, 2016; Roach et al., 2017; Maaß et al., 2019). We propose that this discrepancy arises from task differences: our study involved a duration comparison task, whereas those previous studies primarily employed temporal production tasks, in which participants reproduced a given interval.

Rescaling can also be understood as comparing each stimulus duration to previously experienced durations within the same temporal interval: the first stimulus is evaluated relative to the prior for first stimuli, and the second relative to the prior for second stimuli. The final decision is then based on a comparison of these relative durations. We initially hypothesized that this strategy would occur primarily when comparing stimuli across modalities; however, this was not supported by our data, as the model applying rescaling only to audiovisual stimulus pairs (“2distrib+rescalingAV”) was not the best-fitting model. These results suggest that rescaling may represent a more general strategy employed in discrimination tasks, and further research is needed to test the generality of this mechanism.

The novelty of our approach lies in examining the impact of structural priors on time perception, in addition to contextual priors. We used the well-established overestimation of auditory relative to visual durations as a good example of structural prior. All participants overestimated auditory compared to

visual durations by an average of 7% (approximately 31 ms for a mean duration of 450 ms). This effect aligns with previous reports of auditory overestimation ranging from 6% to 28% (Goldstone, 1968; Goldstone & Lhamon, 1974; Wearden et al., 1998; Penney et al., 2000; Droit-Volet et al., 2004; Ulrich et al., 2006; Droit-Volet et al., 2007; Lustig & Meck, 2011), although it falls toward the lower end of that range. Notably, lower estimates in the literature often come from studies using modeling across different conditions (Penney et al., 2000; Droit-Volet et al., 2004; Droit-Volet et al., 2007; Lustig & Meck, 2011), whereas larger effects are typically reported in studies directly comparing behavioral responses (Goldstone, 1968; Goldstone & Lhamon, 1974; Wearden et al., 1998; Ulrich et al., 2006). We further confirmed this effect by correlating parametric and non-parametric estimates of the structural prior, with the latter yielding a higher overestimation (~34%).

We present a model of temporal discrimination that quantifies both structural and contextual priors, providing a framework for dissecting the underlying neural mechanisms of time perception. Because these biases arise from distinct sources—structural constraints versus contextual adaptation—it is likely that different brain regions and neural oscillatory dynamics contribute to each process (Seriès & Seitz, 2013; Teufel & Fletcher, 2020). This behavioral modeling work represents an essential first step toward uncovering the yet unknown mechanisms by which priors shape our perception.

Reference

Acerbi, L., & Ma, W. J. (2017). Practical Bayesian Optimization for Model Fitting with Bayesian Adaptive Direct Search. *Advances in Neural Information Processing Systems*, *30*. https://papers.nips.cc/paper_files/paper/2017/hash/df0aab058ce179e4f7ab135ed4e641a9-Abstract.html

Acerbi, L., Wolpert, D. M., & Vijayakumar, S. (2012). Internal Representations of Temporal Statistics and Feedback Calibrate Motor-Sensory Interval Timing. *PLOS Computational Biology*, *8*(11), e1002771. https://doi.org/10.1371/journal.pcbi.1002771

Akeroyd, M. (2017). *A binaural cross-correlogram toolbox for MATLAB* [Jeu de données]. The University of Nottingham. https://doi.org/10.17639/nott.320

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433-436.

Burr, D., Banks, M. S., & Morrone, M. C. (2009). Auditory dominance over vision in the perception of interval duration. *Experimental Brain Research*, *198*(1), 49-57. https://doi.org/10.1007/s00221-009-1933-z

Damsma, A., Schlichting, N., & van Rijn, H. (2021). Temporal Context Actively Shapes EEG Signatures of Time Perception. *J. Neurosci.*, *41*(20), 4514. https://doi.org/10.1523/JNEUROSCI.0628-20.2021

Damsma, A., Schlichting, N., van Rijn, H., & Roseboom, W. (2021). Estimating Time : Comparing the Accuracy of Estimation Methods for Interval Timing. *Collabra: Psychology*, *7*(1), 21422. https://doi.org/10.1525/collabra.21422

de Lange, F., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends Cogn Sci.*

Droit-Volet, S., Meck, W. H., & Penney, T. B. (2007). Sensory modality and time perception in children and adults. *Behavioural Processes*, *74*(2), 244-250. https://doi.org/10.1016/j.beproc.2006.09.012

Droit-Volet, S., Tourret ,Stéphanie, & and Wearden, J. (2004). Perception of the duration of auditory and visual stimuli in children and adults. *The Quarterly Journal of Experimental Psychology Section A*, *57*(5), 797-818. https://doi.org/10.1080/02724980343000495

Ellinghaus, R., Bausenhart, K. M., Koc, D., Ulrich, R., & Liepelt, R. (2024a). Order effects in stimulus discrimination challenge established models of comparative judgement : A meta-analytic review of the Type B effect. *Psychonomic Bulletin & Review*, *31*(5), 2275-2284. https://doi.org/10.3758/s13423-024-02479-3

Ellinghaus, R., Bausenhart, K. M., Koc, D., Ulrich, R., & Liepelt, R. (2024b). Order effects in stimulus discrimination challenge established models of comparative judgement : A meta-analytic review of the Type B effect. *Psychonomic Bulletin & Review*, *31*(5), 2275-2284. https://doi.org/10.3758/s13423-024-02479-3

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nat Neurosci.*, *7(7)*, 773-778.

