Conceptually Plausible Bayesian Inference in Interval Timing

Sarah C. Maaß†[1,2,3], Leendert van Maanen[4] and Hedderik van Rijn[1,2]

[1] Department of Experimental Psychology, University of Groningen, Netherlands
[2] Behavioral and Cognitive Neurosciences, University of Groningen, Netherlands
[3] Aging & Cognition Research Group, German Center for Neurodegenerative Diseases (DZNE), Magdeburg, Germany
[4] Department of Experimental Psychology, Utrecht University, Netherlands

†corresponding author:
Sarah C. Maaß
s.c.maass@rug.nl
Grote Kruisstraat 2/1
9712TS Groningen
Netherlands

Abstract

In a world that is uncertain and noisy, perception makes use of optimization procedures that rely on the statistical properties of previous experiences. In Bayesian observer models, these previous experiences are typically modeled by unimodal statistical distributions. Here, we critically assess the validity of the assumptions underlying these models, and propose a model that allows for more flexible, yet conceptually more plausible, modeling of empirical distributions. By representing previous experiences as a mixture of log-normal distributions, this model can be parameterized to mimic different unimodal distributions and thus extends previous instantiations of Bayesian observer models. We fit the Mixture Log-Normal Model to published data of healthy young adults and a clinical population of aged Mild Cognitive Impairment patients and aged-matched controls, and demonstrate that this model better explains behavioral data and provides new insights into the mechanisms that underlie the behavior of a memory-affected clinical population. (146 words)

Keywords:
Interval timing
Central tendency effect
Aging
Bayesian Observer Model
Prior distributions

Conceptually Plausible Bayesian Inference in Interval Timing

In a world that is uncertain and noisy, perception makes use of optimization procedures to reduce the influence of moment-to-moment noise by incorporating statistical properties of previous experiences. This observation holds for the perception of many psychophysical quantities (Martin et al., 2017), including the estimation of distance (e.g., Wiener et al., 2016) and angles (e.g., Petzschner, et al., 2015), object size (Hollingworth, 1910), and duration (e.g., Jazayeri & Shadlen, 2010; Maaß et al., 2019b). These types of optimization procedures assume that when a specific stimulus needs to be reproduced, observers do not only take the current percept into account but also incorporate their prior knowledge of previous similar incidents to form an internal estimate of this stimulus. This process yields more optimal average responses when the perception of quantities is noisy, with the central tendency effect (Hollingworth, 1910) as its prime signature. Hollingworth's work focused on "time sense", describing that reproductions of durations gravitate towards the mean, with durations above the mean being underestimated and durations below the mean overestimated. Later work has linked this central tendency effect with the scalar property: As the noise increases with the size of an interval (Gibbon et al., 1984), the prior experiences will have a relatively larger influence on the longer percepts, and one would therefore expect a stronger central tendency bias for longer durations. Even though the central tendency effect was one of the first timing phenomena described in the literature, a formal account of this phenomenon has only recently been proposed.

In 2010, Jazayeri and Shadlen described how Bayesian principles can be used to construct an elegant mathematical framework in which an observer is assumed to reproduce a duration by integrating the perceived duration, represented as a distribution that can vary in noisiness (the likelihood distribution), with a probability distribution (the prior) representing the earlier observed durations. The multiplication of prior and likelihood results in the posterior distribution of which the mean is taken as the internal estimate of the to be reproduced duration (see Figure 1, left column). The observed reproduced duration is a function of this estimation and a noise component reflecting various noise sources. In Jazayeri and Shadlen's work, the prior is conceptualized as a uniform distribution defined over the stimulus range and effectively acts as a filter constraining the likelihood to the range of durations. To reflect the increased temporal uncertainty for longer intervals, the model assumes a linear scaling of the width of the likelihood as a function of the perceived duration. This implementation of the scalar property (e.g., Gibbon et al., 1984) allows the model to account for the empirical observation that the strength of the central tendency bias is a function of the magnitude of the duration. Similarly, this model accounts for individual differences in central tendency effects by assuming differences in the variability of the temporal percept: the noisier the likelihood, the stronger the impact of the prior, and thus the stronger the central tendency effect (see dark and light lines in the Likelihood and Posterior panels of Figure 1, and https://vanrijn.shinyapps.io/MaassVanMaanenVanRijn-Fig1 for a dynamic simulation demonstrating these effects). The second source of individual differences in Jazayeri and Shadlen's model is the "motor-noise" distribution that was estimated per participant to reflect the noise associated with the mapping from the mean of the posterior distribution to the actual reproduced duration.

Jazayeri and Shadlen's model is a specific instance of a family of Bayesian observer models. These models are characterized by a mapping between stimulus and response, in which the likelihood of an observed stimulus is weighted by a prior distribution over stimuli. Different distributions or functions could be hypothesized for the likelihood, the motor noise components, and the prior. Indeed, over the last couple of years, various Bayesian observer models have been shown to accurately reproduce human behavior in a number of timing tasks (for a review see Shi et al., 2013; Mamassian, & Landy, 2010; Acerbi et al., 2012; Cicchini et al., 2012; Gu et al., 2015; Roach et al., 2017). Most notably, Cichini et al. (2012) proposed an alternative Bayesian observer model in which the prior is represented by a Gaussian distribution of which the width varies as a function of the central tendency effect. This allows the model to partially capture the magnitude of the central tendency effect using a memory-based explanation, instead of solely relying on variability in the likelihood component (see the middle column of Figure 1, and see Narain et al., 2019, for an evaluation of this model).

In the current paper, we will evaluate the Bayesian observer models in terms of their construct validity and assess how the models can explain timing performance in different populations. In this context, construct validity refers to the question whether the computational models accurately reflect the constructs that account for variance in observed performance, similarly to the construct validity of a psychological test (see, for example, Cronbach and Meehl, 1955). Based on these evaluations, we will argue that the assumed prior distributions are, though mathematically elegant, too simple to account for performance in non-idealized laboratory settings and that certain assumptions related to perception and reproduction noise need to be reevaluated. Based on model simulations and comparisons to earlier published datasets of healthy young adults (Maaß & van Rijn, 2018) and an aged sample (Maaß et al., 2019a) that consists of patients diagnosed with Mild-Cognitive Impairment (MCI, a clinical diagnosis considered as precursor to Alzheimer's disease and dementia, Petersen et al., 2001) and healthy controls, we will argue that a prior consisting of a mixture of log-Normal distributions is preferred, and that the motor-noise associated with the reproduction of a duration needs to be estimated separately from the noise associated with the timing process itself.

First Theoretical Consideration: Shape of the Prior

The elegance of the specific Bayesian observer model proposed by Jazayeri and Shadlen (2010) is partially due to a number of simplifying assumptions. For example, the prior is assumed to be a uniform distribution spanning the range of the presented durations. Even though this provides computational simplicity, its theoretical suitability can be questioned as the nature of the central tendency effect is that extreme observations are pulled inwards. Thus, if one assumes that the prior is constructed on the basis of previous posterior distributions (following both Bayesian principles and instance-based explanations of the central tendency effect, e.g., Bausenhart et al., 2014; Taatgen & Van Rijn, 2011; for reviews, see Shi et al. 2013; Van Rijn, 2016), the resulting prior must have a higher mass at the center than at the more extreme values. Cicchini et al. (2012) addressed the issue of the uniform prior, and proposed a Gaussian prior distribution. Even though this distribution has characteristics that are theoretically more plausible than those of a uniformly distributed prior, its theoretical elegance is affected by the necessity to constrain the range of this Gaussian prior to prevent it extending to negative values. Moreover, even though the heavier center of the Gaussian captures the bias towards the mean, the repeated presentation of different durations that are all reasonably accurately reproduced should result in a prior distribution that is probably better characterized as a mixture of distributions centered at the presented durations. This rationale is supported by the observations of Acerbi et al. (2012) who have demonstrated that participants can acquire an internal approximation of more complex empirical distributions that notably deviate from either Gaussian or Uniform distributions as multiple peaks can be observed in the approximated prior distributions (see Figure 5 and the corresponding section in the Results and Discussion for a more elaborative review).

Thus, both theoretical considerations and empirical data suggest that the prior of temporal reproduction yields a mixture distribution. Even though a mixture of Gaussian distributions could be explored, this would require a decision about how to constrain the lower tail of these distributions. Moreover, empirical work has demonstrated that response functions in timing studies are better described by skewed distributions such as the Wald or inverse Gaussian, skewed-Normal, or log-Normal distribution (see, e.g., Oprisan & Buhusi, 2014; Simen et al., 2011; Van Rijn et al., 2014, and the discussion in Cicchini et al., 2012). Given these considerations, a theoretically reasonable prior would consist of a mixture of skewed distributions each representing a stimulus duration. Here, we explore a mixture-based prior that combines a log-Normal distribution for each unique duration that was presented to the participants. We will refer to this model as the Mixture Log-Normal (or MLN) Model.

The shape of the compound mixture distribution is defined by three parameters that can be directly mapped onto constructs. Two parameters are linked to the individual distributions: the location of the mean of the distribution and the distribution’s standard deviation. With respect to the means, we will assume that the location of the middle component corresponds with the duration of the middle stimulus, and that the locations of the other components reflect the central tendency effect, with more extreme distributions pulled inwards as a function of the magnitude of the central tendency effect. This could be a linear shift, but one could also assume a more flexible shift with shorter durations being more affected than longer durations. Here, we will assume that the means of the distributions will be distributed as a geometric series, resembling the type of distribution of the presented stimulus durations. The standard deviation of these components will scale linearly with the estimated means, following the scalar property. We will estimate the standard deviation for the component associated with the median stimulus duration, and derive the standard deviations for the other components. See the right column of Figure 1 for an illustration. The third parameter that could influence the shape of the overall compound mixture is the weight of each of the individual components. Even though this parameter could be used to capture non-uniform stimulus distributions, we will not explore this additional source of variability as the modelled datasets all presented uniform stimulus distributions (also see the section Fitting Empirical Prior Distributions in the Results and Discussion section).

