Ensemble perception in the time domain: evidence in favor of logarithmic encoding of time intervals

Yue Ren[1], Hermann J. Müller[1], Zhuanghua Shi[1*]

[1] General and Experimental Psychology, Psychology Department, LMU Munich, Munich, Germany

* correspondent author, email: strongway@psy.lmu.de

Author Note

This work was supported by the German Research Foundation (DFG) grants SH166/3-2, awarded to ZS.
Correspondence concerning this article should be addressed to Zhuanghua Shi, General and Experimental Psychology, Department of Psychology, LMU Munich, 80802, Munich, Germany.
Email: strongway@psy.lmu.de.

Abstract

Although time perception is based on the internal representation of time, whether the subjective timeline is scaled linearly or logarithmically remains an open issue. Evidence from previous research is mixed: while the classical internal-clock model assumes a linear scale with scalar variability, there is evidence that logarithmic timing provides a better fit to behavioral data. A major challenge for investigating the nature of the internal scale is that the retrieval process required for time judgments may involve a remapping of the subjective time back to the objective scale, complicating any direct interpretation of behavioral findings. Here, we used a novel approach, requiring rapid intuitive, 'ensemble' averaging of a whole set of time intervals, to probe the subjective timeline. Specifically, observers' task was to average a series of successively presented, auditory or visual, intervals in the time range 300–1300 ms. Importantly, the intervals were taken from three sets of durations, which were distributed such that the arithmetic mean (from the linear scale) and the geometric mean (from the logarithmic scale) were clearly distinguishable. Consistently across the three sets and the two presentation modalities, our results revealed subjective averaging to be close to the geometric mean, indicative of a logarithmic timeline underlying time perception.

Introduction

What is the mental scale of time? This is one of the fundamental issues in timing research that has long been posed, yet remains only poorly understood. The well-known internal-clock model implicitly assumes linear coding of time: a central pacemaker generates ticks and an accumulator collects those ticks in a process of linear summation [1,2]. However, the neuronal plausibility of such a coding scheme has been called into doubt: large time intervals would require an accumulator with (near-) unlimited capacity [3], rendering it very costly to implement neuronally [4,5]. Given this, alternative timing models have been proposed that use oscillatory patterns or neuronal trajectories to encode temporal information [6–9]. For example, the striatal beat-frequency model [6,9,10] assumes that time intervals are encoded in the oscillatory firing patterns of cortical neurons, with the length of an interval being discernible, for time judgements, by the similarity of an oscillatory pattern with patterns stored in memory. Neuronal trajectory models, on the other hand, use intrinsic neuronal patterns as markers for timing. However, owing to the 'arbitrary' nature of neuronal patterns, encoded intervals cannot easily be used for simple arithmetic computations, such as the summation or subtraction of two intervals. Accordingly, these models have been criticized for their lack of computational accessibility [11]. Recently, a neural integration model [12–14] adopted stochastic drift diffusion as the temporal integrator which, similar to the classic internal-clock model, starts the accumulation at the onset of an interval and increases until the integrator reaches a decision threshold. To avoid the 'unlimited-capacity' problem encountered by the internal-clock model (see above), the neural integration model assumes the ramping activities reach a fixed decision barrier, though with different drift rates – in particular, a lower drift rate for longer intervals. However, this proposal encounters a conceptual problem: the length of the interval would need to be known at the start of the accumulation. Thus, while a variety of timing models have been proposed, there is no agreement on how time intervals are actually encoded.

There have been many attempts, using a variety of psychophysical approaches, to directly reveal the subjective timeline that underlies our time judgments. However, it turned out that

distinguishing between linear and logarithmic timing is largely constrained by the experimental paradigms adopted [5,15–17]. In temporal bisection, for example, the bisection point was often observed at the geometric mean [18,19], which led to the earliest speculation that the subjective timeline might be logarithmic in nature: If time were coded linearly, the midpoint on the subjective scale should be equidistant from both (the lower and upper) time references, yielding their arithmetic mean. By contrast, given logarithmic coding of time, the midpoint equidistant to both time references on the logarithmic scale would turn out to be their geometric mean, as observed. However, Gibbon and colleagues offered an alternative account for the geometric mean obtained with the bisection task, namely: the midpoint is determined by computing the ratio of the elapsed time T to the Short and Long reference durations (i.e., a comparison between Short/T and T/Long), which also yields the geometric mean [20,21]. Using a modified temporal-bisection procedure and signal-detection theory (SDT), Yi found that logarithmic timing with fixed variability (i.e., timing accuracy is constant in log-space) provided a better fit to the data than a linear representation of time with scalar variability (i.e., accuracy is relative to the interval being timed) [22]. However, Dehaene showed that "linear coding with scalar variability and logarithmic coding with fixed variability lead to the same metric of number similarity, and therefore to the same behavior" (p. 244) [23]. Thus, to date, it remains unsolved whether interval timing is based on a logarithmic or a linear internal representation.

One main factor responsible for the inconsistent findings is that the observer's response to a time interval may not directly reflect the internal representation. For example, an external interval might be encoded and stored (in memory) in a compressed, logarithmic format internally. When that interval is retrieved, it may first have to be decoded (i.e., transformed from logarithmic to linear space) in working memory before any further comparisons. The involvement of decoding processes would complicate drawing direct inferences from empirical data. However, it may be possible to escape such complications by examining basic 'intuitions' of interval timing, which may bypass complex decoding processes. One fundamental perceptual intuition we use all the time is 'ensemble perception' – which refers to the idea that our sensory systems can rapidly extract statistical (summary) properties from a set of similar items, such as

their sum or mean magnitude. For example, Dehaene and colleagues [24] used an individual number-space mapping task to compare Mundurucu, an Amazonian indigenous culture with a reduced number lexicon, to US American educated participants. They found that the Mundurucu group, across all ages, mapped symbolic and nonsymbolic numbers onto a logarithmic scale, whereas educated western adults used linear mapping of numbers onto space – favoring the proposal that the initial intuition of number is logarithmic [24]. And they surmised that the educated Western participants had acquired sophisticated mapping knowledge (i.e., decoding the internal representation) to overwrite the basic intuitive mapping. Moreover, kindergarten and pre-school children also exhibit a non-linear representation of numbers close to logarithmic compression (e.g., they place the number 10 near the midpoint of the 1–100 scale), and this nonlinearity becomes 'linearized' after some 3 to 4 years of schooling [25–27].

