



The timing database: An open-access, live repository for interval timing studies

Turaç Aydoğan¹ · Hakan Karşilar² · Yalçın Akın Duyan³ · Başak Akdoğan⁴ · Alessia Baccarani⁵ · Renaud Brochard⁵ · Benjamin De Corte⁴ · Jonathon D. Crystal⁶ · Bilgehan Çavdaroğlu⁷ · Charles Randy Gallistel⁸ · Simon Grondin⁹ · Ezgi Gür¹ · Quentin Hallez¹⁰ · Joost de Jong¹¹ · Leendert van Maanen¹² · Matthew Matell¹³ · Nandakumar S. Narayanan¹⁴ · Ezgi Özoğlu¹⁵ · Tutku Öztel¹⁶ · Argiro Vatakis¹⁷ · David Freestone¹⁸ · Fuat Balci^{1,16}

Accepted: 5 December 2022 / Published online: 3 January 2023
© The Psychonomic Society, Inc. 2022

Abstract

Interval timing refers to the ability to perceive and remember intervals in the seconds to minutes range. Our contemporary understanding of interval timing is derived from relatively small-scale, isolated studies that investigate a limited range of intervals with a small sample size, usually based on a single task. Consequently, the conclusions drawn from individual studies are not readily generalizable to other tasks, conditions, and task parameters. The current paper presents a live database that presents raw data from interval timing studies (currently composed of 68 datasets from eight different tasks incorporating various interval and temporal order judgments) with an online graphical user interface to easily select, compile, and download the data organized in a standard format. The *Timing Database* aims to promote and cultivate key and novel analyses of our timing ability by making published and future datasets accessible as open-source resources for the entire research community. In the current paper, we showcase the use of the database by testing various core ideas based on data compiled across studies (i.e., temporal accuracy, scalar property, location of the point of subjective equality, malleability of timing precision). The *Timing Database* will serve as the repository for interval timing studies through the submission of new datasets.

Keywords Big data · Experimental psychology · Interval timing · Time perception · Secondary analysis

✉ Fuat Balci
fuat.balci@umanitoba.ca

¹ Department of Biological Sciences, University of Manitoba, Winnipeg, Canada

² Department of Psychology, Özyeğin University, Istanbul, Türkiye

³ Department of Psychology, MEF University, Istanbul, Türkiye

⁴ Department of Psychology, Columbia University, New York, NY, USA

⁵ Département de Psychologie, Université Bourgogne Franche-Comté, Dijon, France

⁶ Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA

⁷ Department of Psychology, University of Toronto – Scarborough, Toronto, Canada

⁸ Center for Cognitive Science, Rutgers University, New Brunswick, NJ, USA

⁹ École de psychologie, Université Laval, Québec, Canada

¹⁰ Institut de Psychologie, Université Lumière Lyon 2, Bron, Lyon, France

¹¹ Department of Experimental Psychology, University of Groningen, Groningen, the Netherlands

¹² Department of Experimental Psychology, Utrecht University, Utrecht, the Netherlands

¹³ Department of Psychological and Brain Sciences, Villanova University, Villanova, PA, USA

¹⁴ University of Iowa, Neuroscience, Iowa City, IA, USA

¹⁵ Department of Psychology, University of Freiburg, Freiburg im Breisgau, Germany

¹⁶ Department of Psychology, Koç University, Istanbul, Türkiye

¹⁷ Department of Psychology, Panteion University of Social and Political Sciences, Athens, Greece