Second Theoretical Consideration: Sources of Noise

A second consideration associated with the Bayesian observer model is that it incorporates two sources of noise, one associated with the perceptual processing of the presented duration (w_m), determining the width of the likelihood, and one associated with the mapping of the posterior to the actual reproduced duration (w_p). In other domains, w_m and w_p are typically conceptualized as representing perceptual noise and non-specific, motor, or response noise (e.g., Petzschner et al., 2015; Wiener et al., 2016). Importantly, the nature of interval timing tasks add a noise component in that the perception of duration is not momentaneous. Instead, a noisy timing system needs to be read out to perceive the duration of an interval, and the same or a similar timing system is involved during reproduction as the timing system needs to generate a cue upon which the motor response can be executed. This suggests that both w_m and w_p should additionally incorporate a “clock noise” component. Thus, w_m captures the perceptual noise associated with perceiving the onset and offset of the presented duration and clock noise associated with the actual timing of the interval, where w_p captures the perceptual noise for the onset of the reproduction phase, the clock noise, and the motor noise associated with the motor movement to mark the end of the reproduction phase. As the scalar property is reflected in a noisier clock estimates with increased durations, the noise associated with both perception and production stages should scale with the relevant duration. As motor noise is more variable and larger than visual perception noise (e.g., Amano et al., 2006), w_m should be smaller than w_p when these parameters are independently estimated as the production stage entails the combination of perceptual and clock noise (i.e., w_m) and the motor noise associated with the execution of the motor movement.

As both parameters were fit independently in Jazayeri and Shadlen’s Bayesian observer model (2010), w_p could take smaller values than w_m and no correlation between both parameters was enforced. In

Cicchini's et al.'s work (2012), neither of the noise parameters were estimated. Their model incorporated an empirical estimate for w_m that was based on each participant's performance on a bisection task, and w_p was fixed for all participants. Similar to Jazayeri and Shadlen's model, this model did not enforce that w_m should be smaller than w_p, nor that they are correlated. Additionally, by keeping w_p fixed over all participants, Cicchini et al. made the simplifying assumption that all sources of noise, including clock noise, are identical for all participants during reproduction. Moreover, as the reproduction noise was not influenced by the estimated duration, Cicchini's model implicitly assumes a timing system that does not adhere to the scalar property during reproduction. However, in the discussion of their model, Cicchini et al. already indicate that the noise should probably be estimated at a per-participant basis. Thus, following Cicchini et al. 's recommendations and Jazayeri and Shadlen's implementation, we do individually estimate the Weber fractions of both clock and motor noise. To provide a fair comparison between models, we extended our implementation of Cicchini et al.'s model to allow for individual differences in both perception and reproduction noise. As Cicchini et al. (2012) estimated reproduction noise at .1 (referred to as just motor noise), and allowed the clock noise to take on values higher than .1, we did not constrain the estimation procedure with respect to the relative values of w_p and w_m for our implementation of their model.

For the MLN Model, we - similarly to Jazayeri and Shadlen and Cicchini et al., scaled the standard deviations of the individual components by their means. Based on the theoretical considerations outlined above, we defined w_p as $w_m + \Delta w_p$, where Δw_p reflects the noise associated with the motor processes. Following earlier work, we assume the production noise to follow a Gaussian distribution, which results in $N(t_s, (w_m \times t_s))$ for the perception stage as clock noise is scaled by the presented duration, and $N(t_e, (w_m \times t_e) + \Delta w_p)$ for the reproduction stage.

Overview of the Paper

In the following sections, we will demonstrate how the MLN Model, including its theoretical constraints on the noise distribution, compares to two baseline models published in literature. These three Bayesian observer models take into account individual differences by assuming variations in clock (i.e., likelihood width) and motor noise, and, for two of the three models, allow for variations in the memory representation of previously perceived durations (i.e., width and/or shape of the prior). The first baseline model is an implementation of the model described in Jazayeri and Shadlen (2010) that assumes a uniform prior constrained over the range of presented durations and allows for an unconstrained estimation of w_m and w_p. We will refer to this model, after the shape of its prior, as the Uniform Model. The second baseline model is our implementation of Cicchini et al. 's (2012) model that assumes a Gaussian prior. Unlike the model reported in Cicchini et al. in which production noise was constant over participants and the perception noise was derived from a separate experiment, our implementation provides more freedom and allows for individual differences in both clock and reproduction noise. For all three models, we will integrate the likelihood and prior numerically over a range extending to 0.5 and 1.5 times the minimum and maximum stimulus durations. In addition, we will estimate a single parameter representing the prior's standard deviation for the Gaussian and MLN Model. As the central tendency effect assumes that the extreme distributions are pulled towards the mean, we estimate an additional parameter representing the inward pull for the means of the outer distributions in the MLN Model. This factor allows for the mean of the distributions of the prior to be located either at the original locations of the empirical distribution (resembling a wide prior, reflecting no central tendency effect), or, in the other extreme case, to all be identical to the median value of the empirical distribution (resembling a more narrow prior, reflecting a maximal central tendency effect).

Figure 1. Schematic representation of Bayesian Inference in interval timing. The left-most column depicts, from top to bottom, how the width of the likelihood determines the magnitude of the central tendency effect, and the middle and right columns depict how a similar effect can be obtained manipulating the prior. In all columns, the top figure depicts the likelihoods associated with perceived short or long durations. The second row depicts the prior, either a Uniform prior (left, cf. Jazayeri & Shadlen, 2010), a Gaussian prior (middle, cf. Cicchini et al., 2012), or the MLN prior. The third row depicts the posterior distribution with the dots reflecting the means of the distributions, which is the estimate (this figure does not depict the influence of the motor noise). The fourth row shows the resulting central tendency effects. In the left column, the light blue lines reflect the posterior resulting from a rather narrow likelihood distribution, representing a fairly accurate internal clock. The dark blue lines represent a noisier clock, resulting in a stronger central tendency effect. The Posterior and Response panels in the middle and right column only show the narrower likelihood distribution for legibility. In the middle plot, the pink, solid line represents a wider, and thus less informative prior, resulting in a relatively small adjustment of the likelihood. The purple, dashed line represents a more constrained prior, resulting in larger adjustments, and thus a larger central tendency effect. In the right plot, the pink, solid line represents a mixture in which three narrow distributions are distributed over most of the stimulus range, whereas the dashed, purple line represents the mixture of three wider distributions that are maximally pulled towards the mean. All mixture components are equally weighted.

The Uniform, Gaussian, and MLN Models will be fit to data from a sample consisting of 68 young healthy adults (undergraduate students) who performed a 1-second estimation task that was used to assess clock variability, and a multi-duration reproduction task (Maaß and Van Rijn, 2018, Figure 2). The multi-duration reproduction task consisted of reproduction of three equiprobable intervals (1.17, 1.4 and 1.68s). The three presented durations are defined as 1.4/1.2, 1.4, 1.4×1.2, following a geometric series with scale factor 1.4, common ratio 1.2, and calculated for terms -1, 0, and 1. As expected, the data demonstrates the central tendency effect, shown in Figure 2A.

Figure 2: Empirical results of the multi-duration reproduction task as reported in Maaß and Van Rijn (2018), Panel A, and in Maaß et al. (2019), Panel B. Panel A represents the data of 68 healthy young adults, Panel B represents the data of 10 Mild Cognitive Impaired patients (MCI) and 25 age-matched, healthy controls (HC). Data in both panels adhere to the central tendency

effect, with a larger central tendency effect of the MCI patients than for the HC in Panel B. t_s: presented interval; t_p: produced interval in seconds.

Furthermore, to test the assumption that a mixture model can provide new meaningful interpretations at group levels, the MLN Model was fit to the data of 10 Mild Cognitive Impaired (MCI) individuals and 25 aged matched healthy controls (HC) from Maaß et al., (2019a, Figure 2B). These participants performed the same 1-second estimation task that was used to assess clock variability and a multi-duration reproduction task as the students sample in Maaß and Van Rijn (2018). Interestingly, the memory impaired population showed a stronger central tendency effect than the healthy controls, The observed negative correlation between the magnitude of the central tendency effect and memory functioning might be considered paradoxical as it implies that with decreasing memory functioning, the prior - which can be seen as the memory for previously perceived stimuli- has stronger influence on subsequent performance (see Maaß et al. 2019a for a discussion). As the temporal production variability as assessed by the 1-second estimation task was similar among all groups, these effects cannot be easily explained by higher clock variance. Thus, the results could indicate that MCI patients have a stronger influence of prior experiences than age-matched HC resulting in stronger central tendency effects. The MLN Model that we present here will quantify these counterintuitive effects by disentangling memory and clock variability processes.

Methods

The Bayesian Observer Model

We fit Bayesian observer models to the data from the experiments introduced above. All models follow the rationale from Jazayeri and Shadlen (2010) that a reproduced temporal interval on a particular trial depends on (1) the relationship between the sample interval t_s and the internal measurement of this interval (t_m), (2) the mapping of t_m onto an internal representation of the perceived duration (t_e) by incorporating the prior, and (3) the incorporation of noise to map the internal representation to the produced interval (t_p). The likelihood of a specific internal representation of the sample interval is modelled as $p(t_m|t_s) = \varphi\left(\frac{t_s-t_m}{w_m t_s}\right)$. Here $\varphi(\cdot)$ represents the standard Normal distribution function, and w_m is the Weber fraction associated with the internal representation of the sample interval (i.e., perception clock noise). The scalar property is captured by the multiplication of w_m and t_s, as this results in increasing uncertainty with longer durations. Likewise, the likelihood of a specific reproduced duration is modelled as $p\left(t_p\middle|t_e\right) = \varphi\left(\frac{t_e-t_p}{w_p t_e}\right)$. Here, w_p is the Weber fraction associated with the reproduction, providing another locus for the scalar property. The mapping from the internal measurement (t_m) to the internal estimate (t_e) depends on the integration of the likelihood over a prior distribution. The specification of the prior distribution will differ according to the different theoretical considerations discussed above, resulting in three distinct models.