Our perceptual intuition works very fast. For example, we quickly form an idea about the average size of apples from just taking a glimpse at the apple tree. In a seminal study by Ariel [28], participants, when asked to identify whether a presented object belonged to a group of similar items, tended to automatically respond with the mean size. Intuitive averaging has been demonstrated for various features in the visual domain [29], from primary ensembles such as object size [30,31], color and grayness [32,33], to high-level ensembles such as facial expression and lifelikeness [34–36]. Rather than being confined to the (inherently 'parallel') visual domain, ensemble perception has also been demonstrated for sequentially presented items, such as auditory frequency, tone loudness, tone duration, and weight [37–40]. In a cross-modal temporal integration study, Chen and colleagues [41] showed that the average interval of a train of auditory intervals can quickly capture a subsequently presented visual interval, influencing visual motion perception.

In brief, our perceptual systems can automatically extract overall statistical properties using very basic intuitions to cope with sensory information overload and the limited capacity of working memory. Thus, given that ensemble perception operates at a fast and low-level stage of processing (possibly bypassing many high-level cognitive decoding processes), using ensemble

perception as a tool to test time perception may provide us with new insights into the internal representation of time intervals.

On this background, we designed an interval-duration averaging task in which observers were asked to compare the average duration of a set of intervals to a standard interval. We hypothesized that if the underlying interval representation is linear, the intuitive average should reflect the arithmetic mean (AM) of the sample intervals. Conversely, if intervals are logarithmically encoded internally and intuitive averaging operates on that level (i.e., without remapping individual intervals from logarithmic to linear scale), we would expect the readout of the intuitive average at the intervals' geometric mean (GM)[1]. Note, though, that the subjective averaged duration may be subject to a general bias, as has often been observed in studies of time estimation [42]. For this reason, we considered it wiser to compare response patterns for multiple sets of intervals to the patterns predicted respectively from the AM and the GM, rather than comparing the subjective averaged duration directly to either the AM or the GM of the intervals. Accordingly, we carefully chose three sets of intervals, for which one set would yield a different average to the other sets according to each individual account (see *Figure 1*). Each set contained five intervals – Set 1: 300, 550, 800, 1050, 1300 ms; Set 2: 600, 700, 800, 900, 1000 ms; and Set 3: 500, 610, 730, 840, 950 ms. Accordingly, Sets 1 and 2 have the same arithmetic mean (800 ms), which is larger than the arithmetic mean of Set 3 (727 ms). And Sets 1 and 3 have the same geometric mean (710 ms), which is shorter than the geometric mean of Set 2 (787 ms). The rationale was that, given that the assumptions of linear and logarithmic representations make distinct predictions for the three sets, we may be able to infer the internal representation by observing the behavioral outcome based on the predictions.

[1] Given a set of time intervals $\{X_1,X_2,\ldots,X_n\}$ which are internally encoded as $\{\log(X_1), \log(X_2), \ldots ,\log(X_n)\}$, an intuitive average of the log-encoded intervals is $\mathrm{Mean}(\log(X))$, and the readout of this average value is $\exp(\mathrm{Mean}(\log(X)))$, which is the geometric mean.

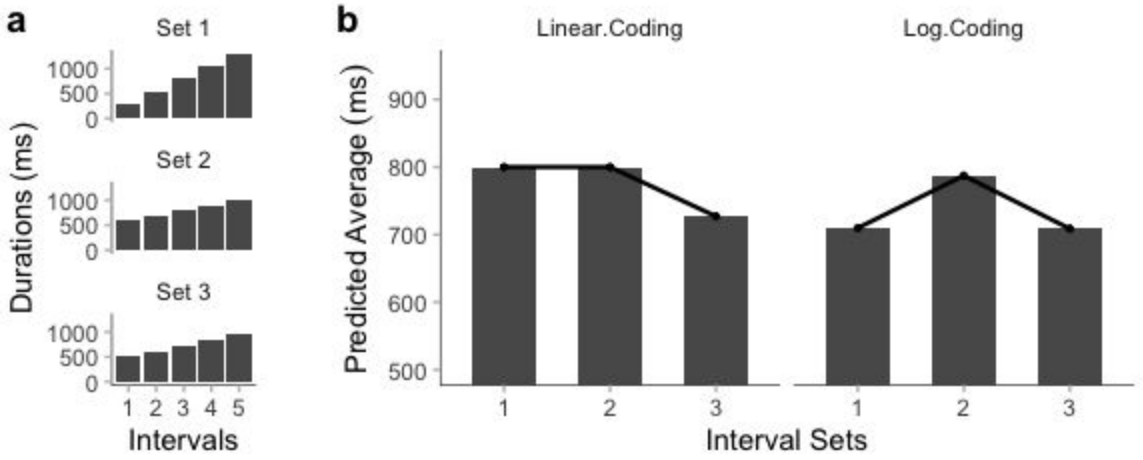

Figure 1. Illustration of three sets of intervals used in the study. (a) Three sets of intervals each of five intervals (Set 1: 300, 550, 800, 1050, 1300 ms; Set 2: 600, 700, 800, 900, 1000 ms; Set 3: 500, 610, 730, 840, 950 ms). The presentation order of the five intervals was randomized within each trial. (b) Predictions of ensemble averaging based on two hypothesized coding schemes: Linear Coding and, respectively, Logarithmic Coding. Sets 1 and 2 have the same arithmetic mean of 800 ms, which is larger than the arithmetic mean of the group 3 (727 ms). Sets 1 and 3 have the same geometric mean of 710 ms, which is smaller than the geometric mean of set 1 (787 ms).

Subjective durations are known to differ between visual and auditory signals [5,43,44], as our auditory system has higher temporal precision than the visual system. Accordingly, the processing strategies may potentially differ between the two modalities. Thus, in order to establish whether the internal representation of time is modality-independent, we tested both modalities using the same interval sets in separate experiments.

Methods

Ethics statement

The methods and experimental protocols were approved by the Ethics Board of the Faculty of Pedagogics and Psychology at LMU Munich, Germany, and are in accordance with the Declaration of Helsinki 2008.

Participants

A total of 32 participants (16 females, mean age of 25.7) were recruited for the study, 16 randomly allocated to each experiment, from the LMU Psychology community. Prior to the experiments, participants gave written informed consent and were paid for their participation of 8 Euros per hour. All reported a normal (or corrected-to-normal) vision, normal hearing, and no somatosensory disorders.