¹⁸ Equip Health, Inc., Carlsbad, CA, USA

Introduction

Interval timing refers to the ability of human and nonhuman animals to estimate, remember, and organize behaviors around intervals lasting seconds to minutes (Buhusi & Meck, 2005; Merchant & de Lafuente, 2014). Time-keeping is important not only as a “perceptual” process but also because of its integral role in other motor and cognitive functions such as motor planning, communication, learning, and decision-making (Balci et al., 2009a; Buhusi & Meck, 2005; Gallistel & Gibbon, 2000). Psychophysical properties of interval timing have revealed remarkable similarities across different species of animals (Meck, 2003). For instance, these studies have shown that subjects, on average, are highly accurate in adapting their timing behavior according to different target intervals (i.e., timed responses are clustered around target intervals). However, studies employing multiple intervals to be timed found that subjects can also deviate from temporal accuracy in systematic ways. These biases are described by *Vierordt's law*, which refers to the tendency of subjects to overestimate short intervals and underestimate long intervals (regression to the mean) observed in different time estimation methods (Lejeune & Wearden, 2009). Similarly, studies employing pharmacological, anatomical, or behavioral manipulations can systematically alter the accuracy of temporally organized behavior (e.g., Meck, 1983). Another consistently demonstrated behavioral benchmark of interval timing is the so-called *scalar property*, which refers to the constant proportionality of the variability in time judgments to the timed target interval (Gibbon, 1977). In other words, the error in temporal estimates scales with the estimated interval. Although numerous studies provided support for the scalar property in timing (Buhusi & Meck, 2005; Lejeune & Wearden, 2006; Wearden, 2016; Wearden & Lejeune, 2008), several others also showed that the scalar property is violated under certain conditions (Gibbon et al., 1997; Grondin, 2010, 2012, 2014; Staddon & Higa, 1999).

Other behavioral benchmarks were revealed based on tasks that required subjects to categorize intervals as “short” or “long” based on their subjective proximity to previously learned reference intervals (e.g., temporal bisection). For instance, studies utilizing the temporal bisection task have shown that the point of subjective equality (PSE; the duration at which both of the dichotomous responses are equally likely) typically ranges between the geometric ($(\prod_{i=1}^n x_i)^{\frac{1}{n}}$) and the arithmetic mean ($\frac{1}{n} \sum_{i=1}^n x_i$) of the reference intervals (Kopec & Brody, 2010). In other words, PSE ranges between the arithmetic mean of the reference durations on the logarithmic scale (equating the ratios) and on the linear scale (equating the

distances), respectively. Thus, the location of the PSE is a theoretically important endpoint as it can help elucidate the quantitative nature of mapping between subjective and objective time (e.g., linear vs. logarithmic) as well as the decision rules that underlie temporal judgments (e.g., difference vs. ratio rule).

Many studies of interval timing build on these types of widely accepted conclusions by using them as benchmarks for validating their tasks, selecting manipulation parameters, analyzing the data, and interpreting the results. Nonetheless, several caveats of such individual studies of interval timing are that they traditionally test a limited number of subjects—typically of only a single species, use a specific interval or a narrow range of intervals (e.g., sub-second vs. supra-second), and adopt a single procedure usually using a simple stimulus from a single modality. Consequently, the conclusions drawn from individual studies are not readily generalizable to other tasks, conditions, task parameters, and/or species. As such, there is a clear need to test the widely accepted effects in the interval timing literature by basing statistical analyses on larger datasets that are composed of previously published/unpublished data, in order to overcome the limitations of individual experiments. Furthermore, there is a need for testing new research questions, which are simply not addressable with the limited degrees of freedom provided by conventional individual studies. For instance, data gathered from individual studies cannot accommodate the investigation of research questions that would entail the analysis of behavioral data spanning milliseconds to minutes; a larger time scale than that employed in most conventional timing paradigms (Lewis & Miall, 2009). As an example, we examine here the extent to which scalar timing holds over such a broad range of intervals, we more precisely examine the accuracy and the nature of temporal mapping, and examine the malleability of temporal sensitivity. Given the substance and the variety of everyday cognitive processes that ordinarily encompass this larger time frame in their operation (e.g., organizing one’s schedule, language comprehension and production, fine-motor preparation and execution), the importance of elucidating such research questions becomes even more evident.