First, the *Uniform Model* is a direct implementation of the Jazayeri and Shadlen (2010) model that assumed a uniform distribution of the prior. The range of the uniform distribution is determined by the stimulus range (i.e., ranging from 1.17 to 1.68 seconds in the experiments we report on here). Second, the *Gaussian Model* assumes a normally distributed prior (cf. Cicchini et al. 2012). The prior is centered at the median stimulus duration (i.e., 1.4 seconds), and truncated below 0. The parameter that is estimated is the standard deviation (p_{sd}) of the normally distributed prior. Third, the *Mixture Log-Normal Model* (MLN Model) incorporates the theoretical considerations that were previously introduced. Specifically, the MLN Model is characterized by a mixture of multiple log-Normal distributions. We assume that the number of mixture components equals the number of empirically presented durations. The standard deviation of the mixture components is estimated under the constraint that they scale linearly with the component means (following the scalar property). Consequently, below we report the standard deviation (p_{sd}) of the middle component and the common ratio p_r of the geometric series that scales the mixture components assuming a scale factor of 1.4 and terms -1, 0, and 1. Finally, we assume that the weight of each mixture component is equal, representing the fact that participants have observed every stimulus an equal number of times.

For all three models, we additionally estimate the Weber fraction parameters w_m and w_p. For the Uniform and Gaussian Models, both w_m and w_p are unconstrained, conforming to the specifications in Jazayeri and Shadlen (2010) and Cicchini et al. (2012). However, the MLN Model incorporates our second theoretical consideration that w_p should reflect the combination of clock noise captured by w_m and additive motor noise. We therefore estimate Δw_p and define w_p as $w_m + \Delta w_p$.

Model Fitting & Model Selection

For each participant, we performed a linear grid search for the optimal set of parameters (e.g., Becsey et al., 1968; Lerman, 1980; Mestdagh et al., 2019). The range as well as the resolution of each parameter are presented in Table 1. The ranges are based on an iterative process to ensure an optimal resolution and to prevent ceiling and floor effects. The number of values tested in the range was chosen to provide a fine enough grid, while still being able to search the full grid in a reasonable amount of time. For each parameter vector we generated 1,000 trials for each stimulus duration, and computed the RMSE between the observed and predicted 10%, 30%, 50%, 70%, and 90% quantiles separately for every participant. As we are mainly interested in the central tendency effects, and not in any linear shifts that represent structural over- or under-reproduction, we subtracted, for each reproduction, the average reproduction time across all trials and added the average stimulus duration (following Cicchini et al., 2012). The integration of the likelihood over the prior distribution was approximated using the Riemann sum/rectangle method over the range 0.585-2.52s, to avoid misspecification where the posterior density approaches zero (implementation details of the models, grid search data, code for all figures, and fitting routines can be found online: https://osf.io/kqjxf/, Maaß et al., 2020).

It is common practice to compare statistical models in terms of their quantitative fit to the data, taking into account any differences in model complexity. Here we take a different approach, as our research question is not about the model that provides the best balance between goodness-of-fit and model complexity, but about which model provides the best understanding of the empirical phenomena observed in healthy and clinical populations while adhering to plausible theoretical considerations. For that reason, we will consider the validity of the estimated parameters vis-a-vis their interpretation to differentiate between models.

Table 1: Parameter ranges for the estimated parameters (with the number of tested equidistant values listed in parenthesis)

Model	w_m	w_p	Δw_p	p_{sd}	p_r
Uniform	[0.005-0.2] (20)	[0.005-0.2] (20)			
Gaussian	[0.005-0.2] (20)	[0.005-0.2] (20)		[0.01-0.3] (20)	
MLN	[0.005-0.2] (20)		[0.0001-0.15] (20)	[0.01-0.3] (20)	[1.02-1.2] (10)

Bayes Factors & Model Averaging

When comparing model fits to the data of individual participants, we will plot the empirical data and the model estimates for the vector of parameters with the lowest RMSE to reflect the best possible fit we could obtain. However, this depiction of the model fit ignores the uncertainty in model selection: it is well possible that another vector of parameters provides similar fits in terms of RMSE. Using Bayesian model averaging we calculated a weighted average of the parameters (weighted by their respective model probability using relative AIC values) that best reflects the information provided by the model fitting processes (e.g., Hinne et al., 2019; Hoeting et al., 1999).

The Bayes factors were computedwith the R package BayesFactor (version 0.9.12-4.2; Morey et al., 2015) using the default prior settings. and are interpreted based on the guidelines provided by Jeffreys (1961, see also van Doorn et al., 2019). The reported Bayes factors summarize the extent to which an observer's opinion of the tested variable should change based on the data. A Bayes factor of 1 indicates that both hypotheses are equally likely under the data and therefore is inconclusive. Bayes factors larger than 1 represent evidence for the alternative hypothesis of an influence of the tested independent variable on the dependent variable, and Bayes factors less than 1 represent evidence for the null hypothesis of no effect of the tested variable.

Results and Discussion

Fitting the Student Sample

First, we will fit the three models to the data of a sample of 68 healthy young adults (undergraduate students). We determined the best fitting vector of parameters per participant by minimizing the RMSE for each quantile and empirical duration per participant and model. Figure 1 depicts in the three leftmost columns the empirical data and the model fits resulting in the lowest RMSE for three participants. In addition, we plotted in the rightmost column of Figure 1 the predicted against empirical reproduced duration for each quantile (coded by color) and empirical duration (each diagonal line representing a t_s). At first sight, all models seem to provide an equally good fit, as no strong deviations can be observed from the diagonal. Yet, when comparing the squared difference between predicted and empirical duration per type of model (RMSEs: 0.032 for Uniform, 0.028 for Gaussian, 0.027 for MLN), there is strong evidence (BF=11.31) that this squared deviation is better predicted when type of model is included as main effect (in addition to the main and interaction effects of quantile, t_s, and empirical duration). This effect is driven by a worse fit for the Uniform Model, as there the Gaussian and MLN Models provide comparable fits (BF=0.055). Thus, although all three models provide a reasonable visual fit to the data, the Gaussian and MLN Models better account for the data than the Uniform Model.

Figure 3: Fit of the models to the empirical data from Maaß and Van Rijn (2018). The three leftmost columns depict the 10%, 50%, and 90% quantile (in blue) and model predictions (in black), for three example participants (t_s: presented interval; t_p: produced interval in seconds). The three rows represent the Uniform, Gaussian, and Mixed Log-Normal Model respectively. The rightmost panel depicts, for each of the three t_s durations (one diagonal line per t_s duration, longest at top), the difference between predicted quantiles and the observed empirical quantiles for all 68 participants, with color scale representing the 10 (black), 30, 50, 70 and 90% (blue) quantiles.

As discussed in the introduction, the Uniform and Gaussian Model do not constrain the estimation of the width of the noise of the reproduction (w_p) in relation to the width of the noise of the perception (w_m) of the duration. As a reproduction consists of the same clock processes associated with the perceptual stage plus an additional motor response, one should expect w_p to be larger than w_m. Figure 2 depicts the difference between w_m and w_p, expressed as Δw_p, for the Uniform and Gaussian Models, and the estimated Δw_p for the MLN Model. As the two leftmost violin plots show, a sizable proportion of the best fitting models per participant assume a *larger* w_m than $w_{p,}$ suggesting that for those participants the reproduction of a duration is associated with less noise than the perception of a duration. For the MLN

Model, w_p is defined to be larger than w_m, as this model estimates Δw_p, as a positive value, that is added to w_m to arrive at w_p. It is important to note that this theoretically implausible result is intrinsic to the Uniform Model as the only way this model can account for the magnitude of the central tendency effect is by manipulation w_m. As the prior is fixed, w_m captures the central tendency effect: A very small w_m yields a very peaked likelihood which mostly falls within the defined uniform prior, and thus only small central tendency effects will be observed, whereas a large w_m will results in large proportions of the likelihood falling outside the range of the prior, resulting in stronger pull effects. Thus, for participants with large central tendency effects, w_m needs to be large. If w_p would be constrained to be larger than w_m, this would result in too dispersed estimates for the extreme quantiles, and thus w_p will often be smaller than w_m when a participant shows a large central tendency effect. Note that for the Gaussian Model, constraining w_p might be an option as central tendency effects might also be captured by manipulating the width of the prior, however, we opted for keeping the models as similar as possible to the earlier published descriptions. This results in two models with parameter estimates that, even though they provide reasonable fits to the data, are difficult to align with the theoretical constructs they represent.

Figure 4: Estimated additional noise (Δw_p) associated with the reproduction of a duration compared to the perception of a duration (w_p) for the three models (Whisker plot represents mean and two standard deviations). For the Uniform and Gaussian Model, Δw_p is calculated by subtracting the estimated w_m from the estimated w_p; for the Mixture Log Normal Model, Δw_p is estimated as additive noise to w_m when the estimated duration is mapped to the reproduced duration. Green dots represent participants for whom the estimated combination of clock and motor noise is *larger* than just clock noise, red dots represent participants for whom the models estimated lower noise for reproduction than for perception.