Stimuli

The experiments were conducted in a sound-isolated cabin, with dim incandescent background lighting. Participants sat approximately 60 cm from a display screen, a 21-inch CRT monitor (refresh rate 100 Hz; screen resolution 800 x 600 pixels). In both experiments, visual stimuli – all gray disks (luminance 21.4 cd/m^2) – were presented in the monitor center against a black screen background (1.6 cd/m^2). In Experiment 1, auditory stimuli (i.e., intervals) were delivered via two loudspeakers positioned just below the monitor, with a left-to-right separation of 40 cm. Brief auditory beeps (10 ms, 60 dB; frequency of 2500 or 3000 Hz, respectively) were presented to mark the beginning and end of the auditory intervals. In Experiment 2, the intervals were marked visually by brief (10-ms) flashes in the display center, of a gray disk (5° of visual angle in diameter) presented at the beginning and end (or between successive) interval(s).

As for the length of the (five) successively presented intervals on a given trial, there were three sets: Set 1: 300, 550, 800, 1050, 1300 ms; Set 2: 600, 700, 800, 900, 1000 ms; and Set 3: 500, 610, 730, 840, 950 ms. These sets were constructed such that Sets 1 and 2 had the same arithmetic mean (800 ms), which is larger than the arithmetic mean of Set 3 (727 ms). And Sets 1 and 3 have the same geometric mean (710 ms), which is shorter than the geometric mean of Set 2 (787 ms). Of note, the order of the five intervals (of the presented set) was randomized on each trial.

Procedure

Two separate experiments were conducted, testing auditory (Experiment 1) and visual stimuli (Experiment 2), respectively. Each trial consisted of two presentation phases: successive

presentation of five intervals, followed by the presentation of a single comparison interval. Participants' task was to indicate, via a keypress response, whether the comparison interval was shorter or longer than the average of the five successive intervals. The response could be given without stress on speed.

In Experiment 1 (auditory intervals), trials started with a fixation cross presented for 500 ms, followed by a succession of five intervals demarcated by six 10-ms auditory beeps. Along with the onset of the auditory stimuli, a '1' was presented on display monitor, telling participants that this was the first phase of the comparison task. The series of intervals was followed by a blank gap (randomly ranging between 800–1200 ms), with a fixation sign '+' on the screen (indicating the transition to comparison phase 2). After the gap, a single comparison duration demarcated by two brief beeps (10 ms) was presented, together with a '2', indicating phase two of the comparison. Following another random blank gap (of 800–1200 ms), a question mark ('?') appeared in the center of the screen, prompting participants to report whether the average interval of the first five (successive) intervals was longer or shorter than the second, comparison interval (Figure 2A). Participants issued their response via the left or right arrow keys (on the keyboard in front of them) using their two index fingers, corresponding to either 'shorter' or 'longer' judgments. To make the two parts 1 and 2 of the interval presentation clearly distinguishable, two different frequencies (2500 and 3000 Hz) were randomly assigned to the first and, respectively, the second set of auditory interval markers.

Experiment 2 (visual intervals) was essentially the same as Experiment 1, except that the intervals were delivered via the visual modality and were demarcated by brief (10-ms) flashes of gray disks in the screen center (see Figure 2B). Also, the visual cue signals used to indicate the two interval presentation phases ('1', '2') in the 'auditory' Experiment 1 were omitted, to ensure participants' undivided attention to the judgment-relevant intervals.

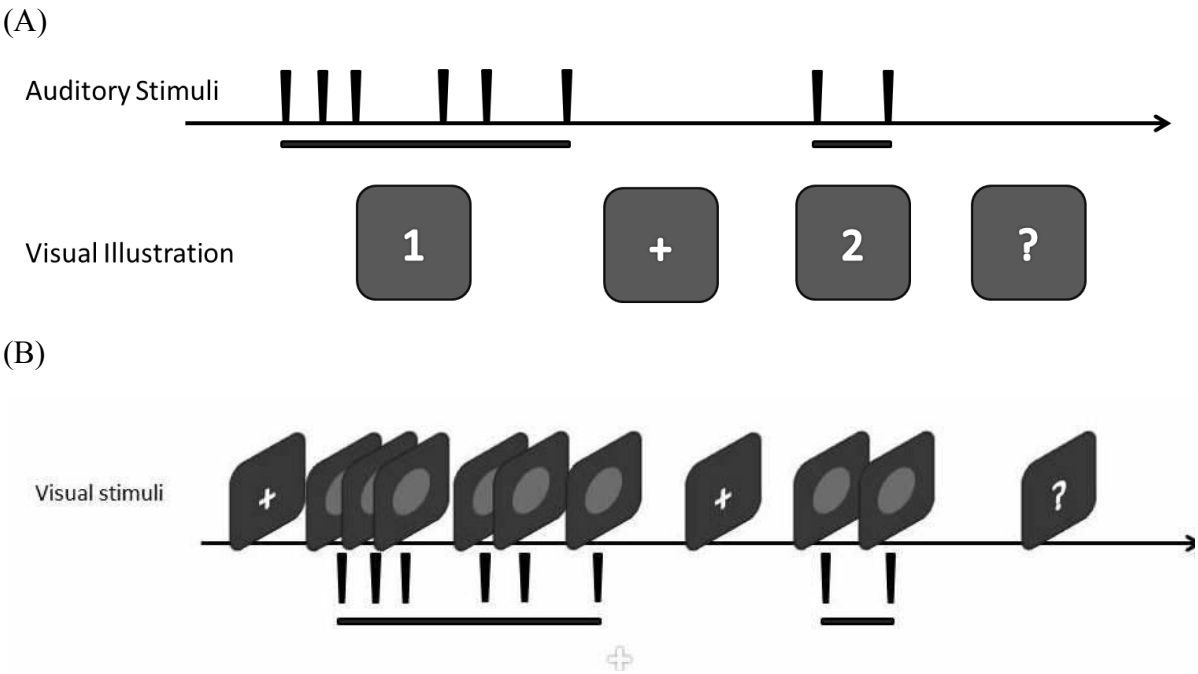

Figure 2. Schematic illustration of a trial in Experiments 1 and 2. (A) In Experiment 1, the first set of stimuli was an auditory sequence of five intervals demarcated by six short (10-ms) auditory beeps of a particular frequency (either 2500 or 3000 Hz). The second, comparison interval was demarcated by two beeps of a different frequency (either 3000 or 2500 Hz). (B) The trial structure was essentially the same in Experiment 2, except that the intervals were marked by briefly presented grey disks in the monitor center (and, given the task required a visual comparison, the two interval presentation phases were not further signaled by the digits '1' and '2'). In both experiments, participants were asked to compare the average of the first set of intervals with the second interval and decide which one was longer.