The umbrella term “big data” has so far been related mostly to social media, text corpora, finance, image and language processing, and geographic data. On the other hand, the availability of similarly large-scale data is also garnering excitement within the cognitive scientific fields such as psychology, neuroscience, and artificial intelligence (Harlow & Oswald, 2016). Regarding human cognition, putting together large-scale data from related areas of research has so far allowed for the formulation of novel theories of memory, linguistics, attention, vision, and decision-making, among others. On the other hand, researchers of human behavior have been argued to have developed a culture of not sharing

data (Munafò et al., 2017). This situation is arguably being exacerbated due to lack of convenient methods of sharing information, underutilization of sustainable forms of data storage, the widespread usage of proprietary data formats, and the diversity of experimental paradigms among laboratories. The *Timing Database* aims to overcome these gaps by promoting and cultivating core and novel analyses of timing behavior/estimates/judgments/ordering by making a large number of datasets accessible as open-source resources for the entire research community¹. The primary novelty of the *Timing Database* is that it constitutes the first repository for data collected from different research groups, incorporating different species, tasks, and task parameters while maximizing the readability and compilation of the data by imposing data structure standards common for all datasets. This database can be used for many fundamental analyses that include but are not limited to the comprehensive between-species comparison of interval timing behaviors, the discovery of novel psychophysical signatures of this function, elucidation of what common and different aspects of interval timing are captured by different tasks, and determining the factors that limit the generalizability of common assumptions regarding timing behavior. The *Timing Database* is currently composed of 68 datasets in a standardized format (over a million trials collected from 2346 subjects), contributed by individual researchers and gathered from publicly available datasets. These data were collected from humans (age range: 5–84 years), mice, and rats using eight different widely utilized tasks, and are kept on the servers of the Open Science Foundation (OSF). The *Timing Database* also offers an online graphical interface to easily select, compile, and download the data of interest (<https://timingdatabase.shinyapps.io/DownloadUI/>). It is designed to be a live and dynamic database that will continually accept submissions of new datasets and will be maintained by the *Timing Research Forum* (TRF; the official scientific community of timing researchers found at <http://timingforum.org>). This paper introduces the details of the database and showcases several multi-experimental analyses demonstrating the core topics (referring to fundamental questions) in interval timing literature.

Details of the database

The *Timing Database* is hosted on the OSF website (<https://osf.io/vrwjz/>). Each dataset is represented by two files—a data file in .csv format and a corresponding readme file in .txt format. All datasets contain the following fields: Species (e.g., human, rat, mouse, pigeon), Subject (ID), Session

(meaningful for multisession tasks), Block, Trial, Stim (stimulus modality used), Procedure (e.g., temporal bisection [TB], time estimation [TE], peak interval [PI]), Duration (set of durations used), Target (target duration used in the corresponding trial), Response (e.g., short/long response in TB, time-stamped responses in PI). In addition to these subfields, there may be additional task/experiment-specific fields depending on the nature of the study. This common template combines different tasks and datasets based on a single format, thereby easing their accessibility and aiding quick and comprehensive analysis.

The *readme* files that accompany individual datasets include critical information that are sufficient for decision-making during the reanalysis of the data. These files contain information including, but not limited to, the contributor of that dataset, any publication(s) associated with the dataset, the type and nature of the stimuli used in the experiment, and the study design including experimental manipulations and subject characteristics.

Crucially, the *Timing Database* contains a variety of tasks that allows a comprehensive reanalysis of the data. Thus, there is a specific format that encompasses all tasks in the database. These tasks include the temporal bisection task (TB), temporal reproduction task (TR), fixed interval/peak interval procedure (FI/PI), temporal order judgment (TOJ), differential reinforcement of low rates of responding (DRL), variable interval (VI), tapping (TA), and temporal estimation (TE).

Temporal reproduction task (TR)

Subjects are presented with a sensory stimulus (e.g., visual and/or auditory) that lasts for a predetermined duration (i.e., target interval). Subjects are then asked to reproduce the same interval by demarcating the reproduced interval with two key presses, a key to stop an automatically started duration, or by a continuous key depression. The comparison of the mean time reproduction shows timing accuracy, whereas the variability of these values (or their coefficient of variation, CV) is used as a measure of timing imprecision (Mioni et al., 2014).