Fitting Empirical Prior Distributions

A distinction between the Uniform and Gaussian Models on the one hand, and the MLN Model on the other is that the latter can accommodate more irregular or complex prior distributions. These distributions could either be elicited by stimuli sampled from non-uniform distributions, or because of observing behavior that does not align with the assumption of an ideal Gaussian or uniform prior. To assess how well the MLN Model fits such prior distributions, we fitted this model to the nonparametrically estimated prior distributions of Experiment 1 reported by Acerbi et al. (2012) that are shown as the dashed lines in Figure 5. The top-left and both bottom panels (Figure 5A, 5C, & 5D) represent the prior derived from the participant's reproductions given either a short (purple/left) or a long (pink/right) uniform stimulus distribution consisting of six stimulus durations. These priors are characterized by a central peak flanked by two lower peaks or shoulders, suggesting (1) that the central tendency effect resulted not just in a behavioral pull, but that this pull is also reflected in the central peak of the prior itself, and (2) that the more extreme stimulus durations might be encoded separately from a representation of the mean of the distribution, resulting in the two subordinate peaks. The dashed lines in the top-right panel (Panel B) represent both a biased (left/green) and an uniform distribution (right/purple), tested over the same range of stimulus durations. For the biased distribution, the second shortest stimulus had a 7/12 probability of being presented, while the other five stimuli durations were each presented with a probability of 1/12. The relative frequency with which the stimuli durations are presented is shown in the small vertical bars just above the x-axis. The estimated prior distributions clearly show multiple modes, again suggesting that a mixture distribution would best describe these priors, with the biased distribution shifted in the direction of the positively biased duration. The bottom two panels (Panel C and D) depict the two uniform empirical distributions with slightly different parameterized mixture log-Normal priors.

Figure 5: The nonparametrically estimated prior distributions of Experiment 1 (Panel A, C, and D) and 2 (Panel B) of Acerbi et al. (2012) and the fits of 10 runs of the MLN Model. For the fits in the top two panels, location and weight of each of the components was independently estimated, in the bottom two panels either weight (C) or location (D) was kept constant. See text for further explanation.

The partly transparent lines plotted in the four panels of Figure 5 represent 10 instances of MLN priors. Each prior consists of 6 component distributions, one for each stimulus duration presented, with a weight between 1/51 and 10/15. The weights, locations, and a single standard deviation were estimated by minimizing the summed squared differences between the nonparametrically estimated prior distributions and the mixture log-Normal prior distribution over the 200 to 1600 ms range (using the Simplex procedure, Nelder & Mead, 1965). In the top row, the standard deviation and both location and weight of each of the components are independently estimated. In the bottom row, the left panel depicts the fit assuming that the weights of each component are equal, whereas the right panel depicts the fit assuming that the locations of the components equate the presented t_s values. The vertical bars extending from 0 downwards represent the location (x value, jittered for visualization purposes) and weight (length of line) of each of the components of the mixture. The small lines at the bottom of each panel represent the relative proportion and location of the presented, empirical durations. By plotting 10 fitted models, the partly transparent lines provide a visual indication of the variability and flexibility of the model fits.

Figure 5A demonstrates that the central peak in the estimated prior distribution is captured by a clustering of relatively heavily weighted component distributions around the mean of the empirical distribution, with the subordinate peaks captured by two smaller clusters to the sides. In Figure 5B, the downwards pointing location and weight indicators show that the central cluster of the biased condition plotted in green is slightly more scattered and shifted leftwards. These observations demonstrate that the MLN prior can fit the qualitative patterns of the nonparametrically estimated priors reported by Acerbi et al. (2012). To assess whether both locations and weights need to be estimated to provide a reasonable fit, Panel C and D represent model fits in which either the locations are estimated but the weights kept constant (Panel C), or, vice versa, the weights were estimated but the locations kept constant. As can be seen by comparing both panels, Panel C fares better in capturing the kurtosis and multimodality of the estimated prior. Panel C can be interpreted as a shift of the mean or location of the prior component that is associated with a particular stimulus duration, a view that fits with the theoretical assumption that the prior reflects the history of posterior values which are also pulled towards the center of the distribution. To conclude, this section demonstrates that the MLN Model is capable of capturing realistic, more complex prior distributions. When comparing Figure 5C and 5D, visual inspection suggests that the simulation that estimates the locations provides a better fit than the simulation with free weights. To prevent unnecessary flexibility, we will therefore estimate a factor representing the pull exerted on the component distributions, yet not estimate the relative weights of the components.

Fitting the MCI Sample

After confirming that the MLN Model matches the data of the young adult population and the more complex empirical priors described in Acerbi et al. (2012), we now assess whether this model can provide additional insight in empirical phenomena and, vice versa, whether the parameters associated with the models relate in a sensible manner to external constructs. Hereto, we fitted the MLN Model to the clinical data collected by Maaß et al. (2019a). In this work, we demonstrated that the participants diagnosed with MCI demonstrated a stronger central tendency effect than age-matched, healthy control (HC) participants. This result might seem paradoxical as MCI status is partially defined by memory dysfunctioning, whereas a stronger central tendency effect assumes an emphasized influence of memory. By fitting the MLN Model to this data, we hope to elucidate the locus of the increased central tendency effect by assessing whether a difference can be observed in the priors associated with both participant groups, or whether this effect is driven by noisier clock systems. Figure 6 presents the fits of the MLN model to three MCI (top row) and three control (bottom row) participants. The right-most column again shows the model fit for all participants, again demonstrating that the deviation between predicted and observed responses is relatively small for all quantiles. Figure 7 plots the distributions of the four parameters of the MLN Model for the MCI and the HC participants, with in the top left corner the Bayes factor representing the evidence in favor of a difference between both distributions as a function of MCI status (assessed with lmBF, Morey et al., 2015). Importantly, while there is a difference observed between both groups for w_m (internal clock noise), there is no evidence for a difference for w_p between both groups. This finding is in line with earlier work that has demonstrated that clock noise increases as a function of cognitive decline (Nichelli et al., 1993; Wearden et al., 1997; Malapani et al., 1998; Caselli et al., 2009; Gooch et al., 2009; Wild-Wall et al., 2009; Turgeon & Wing, 2012). However, and most notably, this contrasts our earlier report in which we argued that there was no difference between both groups in terms of clock noise when we determined clock noise using the 1-second task.

Figure 6: Fit of the MCI Model to the empirical data from Maaß et al. (2019a). The three leftmost columns depict the 10%, 50%, and 90% quantile of the empirical data (blue) and MLN Model predictions (black) for three example participants (t_s: presented interval; t_p: produced interval in seconds). The two rows represent the two participant groups (MCI and HC participants). The rightmost panel depicts, for each of the three t_s durations, the difference between predicted quantiles and the observed empirical quantiles for all HC (n=25) and MCI (n=10) participants, with color scale representing the 10 (black), 30, 50, 70 and 90% (blue) quantiles.

Figure 7: MLN Model parameters for the HC and MCI groups from Maaß et al. (2019a). Violin and whisker plot (representing mean and two standard deviations) represent model parameters estimated by means of grid search and Bayesian model averaging for w_m, w_p, p_r, and p_{sd} by participant group. Dots (jittered for visualization) represent individual participants. The BF value in the top-left of each panel reflects the Bayes Factor in favor of Participant Group predicting that panel's model parameter determined by the lmBF function of the BayesFactor package (Morey et al., 2015).

In the 1-second task, part of both data sets reported in this manuscript (Maaß & Van Rijn, 2018; Maaß et al., 2019a), participants are asked to repeatedly produce a duration of one second by means of a keypress. In our earlier work, we have argued that the measure derived from this task "is related to the width of the likelihood distribution in the multi-duration reproduction task, which has been associated with the noise in the clock parts of the temporal system" (p. 8, Maaß & Van Rijn, 2018). This predicts, in terms of the discussed Bayesian observer models, that the observed measure should correlate with w_m as this measure indexes clock noise, but also with w_p in the Uniform and Gaussian Models, as w_p reflects both clock- and motor noise in these models. However, shown in the columns for the Uniform and Gaussian Model of Table 2, the only reliable correlation is found between w_m and the 1-second task for the Uniform Model, but no reliable results are observed for w_p in either model. Instead, the distribution width parameter of the Gaussian Model (p_{sd}) is negatively correlated to the 1-second variance, indicating that a higher variance during the 1-second task results in stronger pull-effects due to a stronger prior influence. Obviously, this pattern results in the empirical observation of the stronger pull for participants with higher 1-second variance measures (Maaß & Van Rijn, 2018, Maaß et al., 2019a), but it is difficult to conceive a mechanism that can both explain an increased 1-second production variance while at the same time *only* predicting a more narrow prior and no reliable clock-noise effects. Interestingly, the MLN Model does show a correlation between 1-second variability and w_m, but not between 1-second variability and Δw_p. As the latter parameter represents motor-noise, this pattern of results supports the notions forwarded in Maaß and Van Rijn (2018) that the 1-second variability measure captures clock noise, and provides conceptual support for the internal validity of the MLN Model. However, no reliable relations can be found for the MCI sample. This could be due to the smaller number of participants, or because this sample consists of multiple subsamples (i.e., healthy aged, non-diagnosed but memory-affected, and MCI-diagnosed participants, Maaß et al., 2019a). As analysing these subsamples did not provide qualitatively different results, we refrain from interpretation, but recommend to reassess this relation when data of a larger sample of participants are available. Obviously, an alternative explanation is that the results of the 1-second task are influenced by motor-noise, which might be stronger in the aged population than in the young-adult population. The results reported in Table 2 hint in that direction, as where there is positive evidence for w_m and negative evidence for w_p relating to 1-second production noise in the young-adult population, the effects are reversed for the aged population. This argues that stronger effects of motor noise in aged populations could explain the disparity between the results in Figure 7A (where we find w_m differences between MCI and HC participants) and the earlier reported (Maaß et al., 2019a) absence of clock noise differences between these groups.