In order to obtain, in an efficient manner, reliable estimates of both the point of subjective equality (PSE) and the just noticeable difference (JND) of the psychometric function of the interval comparison, we employed the updated maximum-likelihood adaptive procedure from the UML toolbox for Matlab [45]. The comparison interval was set to 500 ms initially, and the UML adaptive procedure then determined the next comparison interval based on the participant's response. Moreover, to mitigate habituation and expectation effects, the sequences of comparison intervals for three different sets were presented randomly intermixed across trials, though with concurrent tracking of the three separate adaptive procedures.

Prior to the testing session, participants were given verbal instructions and then familiarized with the task in a practice block of 30 trials (10 comparison trials for each set). Of note, upon receiving the instruction, most participants spontaneously voiced concern about the difficulty of the judgment they were told to make. However, after performing just a few trials of the training block, they all expressed confidence that the task was easily doable after all, and they all went on to complete the experiment successfully. In the formal testing session, each of the three sets was tested for 80 times, yielding a total of 240 trials per each experiment. The whole experiment took approximately 60 minutes to complete.

Results

Figure 3 depicts the mean PSEs for three sets of intervals, separately for the auditory Experiment 1 and the visual Experiment 2. In both experiments, the estimated average from the three sets showed a similar pattern, with the mean of Set 2 being larger than the means of both Set 1 and Set 3. Repeated-measures ANOVAs conducted for both experiments separately revealed the Set (main) effect to be significant both for Experiment 1, $F(2,30) = 10.1, p < .001, \eta_g^2 = .065$, and Experiment 2, $F(2,30) = 7.31, p < .01, \eta_g^2 = .012$. Post-hoc Bonferroni-corrected comparisons confirmed the Set effect was mainly due to the mean being highest with Set 2. In more detail, for the auditory experiment (Figure 3a), the mean of Set 2 was larger than those of Set 1 (p=.01) and Set 3 (p<.001), while there was no significant difference between the latter (p=.42). The result pattern was the same for the visual experiment (Figure 3b), with Set 2 generating a larger mean than both Set 1 (p=.007) and Set 3 (p<.001), with no reliable difference between the latter (p=.08). These results are consistent with one of our predictions, namely, that the main averaging process for providing perceptual summary statistics is based on the geometric mean, in both the visual and auditory modalities.

In addition, for the three sets of auditory intervals (Experiment 1), the mean PSEs did not differ significantly from their correspondent geometric means (paired t-test pooled across the three sets), $t(47) = 1.71, p = .093$, but were significant smaller than the arithmetic means, $t(47) = 3.87, p < .001$. In Experiment 2, the mean PSEs for all three visual interval sets were

significantly smaller than both the geometric [$t(47) = 4.13, p < .001$] and arithmetic means [$t(47) = 5.62, p < .001$]. Additionally, the estimated mean durations were smaller for the visual (Experiment 2) compared to the auditory intervals (Experiment 1) , $F(1, 92) = 5.972, p = .0164$. The latter effect is consistent with previous reports that auditory intervals are often perceived as longer than physically equivalent visual intervals [43,46].

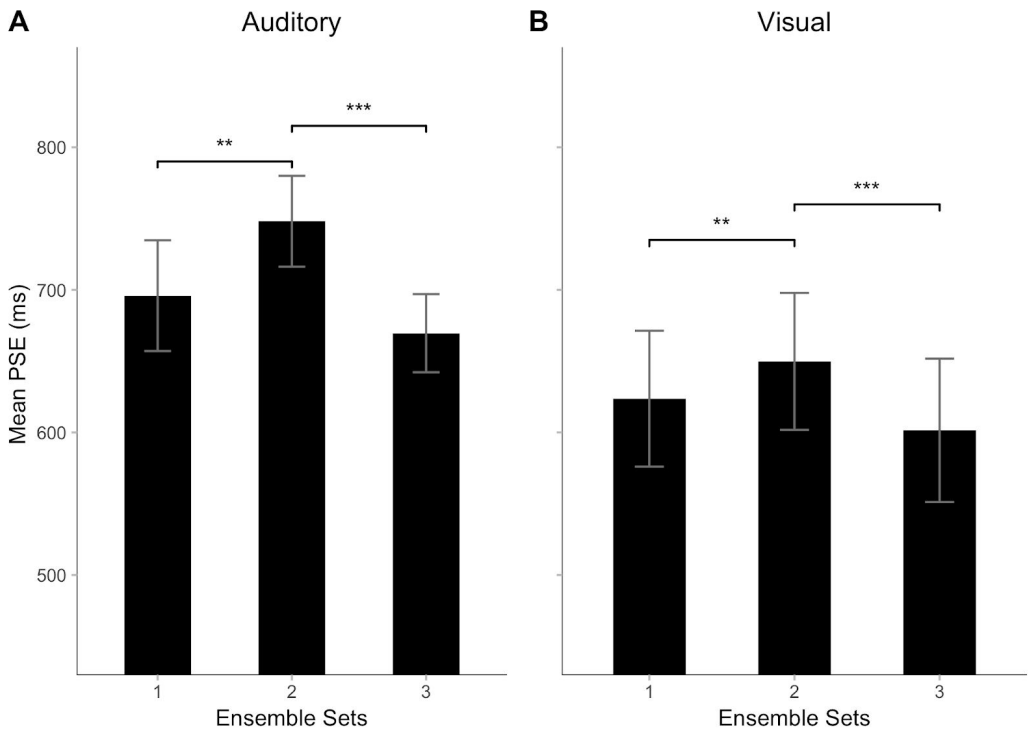

*Figure 3: Mean PSEs of the three tested sets of intervals, separately for Experiment 1 (A) and Experiment 2 (B). ** denotes p<.01, and *** p<.001.*

Another key parameter providing an indicator of an observer's temporal sensitivity (resolution) is given by the just noticeable difference (JND), defined as the interval difference between the 50%- and 75%-thresholds estimated from the psychometric function. Figure 4 depicts the JNDs obtained in Experiments 1 and 2, separately for the three sets of intervals. Repeated-measures ANOVAs, with Set as the main factor, failed to reveal any differences among the three sets, for either experiment [Experiment 1: $F(2, 30) = 0.78$, $p = .47$, $\eta_g^2 = .023$;

Experiment 2: $F(2, 30) = 0.296$, $p = .746$, $\eta_g^2 = .004$]. Comparison across Experiments 1 and 2, however, revealed the JNDs to be significantly smaller for auditory than for visual interval averaging, $F(1, 92) = 8.93$, $p < .01$; that is, temporal resolution turned out higher for the auditory than for the visual modality, consistent with the literature [47].