Goldstone, S. (1968). Production and reproduction of duration : Intersensory comparisons. *Percept Mot Skills*, *26(3)*, 755-760.

Goldstone, S., & Lhamon, W. T. (1972). Auditory-visual differences in human temporal judgment. *Perceptual and Motor Skills*, *34*(2), 623-633. https://doi.org/10.2466/pms.1972.34.2.623

Goldstone, S., & Lhamon, W. T. (1974). Studies of Auditory-Visual Differences in Human Time Judgment : 1. Sounds are Judged Longer than Lights. *Perceptual and Motor Skills*, *39*(1), 63-82. https://doi.org/10.2466/pms.1974.39.1.63

Grabot, L., & Kayser, C. (2020). Alpha Activity Reflects the Magnitude of an Individual Bias in Human Perception. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *40*(17), 3443-3454. https://doi.org/10.1523/JNEUROSCI.2359-19.2020

Grabot, L., & van Wassenhove, V. (2017). Time order as psychological bias. *Psychological Science*, *28(5)*, 670-678.

Ipser, A., Agolli, V., Bajraktari, A., Al-Alawi, F., Djaafara, N., & Freeman, E. D. (2017). Sight and sound persistently out of synch : Stable individual differences in audiovisual synchronisation revealed by implicit measures of lip-voice integration. *Scientific Reports*, *7(46413)*.

Ipser, A., Karlinski, M., & Freeman, E. D. (2018). Correlation of individual differences in audiovisual asynchrony across stimuli and tasks : New constraints on Temporal Renormalization theory. *Journal of Experimental Psychology: Human Perception and Performance*.

Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nature Neuroscience*, *13*(8), 1020-1026.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception*, *36*(14), 1-16.

Lustig, C., & Meck, W. H. (2011). Modality differences in timing and temporal memory throughout the lifespan. *Brain and Cognition*, *77*, 298-303.

Maaß, S. C., Schlichting, N., & van Rijn, H. (2019). Eliciting contextual temporal calibration : The effect of bottom-up and top-down information in reproduction tasks. *Acta Psychologica*.

Mamassian, P., & Landy, M. S. (2010). It's that time again. *Nature Neuroscience*, *13*(8), 914-916. https://doi.org/10.1038/nn0810-914

Matthews, W. J., & Meck, W. H. (2014). Time perception : The bad news and the good. *WIREs Cog Sci.*, *5(4)*, 429-446.

Miyazaki, M., Yamamoto, S., Uchida, S., & Kitazawa, S. (2006). Bayesian calibration of simultaneity in tactile temporal order judgment. *Nature Neuroscience*, *9(7)*, 875-877.

Murai, Y., & Yotsumoto, Y. (2016). Timescale- and Sensory Modality-Dependency of the Central Tendency of Time Perception. *PloS One*, *11*(7), e0158921.

Penney, T. B., Gibbon, J., & Meck, W. H. (2000). Differential effects of auditory and visual signals on clock speed and temporal memory. *Journal of Experimental Psychology. Human Perception and Performance*, *26*(6), 1770-1787.

Raviv, O., Ahissar, M., & Loewenstein, Y. (2012). How Recent History Affects Perception : The Normative Approach and Its Heuristic Approximation. *PLoS Computational Biology*, *8*(10), e1002731. https://doi.org/10.1371/journal.pcbi.1002731

Roach, N. W., McGraw, P. V., Whitaker, D. J., & Heron, J. (2017). Generalization of prior information for rapid Bayesian time estimation. *Proceedings of the National Academy of Science*, *114*(2), 412-417. https://doi.org/10.1073/pnas.1610706114

Sanford, E. C. (1888). Personal Equation. *A J Psychol*, *2(1)*, 3-38.

Seriès, P., & Seitz, A. (2013). Learning what to expect (in visual perception). *Frontiers in Human Neuroscience*, *7*. https://www.frontiersin.org/articles/10.3389/fnhum.2013.00668

Singh, G. S., & Acerbi, L. (2024). PyBADS : Fast and robust black-box optimization in Python. *Journal of Open Source Software*, *9*(94), 5694. https://doi.org/10.21105/joss.05694

Teufel, C., & Fletcher, P. C. (2020). Forms of prediction in the nervous system. *Nature Reviews Neuroscience*, *21*(4), 231-242. https://doi.org/10.1038/s41583-020-0275-5

Ulrich, R., Nitschke, J., & Rammsayer, T. (2006). Crossmodal temporal discrimination : Assessing the predictions of a general pacemaker-counter model. *Perception & Psychophysics*, *68*(7), 1140-1152.

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 B: Comparative and Physiological Psychology*, *51B*(2), 97-120.

Wearden, J. H., & Jones, L. A. (2021). Judgements of the Duration of Auditory and Visual Stimuli. *Timing & Time Perception*, *9*(2), 199-224. https://doi.org/10.1163/22134468-bja10008

Yamamoto, S., Miyazaki, M., Iwano, T., & Kitazawa, S. (2012). Bayesian Calibration of Simultaneity in Audiovisual Temporal Order Judgments. *PLOS One*, *7(7)*(e40379).

Yon, D., de Lange, F. P., & Press, C. (2019). The Predictive Brain as a Stubborn Scientist. *Trends in Cognitive Sciences*, *23*(1), 6-8.