Table 2: Correlations between 1-second task measure to predict clock noise (see Maaß & Van Rijn, 2018) and the model parameters. The w_p / Δw_p row lists w_p for the Uniform and Gaussian Models, and Δw_p for the MLN Model. The *r* column lists Pearson's product moment correlation coefficient, the BF column lists the Bayes Factor determined by Jeffreys (1961) test for linear correlation as implemented in the correlationBF function of the BayesFactor package (Morey et al., 2015).

	Maaß & Van Rijn (2018) (n=57, young healthy adults)						Maaß et al., (2019a) (n=30, MCI & HC)	
Model	**Uniform**		**Gaussian**		**MLN**		**MLN**	
	r	BF	*r*	BF	*r*	BF	*r*	BF
$\boldsymbol{w_m}$	**.38**	**16.247**	.21	0.907	**.32**	**5.046**	.19	0.622
$\boldsymbol{w_p}$ / $\boldsymbol{\Delta w_p}$	.20	0.845	.26	1.885	.11	0.403	.33	1.649
$\boldsymbol{p_{sd}}$			**-.39**	**22.936**	-.19	0.763	.36	2.210
$\boldsymbol{p_r}$					-.15	0.546	.24	0.837

Apart from the difference between MCI and HC for the w_m parameter, Figure 7C also depicts a difference for the p_r parameter which reflects the spread of the individual components contributing to the MLN prior. The smaller p_r values estimated for the MCI participants indicate that their prior is more compact, resulting in a stronger memory-based central tendency effect. This is in line with the conclusions drawn on the statistical analysis of the behavioral data reported in Maaß et al. (2019a), and thus provides additional support for the hypothesis that in this memory-affected population the internal representation of earlier experiences is weighted more strongly than in healthy control populations. Whereas Bayesian observer models are typically considered to represent optimally as the integration with the prior compensates for a lack of accuracy, in this clinical population compensation might be needed to counter the decay of the memory trace of the current interval.

General Discussion and Conclusion

We demonstrated and argued that Bayesian observer models that allow for interindividual variability in the shape of the prior outperform a Bayesian observer model that assumes a uniform prior, but that only the MLN Model adheres to theoretical constraints regarding perception and production noise. In addition, this more realistic MLN Model provides new insights regarding the differences between Mild-Cognitive Impaired patients and healthy controls in a context task. Whereas in our original work we attributed the increased context effects uniquely to a more narrow prior for MCI patients, the computational cognitive model demonstrated that the combination of a narrower prior *and* a noisier internal clock drive the observed patterns. The components that make up the prior are more strongly pulled inwards, rather than represented by a narrower distribution (as would be the case in a Gaussian model). These results thus refine the conclusions that were drawn on the basis of a statistical analysis of the behavioral data (Maaß et al., 2019, see also Rueda & Schmitter-Edgecombe, 2009) and exemplifies the observation by Paraskevoudi et al. (2018) that insight in deficient timing processes hinges upon a formalized approach that allows for dissecting whether memory or clock-noise processes drive deviations in timing performance.

Even though we presented the Gaussian and MLN Models as two distinct models, the former could be seen as a special case of the latter. That is, apart from the distributional differences between a Gaussian and Log-Normal, the Gaussian Model can be approximated by an MLN Model with $r = 1$, as this would assume completely overlapping mixture prior distributions. However, the model simulations estimated r to be 1.05 or higher, indicating that even in the participants with the strongest prior-based central tendency effect, the components were not fully overlapping (i.e., a $r = 1.05$ results in distribution means of 1.33, 1.4, and 1.47). Consequently, the estimated distributions are more leptokurtic than what would be obtained with a Normal distribution.

In addition, the Gaussian distribution has been shown to fail to predict empirical patterns observed in studies that use a dense representation of intervals. In a number of such studies, it has been demonstrated that the central tendency effect may not be a linear deviation from the target duration, but rather follows a more sigmoid pattern, where more extreme data points gravitate towards the mean more than less extreme data points (Narain et al., 2018; Sohn et al., 2019). Narain et al. demonstrated that the sigmoid pattern is conditional on a uniform prior as a Gaussian prior does not yield sufficient curvature. Our work extends the observations forwarded by Narain et al., as it supports the observation that a prior distribution with a sudden transition from negligible prior density to a higher prior density yields a sigmoid pattern in central tendency. In the extreme case, when many underlying components are assumed to contribute to the prior, the resulting MLN prior distribution mimics a uniform distribution (Figure 8A). The degree of mimicry to the uniform prior depends on the width of each component, with narrower components resulting in a distribution that more closely resembles the step-shaped boundaries of the uniform distribution closer. When wider components are assumed, the resulting prior more resembles a Gaussian distribution, yielding a weakening of the sigmoidal pattern (Figure 8B). However, a high density of intervals is not required to observe the sigmoidal pattern. When, following the experimental paradigm of Sohn et al. (2019), a smaller number of underlying components (Figure 8C) is assumed, the MLN Model still results in sigmoidal response patterns (Figure 8D), the strength of which is again a function of the width of the mixture components. Thus, the MLN Model reveals that it is not the uniform-like prior distribution per se that yields the sigmoid pattern in central tendency, but a sudden transition from negligible to increased prior densities. Moreover, it suggests that the strength of the sigmoid pattern appears as a result of individual differences in timing behavior: Individuals with more precise timing capabilities may develop prior distributions that have narrower mixture components, required for a sudden transition of prior density to demonstrate strong curvature effects.

Figure 8: Simulations with the MLN Model show that the model predicts complex behavioral patterns. Simulations plotted in the top row are based on an MLN prior consisting of 21 equidistant mixture components. With a sufficiently low standard deviation (narrow components drawn in black/grey lines), the resulting mixture resembles a uniform distribution (solid line). A larger standard deviation (wide components) results in a more unimodal distribution (dashed line). The right panel depicts that an increased width of the components results in increased nonlinearity in the central tendency effect. The simulations in the bottom row are based on an MLN prior consisting of 5 mixture components (similar to Sohn et al., 2019). If the components have a sufficiently low standard deviation, the resulting mixture has a density that is negligible outside the stimulus range (solid purple line). A larger standard deviation (wide components) results in a more unimodal distribution (dashed line). Again, the different shapes of the MLN prior predict differences in the central tendency effect.

One challenge when comparing different computational cognitive models, or variants of a single computational cognitive model, is to account for the degrees of freedom provided by additional parameters or assumed processing steps (Pit & Myung, 2002; Myung, 2000). The same holds for the three models presented in this manuscript, as it is not straightforward to provide an objective assessment of the exact number of degrees of freedom for each model. One could, for example, consider the parameters determining the width of uniform prior as free-but-not-manipulated parameters in the Uniform Model (cf. Taatgen & Anderson, 2008, Van Rijn et al., 2016): One could imagine an uniform prior to either extend beyond the maximum range of the presented durations if one assumed that a highly noisy internal clock blends out the perceived durations, or assume a more narrow uniform prior if one assumes that the estimated (or reproduced) durations would provide the basis for the prior. One could argue that this latter assumption is more in line with Bayesian reasoning as it fits well with the theoretical assumption that the model's posterior on a current trial contributes to the prior on future trials (e.g., Shi et al., 2013). As the posterior will demonstrate the central tendency effect, one could hypothesize that the prior should also be narrower. Simply iterating over the number of estimated parameters is also problematic when considering w_m and w_p. All three models assume two parameters for the noise associated with the perception and production stages, but only the MLN Model constrains the latter to be larger than the former, reducing the model's flexibility. Instead of providing formal comparisons based on model complexity or flexibility, here we compared the three different models on construct validity. Even though the mathematical simplicity of the Uniform and Gaussian Models could be taken as arguments in favor of their adoption, and both models fare well in cases of well-designed and balanced laboratorium paradigms, neither model allows for capturing more complex behavioral patterns that can readily be observed in empirical studies (e.g., Acerbi et al., 2012). Moreover, especially in the case of the Uniform Model as implemented here, central tendency effects are defined as a function of clock noise, as only by increasing the width of the likelihood can this model explain stronger central tendency effects. This rules out the possibility that memory-based explanations drive observed central tendency effects, which we argue is an important explanatory variable when considering the behavioral patterns observed in clinical populations. Additionally, a fixed prior also means that in the case of a strong central tendency effect but limited reproduction variance, the Uniform Model has to assume a highly variable clock for the perception phase to account for the strong central tendency effect, but a very precise clock for the reproduction phase to capture the lack of variance. This dissociation between perception and reproduction clock noise is, to our knowledge, not supported by any empirical evidence. In sum, the models presented in this manuscript provide a hierarchy of conceptually plausible models, and demonstrate that the different components of the MLN Model provide the best tools to understand the clock and memory mechanisms involved in interval timing.

Conclusions

To conclude, a Bayesian observer model that assumes a prior that consists of a mixture of log-Normal distributions outperforms a Bayesian observer model based on a Uniform prior, and can fit *and* explain a number of empirical phenomena not captured by either a model based on an uniform or Gaussian prior. The MLN Model does not just explain how Bayesian integration could take place, but also provides a theoretically sensible foundation for the shape of the prior. By fitting the prior to the observed behavior, the MLN Model provides a mechanistic account of how memory influences Bayesian integration. In addition, this model allows for a functional separation of clock noise, which is mostly driving the noise during the perception phase, and motor noise, which is an additional noise component during the reproduction phase. With respect to the distribution of the prior, specific parameterizations of the MLN prior resemble the uniform and Gaussian priors that have been proposed earlier, indicating that the MLN Model can capture the phenomena earlier ascribed to these more specific models while adhering to more stringent construct validity criteria. Thus, even though the MLN Model is more complex than the existing models, theoretical and empirical considerations justify this model over the simpler models. Lastly, we demonstrated that this MLN Model allows for a more precise interpretation of the behavioral results in a clinical population, paving the way for the utilization of computational cognitive models (cf., Huys et al., 2016) to assess the relative contribution of memory and clock components in declining performance in clinical populations.