Thus, taken together, evaluation of both the mean and sensitivity of the participants' interval estimates demonstrated not only that ensemble coding in the temporal domain is accurate and consistent, but also that the geometric mean is used as the predominant averaging scheme for performing the task.

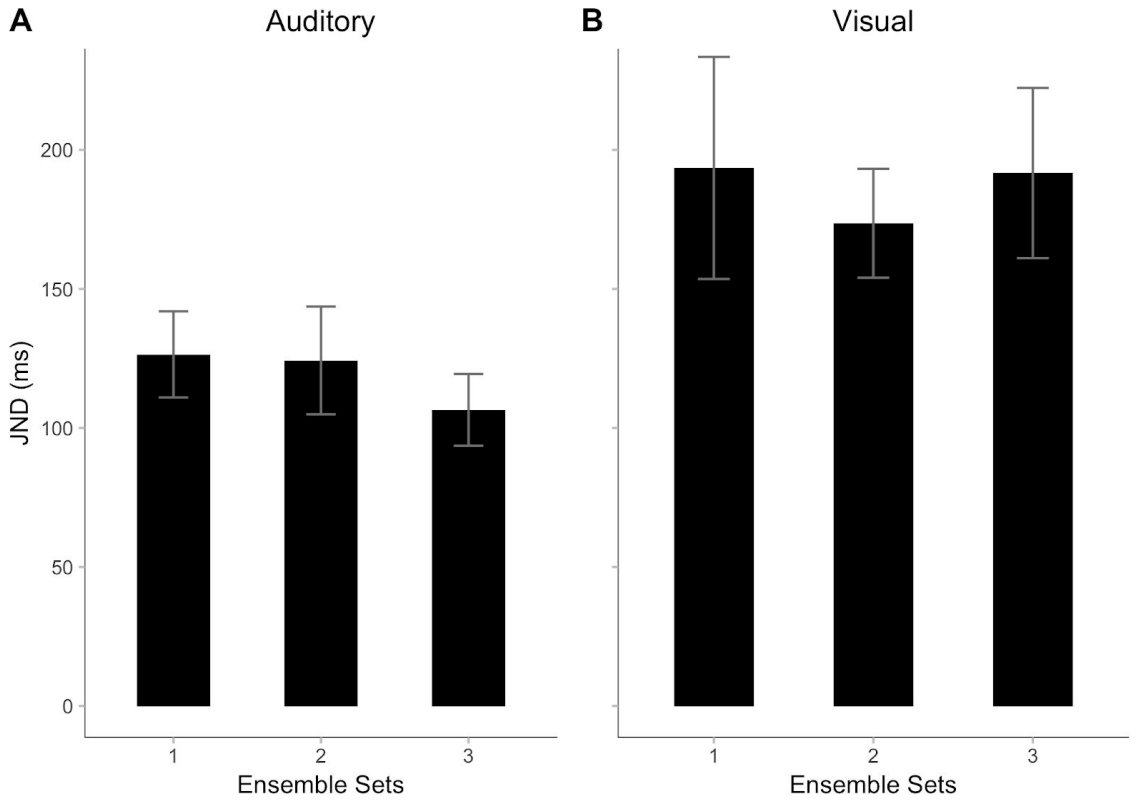

Figure 4: The mean JNDs of the three tested sets of intervals, separately for Experiment 1 (A) and Experiment 2 (B).

Discussion

The present study aimed to reveal the internal encoding of subjective time by examining ensemble averaging in the time domain. We hypothesized that ensemble statistical summary processes work at a very fast, low level of coding and might possibly bypass many high-level cognitive decoding strategies; in other words, ensemble averaging of time intervals may reflect the fundamental internal representation of time. If the internal representation of the subjective timeline is logarithmic, then basic intuitive averaging should be close to the geometric mean (see Footnote 1). By contrast, ensemble averaging of linearly encoded intervals should be close to the arithmetic mean. Proceeding from this hypothesis, we compared three sets of intervals, across which the geometric and arithmetic means exhibit different patterns (see *Figure 1*). The pattern we found was essentially the same with both auditory (Experiment 1) and visual intervals (Experiment 2); and, critically, it did match the pattern expected for the geometric – rather than that for the arithmetic – mean (see comparison between the predictions in *Figure 1* and the behavioral data in *Figure 3*). This supports our hypothesis that, regardless of the sensory modality, intuitive ensemble averaging of time intervals is based on logarithmically coded time. In other words, at least in the short range of intervals (from 300 to 1300 ms), the subjective timeline is likely logarithmically scaled.

Unlike estimating the mean of visual ensembles (e.g., mean object size or mean facial expression), using an averaging task in the temporal domain raises a pragmatic concern: how do we actually average time across time – or, in Wearden and Jones's [48] words: 'can people do this at all?' (p. 1295). In their study, participants were asked to average three consecutively presented durations and compare their mean to that of three subsequently presented durations; Wearden and Jones[48] found that not only could participants accurately extract the arithmetic mean, but the estimated means also remained indifferent to variations in the spacing of the sample durations. In the current study, by adopting the averaging task for multiple temporal intervals (>3), we resolved the problem encountered with the temporal bisection task, namely, that, as outlined in

the Introduction, the finding of the geometric (bisection) mean may be the outcome of a ratio comparison [20,21], rather than reflecting the internal subjective timeline. Specifically, we hypothesized that temporal ensemble perception may be indicative of a fast and intuitive process likely involving two stages: first, transformation, either linearly or nonlinearly, of the sample values onto a subjective scale [17,49,50] and storage in short-term (or working) memory (STM); second, estimation of the average of the multiple intervals on the subjective scale and then remapping from the subjective to the objective scale. One might surmise that the most efficient form of encoding would be linear, avoiding the need for nonlinear transformation. However, this is at variance with our finding that the averaging patterns observed for both visual and auditory intervals follow the pattern predicted by logarithmic encoding. The use of logarithmic encoding may be owing to the limited capacity of STM: uncompressed intervals require more space ('bits') to store, as compared to compressed intervals on a logarithmic scale. The brain appears to have chosen the latter for efficient STM storage in the first stage. However, nonlinear, logarithmic encoding in stage 1 could give rise to a computational cost for the averaging process in stage 2: averaging intervals on the objective, external scale would require the individual encoded intervals to be first transformed back from the subjective to the objective scale, which, due to being computationally expensive, would reduce processing speed. By contrast, arithmetic averaging on the subjective scale would be computationally efficient, as it requires only one step of remapping – of the subjective averaged interval onto the objective scale. Intuitive ensemble processing of time appears to have opted for the latter, ensuring computational efficiency. Thus, given the subjective scale is logarithmic, intuitive averaging would yield the geometric mean.