Acknowledgements: none

Contributions: All authors contributed to the development of ideas, theoretical underpinnings, modeling of data and to the preparation of the manuscript.

Competing Interests: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding: Maaß and Van Rijn were supported by the research program "Interval Timing in the Real World: A functional, computational and neuroscience approach" with project number 453-16-005, awarded to Van Rijn, which is financed by the Netherlands Organization for Scientific Research (NWO).

Data Availability Statement: The datasets analyzed for this study can be found in the Open Science Framework (https://osf.io/kqjxf/).

References

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). doi:10.1371/journal.pcbi.1002771

Amano, K., Goda, N., Nishida, S. Y., Ejima, Y., Takeda, T., & Ohtani, Y. (2006). Estimation of the timing of human visual perception from magnetoencephalography. Journal of Neuroscience, 26(15). 3981-3991. doi:10.1523/JNEUROSCI.4343-05.2006

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. doi:10.1016/j.actpsy.2013.06.011

Becsey, J. C., Berke, L., & Callan, J. R. (1968). Nonlinear least squares methods: A direct grid search approach. *Journal of Chemical Education*, *45*(11), 728. doi:10.1021/ed045p728

Caselli, L., Iaboli, L., & Nichelli, P. (2009). Time estimation in mild Alzheimer's disease patients. *Behavioral and Brain Functions*, *5*(1). 32. doi:10.1186/1744-9081-5-32

Cicchini, G. M., Arrighi, R., Cecchetti, L., Giusti, M., & Burr, D. C. (2012). Optimal encoding of interval timing in expert percussionists. *Journal of Neuroscience*, *32*(3), 1056-1060. doi:10.1523/JNEUROSCI.3411-11.2012

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin*, *52*(4), 281. doi:10.1037/h0040957

Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. *Annals of the New York Academy of sciences*, *423*(1), 52-77. doi:10.1111/j.1749-6632.1984.tb23417.x

Gooch, C. M., Stern, Y., & Rakitin, B. C. (2009). Evidence for age-related changes to temporal attention and memory from the choice time production task. *Aging, Neuropsychology, and Cognition*, *16*(3), 285-310. doi:10.1080/13825580802592771

Gu, B. M., Jurkowski, A. J., Lake, J. I., Malapani, C., & Meck, W. H. (2015). Bayesian models of interval timing and distortions in temporal memory as a function of Parkinson's disease and dopamine-related error processing. In *Time Distortions in Mind* (pp. 281-327). Brill. doi:10.1163/9789004230699_012

Hinne, M., Gronau, Q. F., van den Bergh, D., & Wagenmakers, E. J. (2019). A conceptual introduction to Bayesian model averaging. doi:10.31234/osf.io/wgb64

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. *Statistical science*, 382-401. doi:10.1214/ss/1009212519

Hollingworth, H. L. (1910). The central tendency of judgment. *The Journal of Philosophy, Psychology and Scientific Methods*, *7*(17), 461-469. doi:10.2307/2012819

Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nature neuroscience*, *13*(8), 1020. doi:10.1038/nn.2590

Jeffreys, H. (1998). *The theory of probability*. OUP Oxford.

Lerman, P. M. (1980). Fitting segmented regression models by grid search. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *29*(1), 77-84. doi:10.2307/2346413

Maaß, S. C., & Van Rijn, H. (2018). 1-s productions: A validation of an efficient measure of clock variability. *Frontiers in human neuroscience*, *12*, 519. doi:10.3389/fnhum.2018.00519

Maaß, S. C., Riemer, M., Wolbers, T., & van Rijn, H. (2019a). Timing deficiencies in amnestic Mild Cognitive Impairment: Disentangling clock and memory processes. *Behavioural brain research*, *373*, 112110. doi:10.1016/j.bbr.2019.112110

Maaß, S. C., Schlichting, N., & van Rijn, H. (2019b). Eliciting contextual temporal calibration: The effect of bottom-up and top-down information in reproduction tasks. *Acta psychologica*, *199*, 102898. doi:10.1016/j.actpsy.2019.102898

Maaß, S.C., van Maanen, L., & van Rijn, H. (2020, April 15). Conceptually Plausible Bayesian Inference in Interval Timing. doi:10.17605/OSF.IO/KQJXF. Retrieved from osf.io/kqjxf.

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

Martin, B., Wiener, M., & van Wassenhove, V. (2017). A Bayesian perspective on accumulation in the magnitude system. *Scientific reports*, *7*(1), 1-14. doi:10.1038/s41598-017-00680-0

Mestdagh, M., Verdonck, S., Meers, K., Loossens, T., & Tuerlinckx, F. (2019). Prepaid parameter estimation without likelihoods. *PLoS computational biology*, *15*(9), e1007181. doi:10.1371/journal.pcbi.1007181

Morey, R.D., & Rouder, J.N. (2018). BayesFactor: Computation of Bayes Factors for Common Designs. R package version 0.9.12-4.2. https://CRAN.R-project.org/package=BayesFactor

Myung, I. J. (2000). The importance of complexity in model selection. *Journal of mathematical psychology*, *44*(1), 190-204. doi:10.1006/jmps.1999.1283

Narain, D., Remington, E. D., De Zeeuw, C. I., & Jazayeri, M. (2018). A cerebellar mechanism for learning prior distributions of time intervals. *Nature communications*, *9*(1), 1-12. doi:10.1038/s41467-017-02516-x

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The computer journal*, *7*(4), 308-313. doi:10.1093/comjnl/7.4.308

Nichelli, P., Venneri, A., Molinari, M., Tavani, F., & Grafman, J. (1993). Precision and accuracy of subjective time estimation in different memory disorders. *Cognitive Brain Research*, *1*(2), 87-93. doi:10.1016/0926-6410(93)90014-V

Oprisan, S. A., & Buhusi, C. V. (2014). What is all the noise about in interval timing?. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1637), 20120459. doi:10.1098/rstb.2012.0459

Paraskevoudi, N., Balcı, F., & Vatakis, A. (2018). "Walking" through the sensory, cognitive, and temporal degradations of healthy aging. *Annals of the New York Academy of Sciences*, *1426*(1), 72-92. doi:10.1111/nyas.13734

Petersen, R. C., Doody, R., Kurz, A., Mohs, R. C., Morris, J. C., Rabins, P. V., ... & Winblad, B. (2001). Current concepts in mild cognitive impairment. Archives of neurology, 58(12), 1985-1992. doi:10.1001/archneur.58.12.1985

Petzschner, F. H., Glasauer, S., & Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends in cognitive sciences*, *19*(5), 285-293. doi:10.1016/j.tics.2015.03.002

Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in cognitive sciences*, *6*(10), 421-425. doi:10.1016/S1364-6613(02)01964-2

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 Sciences*, *114*(2), 412-417. doi:10.1073/pnas.1610706114

Rueda, A. D., & Schmitter-Edgecombe, M. (2009). Time estimation abilities in mild cognitive impairment and Alzheimer's disease. *Neuropsychology*, *23*(2), 178. doi:10.1037/a0014289

Simen, P., Balci, F., Desouza, L., Cohen, J. D., & Holmes, P. (2011). Interval timing by long-range temporal integration. *Frontiers in integrative neuroscience*, *5*, 28. doi:10.3389/fnint.2011.00028

Shi, Z., Church, R. M., & Meck, W. H. (2013). Bayesian optimization of time perception. *Trends in cognitive sciences*, *17*(11), 556-564. doi:10.1016/j.tics.2013.09.009

Sohn, H., Narain, D., Meirhaeghe, N., & Jazayeri, M. (2019). Bayesian computation through cortical latent dynamics. *Neuron*, *103*(5), 934-947. doi:10.1016/j.neuron.2019.06.012

Taatgen, N. A., & Anderson, J. R. (2008). Constraints in cognitive architectures. *Cambridge handbook of computational psychology*, 170-185. doi:10.1017/CBO9780511816772.009

Taatgen, N., & van Rijn, H. (2011). Traces of times past: representations of temporal intervals in memory. *Memory & cognition*, *39*(8), 1546-1560. doi:10.3758/s13421-011-0113-0

Turgeon, M., & Wing, A. M. (2012). Late onset of age-related difference in unpaced tapping with no age-related difference in phase-shift error detection and correction. *Psychology and aging*, *27*(4), 1152. doi:10.1037/a0029925

van Doorn, J., van den Bergh, D., Bohm, U., Dablander, F., Derks, K., Draws, T., ... & Ly, A. (2019). The JASP guidelines for conducting and reporting a Bayesian analysis. doi:10.31234/osf.io/yqxfr van Rijn, H. (2016). Accounting for memory mechanisms in interval timing: a review. *Current Opinion in Behavioral Sciences*, *8*, 245-249. doi:10.1016/j.cobeha.2016.02.016

van Rijn, H., Borst, J., Taatgen, N., & van Maanen, L. (2016). On the necessity of integrating multiple levels of abstraction in a single computational framework. *Current Opinion in Behavioral Sciences*, *11*, 116-120. doi:10.1016/j.cobeha.2016.07.007

Van Rijn, H., Gu, B. M., & Meck, W. H. (2014). Dedicated clock/timing-circuit theories of time perception and timed performance. In *Neurobiology of interval timing* (pp. 75-99). Springer, New York, NY. doi:10.1007/978-1-4939-1782-2_5

Wearden, J. H., Wearden, A. J., & Rabbitt, P. M. (1997). Age and IQ effects on stimulus and response timing. Journal of Experimental Psychology: Human Perception and Performance, 23(4), 962. doi:10.1037/0096-1523.23.4.962

Wiener, M., Michaelis, K., & Thompson, J. C. (2016). Functional correlates of likelihood and prior representations in a virtual distance task. *Human brain mapping*, *37*(9), 3172-3187. doi:10.1002/hbm.23232

Wild-Wall, N., Willemssen, R., & Falkenstein, M. (2009). Feedback-related processes during a time-production task in young and older adults. *Clinical Neurophysiology*, *120*(2), 407-413. doi: 10.1016/j.clinph.2008.11.007

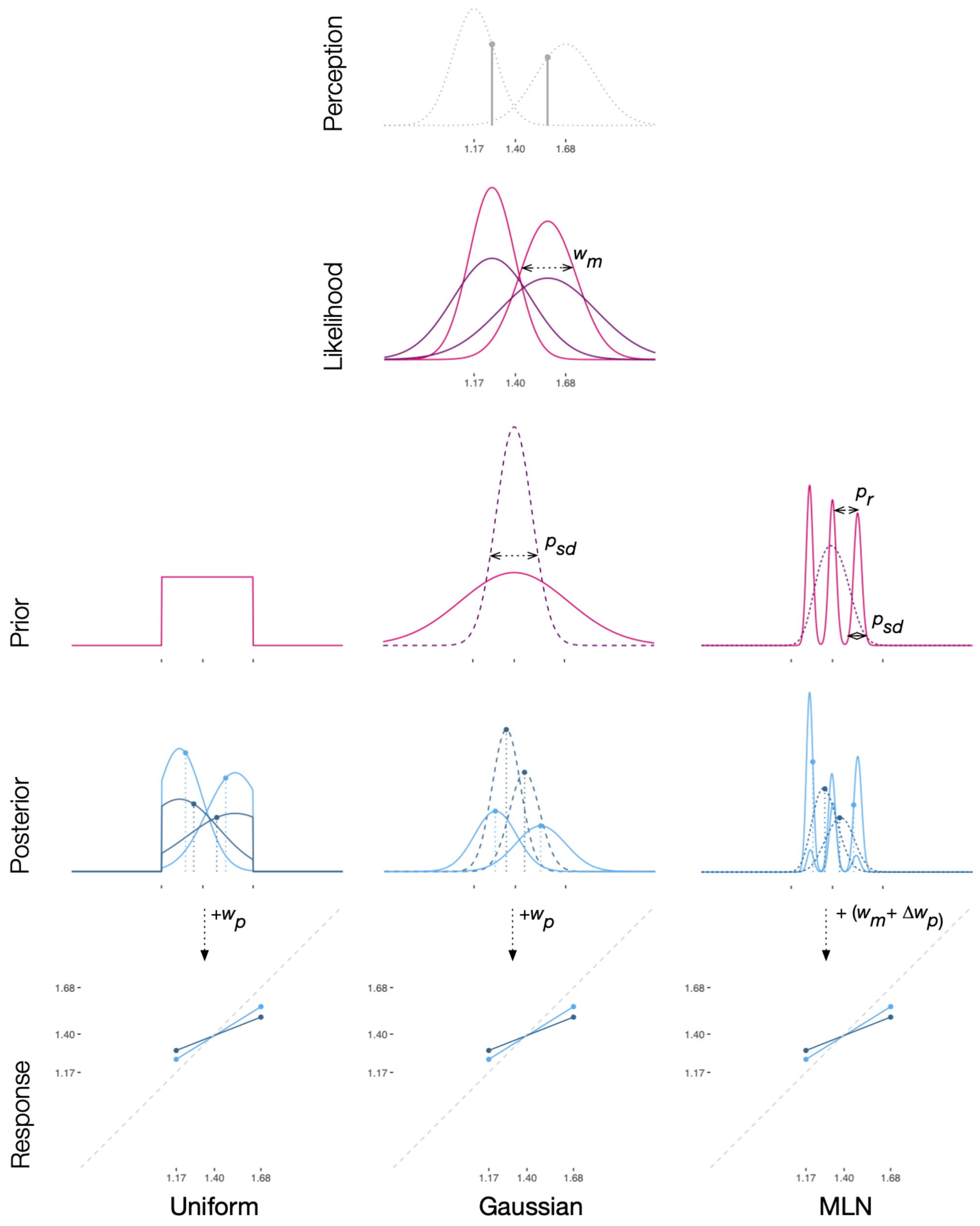

Perception
1.17
1.40
1.68
Likelihood
w_m
1.17
1.40
1.68
Prior
p_{sd}
p_r
p_{sd}
Posterior
$+w_p$
$+w_p$
$+ (w_m + \Delta w_p)$
Response
1.68
1.40
1.17
1.17
1.40
1.68
Uniform
Gaussian
MLN