It could, of course, be argued that participants may adopt alternative weighting schemes to simple (equally weighted) arithmetic or geometric averaging. For example, the weight of an interval in the averaging process might be influenced by the length of that interval or/and the position of that interval in the sequence. Thus, for example, a long interval might engage more attention than a short interval, and weights are assigned to the intervals in accordance to their lengths. We simulated this alternative averaging strategy using an ideal observer model, which assumes that the observer encodes the presented intervals veridically (i.e., linearly), but assigns

weights according to the proportion of a given interval duration to the total duration. As illustrated in *Figure 5* (left panel), if this linear weighted averaging process (WA) were adopted as the dominant strategy for computing the mean interval, the estimated mean for Set 1 should be the largest one among the three sets. However, the observed data clearly are at variance with this model.

As has previously been demonstrated, more recent events in a temporal sequence tend to be more heavily weighted than earlier ones, which is known as recency effect [51]. Accordingly, the position of the intervals might influence the averaging on individual trials – though this position effect would be averaged out as a result of the random order of the intervals across trials. Nevertheless, we simulated two possible position-based averaging processes: the weighted moving average (WMA) and the exponential moving average (EMA). In the WMA, the weight values were calculated by constructing a linearly decreasing function (from recent to earlier events) relating the interval duration to its corresponding temporal order (as numeric values). To illustrate, in our data, we assigned 5/15, 4/15, …, 1/15 to the time order, representing the weights of the fifth to the first presented interval. The resulting weighted moving average (WMA) for the different interval sets is depicted in *Figure 5* (right panel). In EMA, we constructed a similar monotonic function by adding an exponential rate (α) in calculating the weights. That is, instead of progressing being linearly, the weight assigned to each interval decreases exponentially as the time order increases, but never reaches zero. *Figure 5* (middle panel) presents the EMA simulated results for the three interval sets (with $\alpha = 0.9$ as the initial weight). Comparison with Figure 3 shows that none of those weighting schemes is consistent with the observed behavioral pattern.

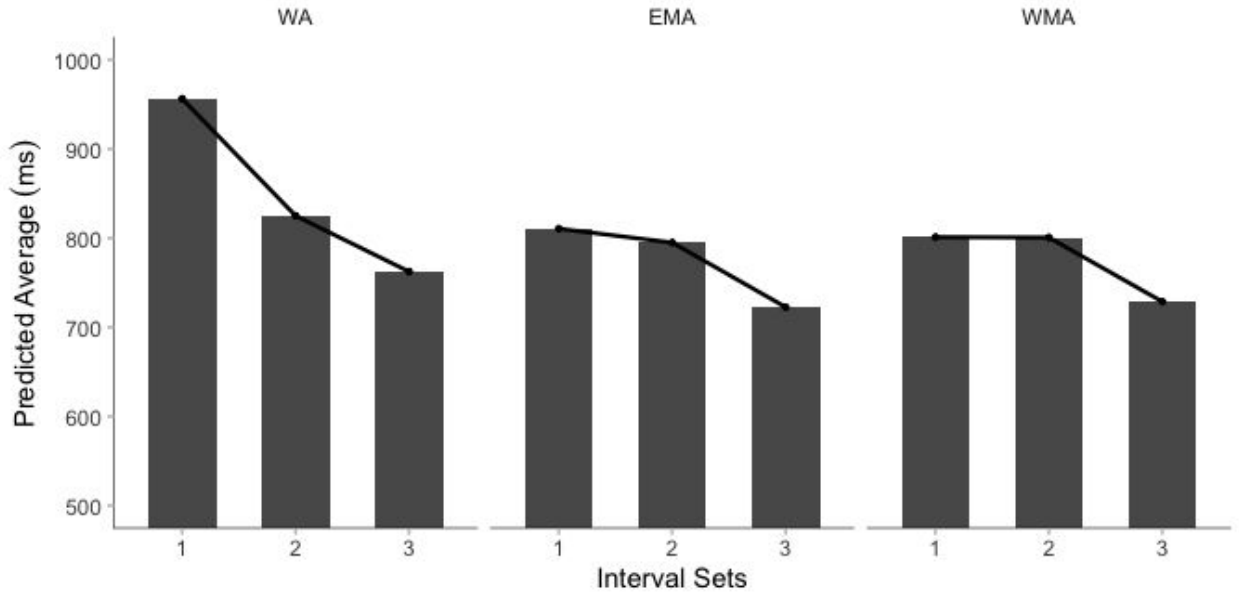

Figure 5: Predictions based on observers adopting three different strategies in the averaging task: Weighted Average (WA), Exponential Moving Average (EMA), and Weighted Moving Average (WMA), respectively.

In short, we examined three alternative averaging schemes (WA, EMA, and WMA) and found that none of them could explain the patterns we observed in Experiments 1 and 2. We are, thus, fairly confident that we can exclude those alternatives. Of course, there are many other, unusual weighting schemes one could think of. However, based on Occam's razor, our observed data patterns favor the simple geometric averaging account.

Logarithmic representation of stimulus intensity, such as of loudness or weight, has been proposed by Fechner over one and a half centuries ago [52], based on the fact that the JND measured at a given intensity is proportionate to the stimulus intensity (Weber's law). It has been shown that, for the same amount of information (quantized levels), the logarithmic scale provides the minimal expected relative error that optimizes communication efficiency, under the constraint that neural storage of sensory or magnitude information in general is capacity-limited [53]. In this respect, logarithmic timing provides a good solution for coping with a limited STM capacity to represent longer periods of time. However, as argued by Gallistel [54], logarithmic

encoding makes valid computations problematic: "Unless recourse is had to look-up tables, there is no way to implement addition and subtraction, because the addition and subtraction of logarithmic magnitudes corresponds to the multiplication and division of the quantities they refer to" (p. 8). It is likely that the ensuing computational complexity pushed intuitive ensemble averaging onto the internal, subjective scale – rather than the external, objective scale, which would have required multiple nonlinear transformations. Thus, our results join the increasing number of studies suggesting that, like other magnitudes [23,55], time is represented internally on a logarithmic scale, and intuitive averaging processes are likely bypassing higher-level cognitive computations. It should be noted that higher-level computations based on the external, objective scale can be acquired through training and education, and this is linked to mathematical ability [23,56,57]. Interestingly, a recent study also reported that, under dual-task conditions with an attention-demanding secondary task taxing visual working memory, the mapping of number onto space changed from linear to logarithmic [58] – providing convergent support for our hypothesis of an intuitive judgment strategy that can operate with a minimum of cognitive resources.

Another interesting finding of the present study concerns the overall underestimation of the (objective) mean interval durations, which was evident with all three sets of intervals and for both modalities (though it was particularly marked with visual intervals). The general underestimation is consistent with the subjective shortening effect – a source of bias reducing individual durations in memory [59,60]. The underestimation was more pronounced in the visual as compared to the auditory modality, which is consistent with the classic modality effect, of auditory events being judged as longer than visual events. This effect has been heavily investigated, the dominant idea being that temporal information is processed with higher resolution in the auditory than in the visual domain [43,61–63]. Thus, rather than examining whether the estimated mean was closer to the AM or GM for each ensemble set, we instead focused on the global estimation pattern across multiple sets and examined whether the patterns observed conformed with any of our hypotheses. With a consistent pattern found with all three sets (Figure 3) and underpinned by strong statistical power, it is reasonable to conclude that the obtained data

genuinely reflect an intuitive process of averaging multiple durations, where the average lies close to their geometric mean.

In summary, the present study provides behavioral evidence supporting a logarithmic representation of subjective time, and intuitive ensemble averaging is based on the geometric mean. Although the validity of behavioral studies has been increasingly acknowledged in recent years, lacking direct information about the neural basis makes it difficult to substantiate the inner timeline. We argue that achieving a full understanding of human timing requires a concerted research effort from both the psychophysical and neural perspectives.

REFERENCES

1. Gibbon, J. Scalar expectancy theory and Weber's law in animal timing. *Psychol. Rev.* **84**, 279–325 (1977).
2. Church, R. M. Properties of the Internal Clock. *Ann. N. Y. Acad. Sci.* 566–582 (1984).
3. Buhusi, C. V. & Meck, W. H. What makes us tick? Functional and neural mechanisms of interval timing. *Nat. Rev. Neurosci.* **6**, 755–765 (2005).
4. Eagleman, D. M. & Pariyadath, V. Is subjective duration a signature of coding efficiency? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **364**, 1841–1851 (2009).
5. Matthews, W. J. & Meck, W. H. Time perception: the bad news and the good. *Wiley Interdisciplinary Reviews: …* **5**, 429–446 (2014).
6. Matell, M. S. & Meck, W. H. Cortico-striatal circuits and interval timing: coincidence detection of oscillatory processes. *Cognitive Brain Research* **21**, 139–170 (2004).
7. Matell, M. S., Meck, W. H. & Nicolelis, M. a. L. Interval timing and the encoding of signal duration by ensembles of cortical and striatal neurons. *Behav. Neurosci.* **117**, 760–773 (2003).
8. Buonomano, D. V. & Karmarkar, U. R. How Do We Tell Time ? *Neuroscientist* 42–51 (2002).
9. Gu, B. M., van Rijn, H. & Meck, W. H. Oscillatory multiplexing of neural population codes for

interval timing and working memory. *Neurosci. Biobehav. Rev.* **48**, 160–185 (2015).

10. Oprisan, S. a. & Buhusi, C. V. Modeling pharmacological clock and memory patterns of interval timing in a striatal beat-frequency model with realistic, noisy neurons. *Front. Integr. Neurosci.* **5**, 52 (2011).
11. Wilkes, J. T. & Gallistel, C. R. Information Theory, Memory, Prediction, and Timing in Associative Learning. *Computational Models of Brain and Behavior* (2016).
12. Simen, P., Balci, F., de Souza, L., Cohen, J. D. & Holmes, P. A model of interval timing by neural integration. *J. Neurosci.* **31**, 9238–9253 (2011).
13. Balci, F. & Simen, P. Decision processes in temporal discrimination. *Acta Psychol.* **149**, 157–168 (2014).
14. Simen, P., Vlasov, K. & Papadakis, S. Scale (in)variance in a unified diffusion model of decision making and timing. *Psychol. Rev.* **123**, 151–181 (2016).
15. Staddon, J. E. & Higa, J. J. Time and memory: towards a pacemaker-free theory of interval timing. *J. Exp. Anal. Behav.* **71**, 215–251 (1999).
16. Gibbon, J. multiple time scales is well named. *J. Exp. Anal. Behav.* 272–275 (1999).
17. Johnson, K. O., Hsiao, S. S. & Yoshioka, T. Neural coding and the basic law of psychophysics. *Neuroscientist* **8**, 111–121 (2002).
18. Church, R. M. & Deluty, M. Z. Bisection of temporal intervals. *J. Exp. Psychol. Anim. Behav. Process.* **3**, 216–228 (1977).
19. Stubbs, D. A. Scaling of stimulus duration by pigeons. *J. Exp. Anal. Behav.* **26**, 15–25 (1976).
20. Allan, L. G. & Gibbon, J. Human bisection at the geometric mean. *Learn. Motiv.* **22**, 39–58 (1991).
21. Allan, L. G. The influence of the scalar timing model on human timing research. *Behav. Processes* **44**, 101–117 (1998).

22. Yi, L. Do rats represent time logarithmically or linearly? *Behav. Processes* **81**, 274–279 (2009).

23. Dehaene, S. Subtracting pigeons: logarithmic or linear? *Psychological science* vol. 12 244–6; discussion 247 (2001).

24. Dehaene, S., Izard, V., Spelke, E. & Pica, P. Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science* **320**, 1217–1220 (2008).

25. Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S. & Zorzi, M. Numerical estimation in preschoolers. *Dev. Psychol.* **46**, 545–551 (2010).

26. Booth, J. L. & Siegler, R. S. Developmental and individual differences in pure numerical estimation. *Dev. Psychol.* **42**, 189–201 (2006).

27. Barth, H. C. & Paladino, A. M. The development of numerical estimation: evidence against a representational shift. *Dev. Sci.* **14**, 125–135 (2011).