Figure 1

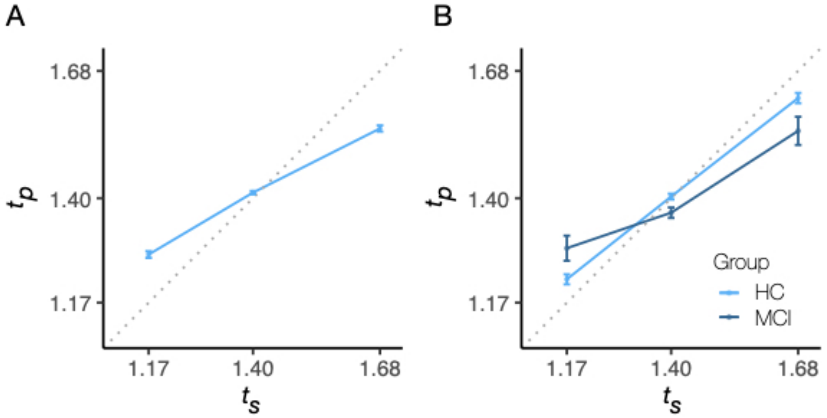


Figure 2

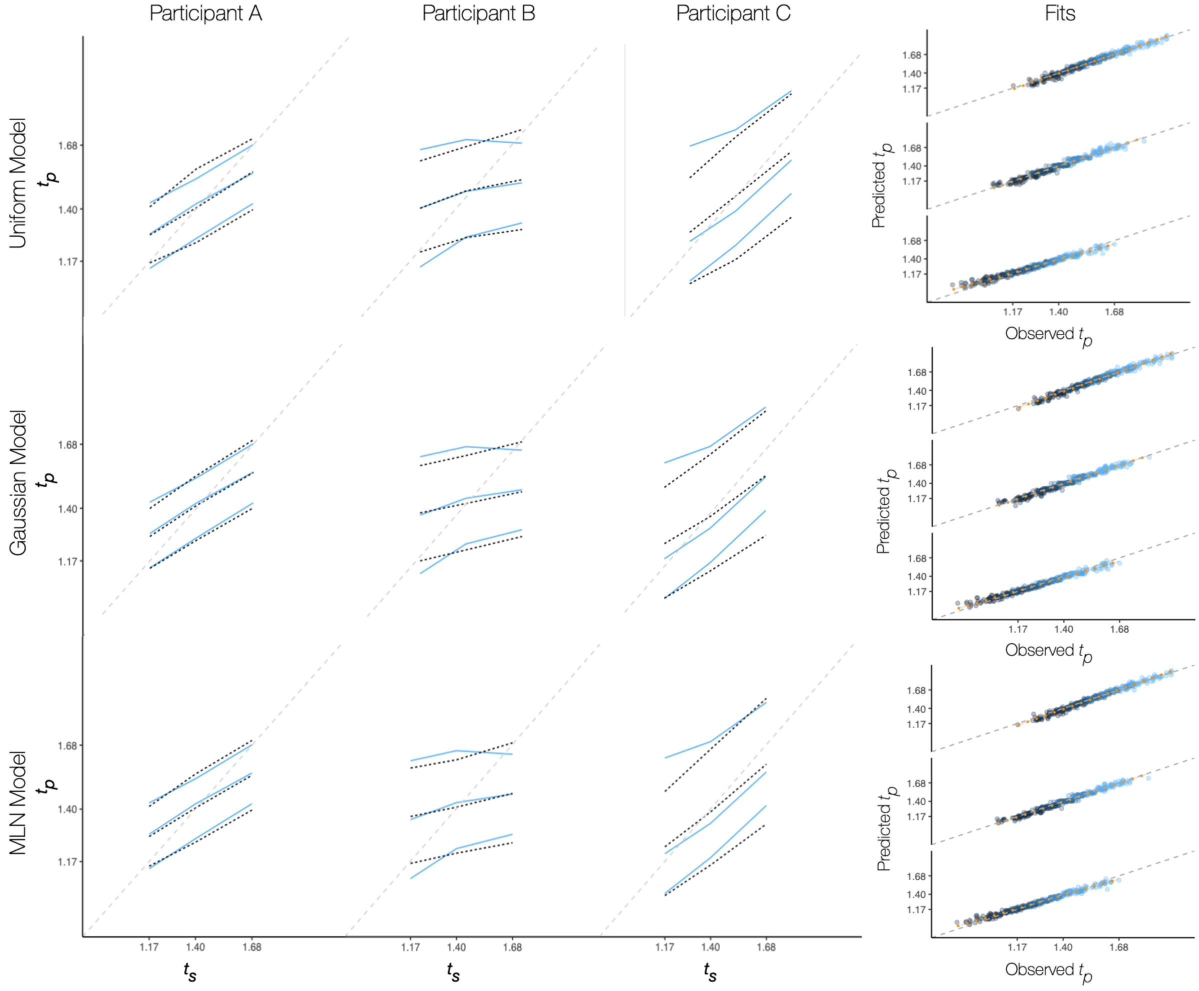


Figure 3

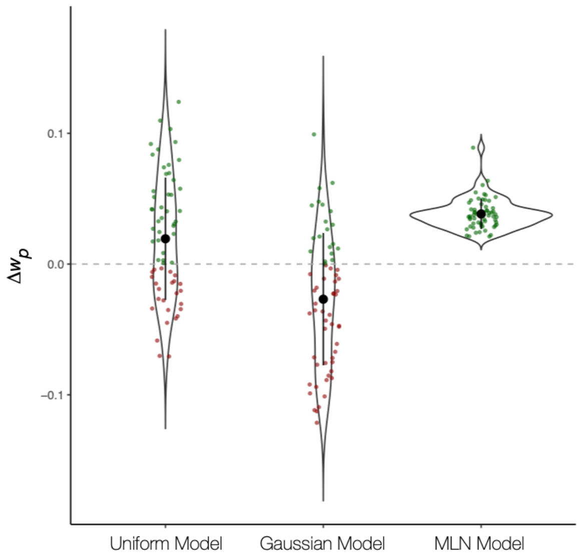
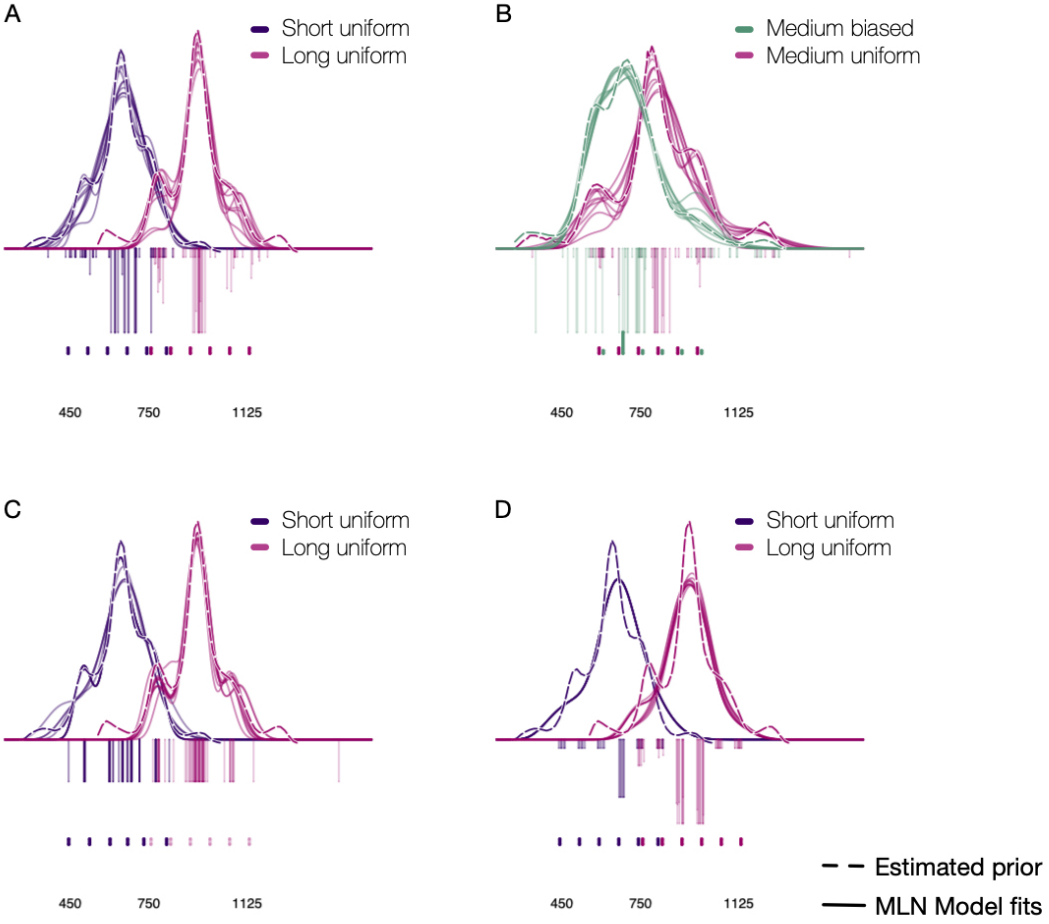


Figure 4



Fi

Figure 5

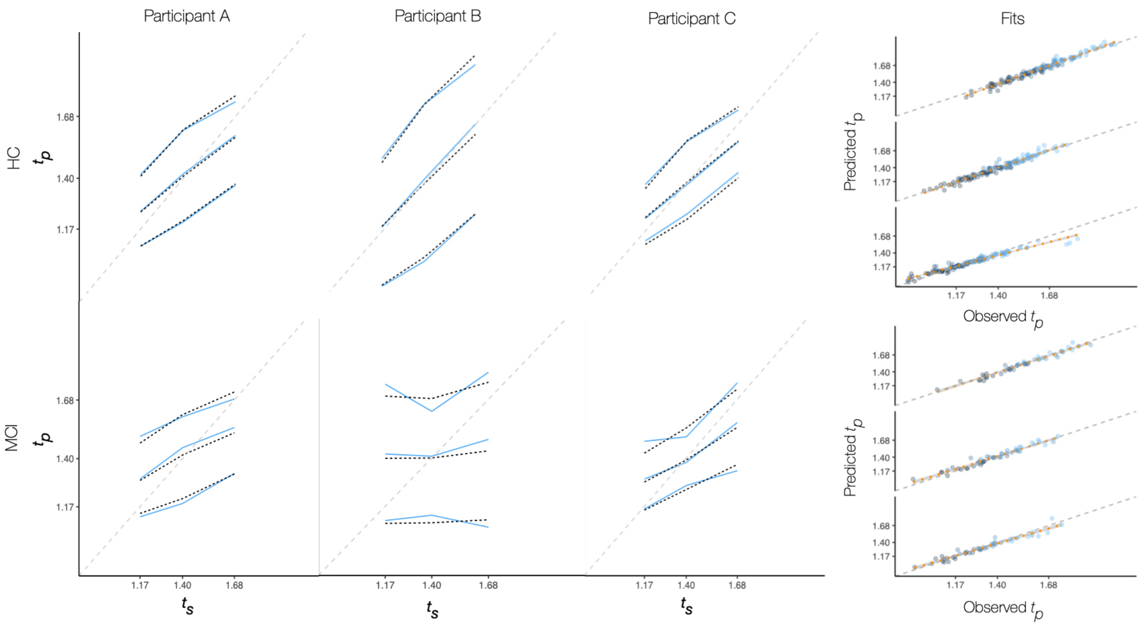


Figure 6

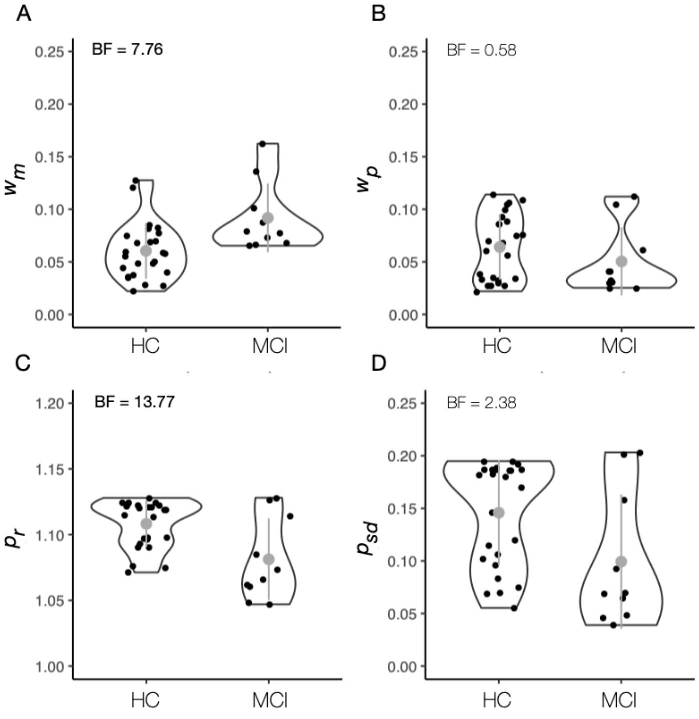


Figure 7

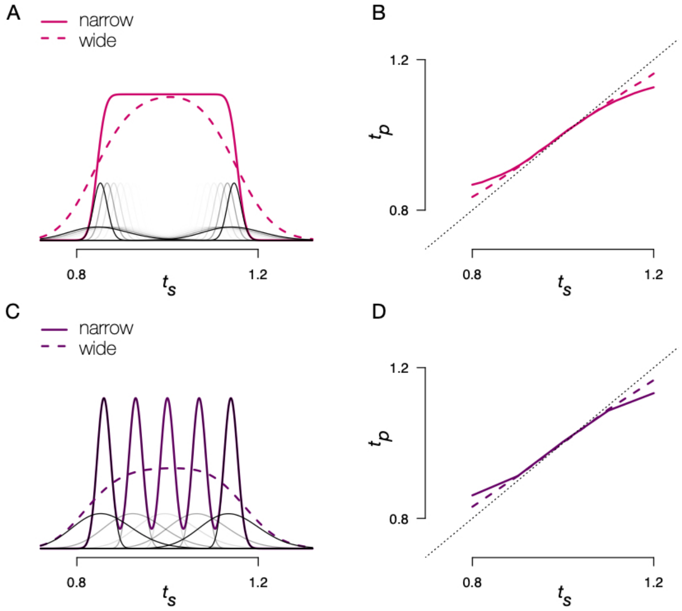

A
narrow
wide
0.8
1.2
t_s
B
1.2
0.8
t_p
0.8
1.2
t_s
C
narrow
wide
0.8
1.2
t_s
D
1.2
0.8
t_p
0.8
1.2
t_s

Figure 8