28. Ariely, D. Seeing sets: Representation by Statistical Properties. *Psychol. Sci.* **12**, 157–162 (2001).

29. Whitney, D. & Yamanashi Leib, A. Ensemble Perception. *Annu. Rev. Psychol.* **69**, 105–129 (2018).

30. Chong, S. C. & Treisman, A. Representation of statistical properties. *Vision Res.* **43**, 393–404 (2003).

31. Chong, S. C. & Treisman, A. Statistical processing: Computing the average size in perceptual groups. *Vision Res.* **45**, 891–900 (2005).

32. Haberman, J. & Whitney, D. Rapid extraction of mean emotion and gender from sets of faces. *Curr. Biol.* 751–753 (2007).

33. Haberman, J., Brady, T. F. & Alvarez, G. A. Individual differences in ensemble perception reveal multiple, independent levels of ensemble representation. *J. Exp. Psychol. Gen.* **144**, 432–446 (2015).

34. Haberman, J. & Whitney, D. Ensemble perception: Summarizing the scene and broadening the limits of visual processing. *From perception to consciousness: ...* 1–21 (2012).

35. Kramer, R. S. S., Ritchie, K. L. & Burton, A. M. Viewers extract the mean from images of the same person: A route to face learning. *J. Vis.* **15**, 1–9 (2015).

36. Leib, A. Y., Kosovicheva, A. & Whitney, D. Fast ensemble representations for abstract visual impressions. *Nat. Commun.* **7**, 13186 (2016).

37. Curtis, D. W. & Mullin, L. C. Judgments of average magnitude: Analyses in terms of the functional measurement and two-stage models. *Perception & Psychophysics* vol. 18 299–308 (1975).

38. Piazza, E. A., Sweeny, T. D., Wessel, D., Silver, M. A. & Whitney, D. Humans Use Summary Statistics to Perceive Auditory Sequences. *Psychological Science* vol. 24 1389–1397 (2013).

39. Anderson, N. H. Application of a weighted average model to a psychophysical averaging task. *Psychonomic Science* vol. 8 227–228 (1967).

40. Schweickert, R., Han, H. J., Yamaguchi, M. & Fortin, C. Estimating averages from distributions of tone durations. *Atten. Percept. Psychophys.* **76**, 605–620 (2014).

41. Chen, L., Zhou, X., Müller, H. J. & Shi, Z. What you see depends on what you hear: Temporal averaging and crossmodal integration. *J. Exp. Psychol. Gen.* **147**, 1851–1864 (2018).

42. Shi, Z., Ganzenmüller, S. & Müller, H. J. Reducing Bias in Auditory Duration Reproduction by Integrating the Reproduced Signal. *PLoS One* **8**, e62065 (2013).

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

44. Wearden, J. H. When do auditory/visual differences in duration judgments occur? *Q. J. Exp. Psychol.* **59**, 1709–1724 (2006).

45. Shen, Y., Dai, W. & Richards, V. M. A MATLAB toolbox for the efficient estimation of the psychometric function using the updated maximum-likelihood adaptive procedure. *Behav. Res.*

Methods **47**, 13–26 (2015).

46. Ganzenmüller, S., Shi, Z. & Müller, H. J. Duration reproduction with sensory feedback delay: differential involvement of perception and action time. *Front. Integr. Neurosci.* **6**, 1–11 (2012).
47. Shipley, T. AUDITORY FLUTTER-DRIVING OF VISUAL FLICKER. *Science* **145**, 1328–1330 (1964).
48. Wearden, J. H. & Jones, L. A. Is the growth of subjective time in humans a linear or nonlinear function of real time? *Q. J. Exp. Psychol.* **60**, 1289–1302 (2007).
49. Gibbon, J. The Structure Of Subjective Time: How Time Flies. *Psychology of Learning and Motivation* 105–135 (1986) doi:10.1016/s0079-7421(08)60017-1.
50. Taatgen, N. A., van Rijn, H. & Anderson, J. An integrated theory of prospective time interval estimation: the role of cognition, attention, and learning. *Psychol. Rev.* **114**, 577–598 (2007).
51. Hubert-Wallander, B. & Boynton, G. M. Not all summary statistics are made equal: Evidence from extracting summaries across time. *Journal of Vision* vol. 15 5 (2015).
52. Fechner, G. T. *Elemente der Psychophysik*. (1860).
53. Sun, J. Z., Wang, G. I., Goyal, V. K. & Varshney, L. R. A framework for Bayesian optimality of psychophysical laws. *Journal of Mathematical Psychology* vol. 56 495–501 (2012).
54. Gallistel, C. R. Mental Magnitudes. in *Space, time, and number in the brain: Searching for the foundations of mathematical thought* (eds. Dehaene, S. & Brannon, E. M.) 3–12 (Elsevier, 2011).
55. Roberts, W. A. Evidence that pigeons represent both time and number on a logarithmic scale. *Behavioural Processes* vol. 72 207–214 (2006).
56. Libertus, M. E., Feigenson, L. & Halberda, J. Preschool acuity of the approximate number system correlates with school math ability. *Dev. Sci.* **14**, 1292–1300 (2011).
57. Anobile, G. *et al.* Spontaneous perception of numerosity in pre-school children. *Proc. Biol. Sci.* **286**,

20191245 (2019).

58. Anobile, G., Cicchini, G. M. & Burr, D. C. Linear mapping of numbers onto space requires attention. *Cognition* **122**, 454–459 (2012).

59. Meck, W. H. Selective adjustment of the speed of internal clock and memory processes. *J. Exp. Psychol. Anim. Behav. Process.* **9**, 171–201 (1983).

60. Spetch, M. L. & Wilkie, D. M. Subjective shortening: A model of pigeons' memory for event duration. *Journal of Experimental Psychology: Animal Behavior Processes* vol. 9 14–30 (1983).

61. Gu, B. M., Cheng, R. K., Yin, B. & Meck, W. H. Quinpirole-induced sensitization to noisy/sparse periodic input: temporal synchronization as a component of obsessive-compulsive disorder. *Neuroscience* **179**, 143–150 (2011).

62. Shi, Z., Church, R. M. & Meck, W. H. Bayesian optimization of time perception. *Trends Cogn. Sci.* **17**, 556–564 (2013).

63. Penney, T. B., Gibbon, J. & Meck, W. H. Differential effects of auditory and visual signals on clock speed and temporal memory. *J. Exp. Psychol. Hum. Percept. Perform.* **26**, 1770–1787 (2000).