Temporal bisection task (TB)

Subjects are initially trained to discriminate two reference intervals as short and long (Vatakis et al., 2018). They are then tested on their discrimination performance with a larger set of intervals that include both references in addition to a number of intermediate intervals. They are asked to categorize each of these intervals as “short” or “long” based on their subjective similarity to the reference durations. Responses to these intermediate intervals are unreinforced or non-differentially reinforced. Fits of sigmoidal functions

¹ Note that temporal order judgments (TOJ), which do not rely on interval timing, are also incorporated in the database.

to the proportion of long categorization responses as a function of the test intervals constitute the psychophysical function. The interval that corresponds to equal likelihood of both choices (i.e., 50% chance of long categorization) on the psychophysical function is referred to as the point of subjective equality (PSE). The normalized steepness of the psychophysical function denotes timing sensitivity (e.g., Weber's fraction, WF). Behavioral variations detected by virtue of these two parameters allow for tractable discussions regarding possible modulations in the so-called internal clock (Allman et al., 2014; Treisman, 1963). Although it is not commonly included in the conventional analyses of TB data, reaction times of short and long judgments are also differentially modulated by test intervals (Balci & Simen, 2014).

Synchronization and continuation tapping task (TA)

Subjects are asked to tap in synchrony with a rhythmic stimulus (e.g., metronome clicks) and to continue tapping at the same frequency after the rhythmic stimulus presentation has stopped. The inter-tap intervals and their normalized variability are used as the main unit of analysis in this task.

Temporal order judgment task (TOJ)

Subjects report the onset sequence of two events (i.e., order) as a function of different degrees of mismatch between the event onset times. Based on the assumption of varying levels of temporal resolution whilst timing an interval (akin to spatial resolution in vision research see Sternberg & Knoll, 1973), the task attempts to find the minimum duration of separation necessary for two events to be perceived in correct temporal order.

Temporal estimation task (TE)

Subjects are asked to indicate (either verbally or by using a slider/knob) how long a stimulus duration is. Thus, different from the tasks summarized above, timed judgments are not manifested by point responses. Similar to the TR task, the accuracy and the precision of estimates in the TE task are evaluated based on the proximity of the estimates to the actual experienced sample durations and their variability.

Differential reinforcement of low rates of responding task (DRL)

Subjects are trained to wait for a minimum amount of time between each response. Each response resets the elapsed time but only those emitted after the minimum wait time are reinforced. The mean and variability of the inter-response times are used as the critical units of analysis for

the optimality of the waiting response and timing uncertainty, respectively. Data from the DRL have been used for investigating the optimality of timed decisions (Çavdaroglu et al., 2014).

Fixed interval/peak interval task (FI/PI)

In the discrete FI schedule, subjects' first response following the fixed delay after the onset of a discriminative stimulus is reinforced. Well-trained subjects typically wait (e.g., until half of the FI elapses) before starting to respond in anticipation of the reward availability. The PI procedure includes FI trials but also probe trials in which the discriminative stimulus is presented much longer than the FI and the reward is omitted. These probe trials enable the characterization of timed responses without contamination by reward delivery. On a single trial, the points at which subjects abruptly increase the rate of responding are referred to as the "start time," and during probe trials, the trial time at which the rate of responding abruptly decreases is referred to as the "stop time." A more common but less refined way of characterizing timing performance requires dividing the trial duration into small time bins and calculating response rates for each. The proximity of the peak location of the resultant response curve to the target time is a measure of timing accuracy, whereas its spread is an index of timing uncertainty or imprecision.

Variable interval schedule (VI)

In the continuous VI task, the first response after a variable interval since the last response is reinforced. When the VI is an exponentially distributed random variable, this schedule constitutes an ideal control for time-based anticipation of reward availability. Discrete versions of the VI schedules can also be used. In this case, the reward availability becomes contingent upon the first response after a variable delay since the onset of the discriminative stimulus.

Example uses of the timing database

The *Timing Database* is designed to provide information at a granular level to support model testing, analysis of trial-to-trial adjustments, and learning, in addition to coarser-level descriptive analyses. The large size of the samples brings high statistical power to the reanalysis efforts. Here we demonstrate the example uses of the database by testing several core assumptions of the timing literature. To that end, we first tested whether timing behavior is accurate over a wide range of intervals using the temporal reproduction (TR) and the peak interval (PI) tasks in humans and nonhuman animals, respectively. We then tested whether the oft-reported *scalar property* holds in two interval timing tasks (e.g., TR,

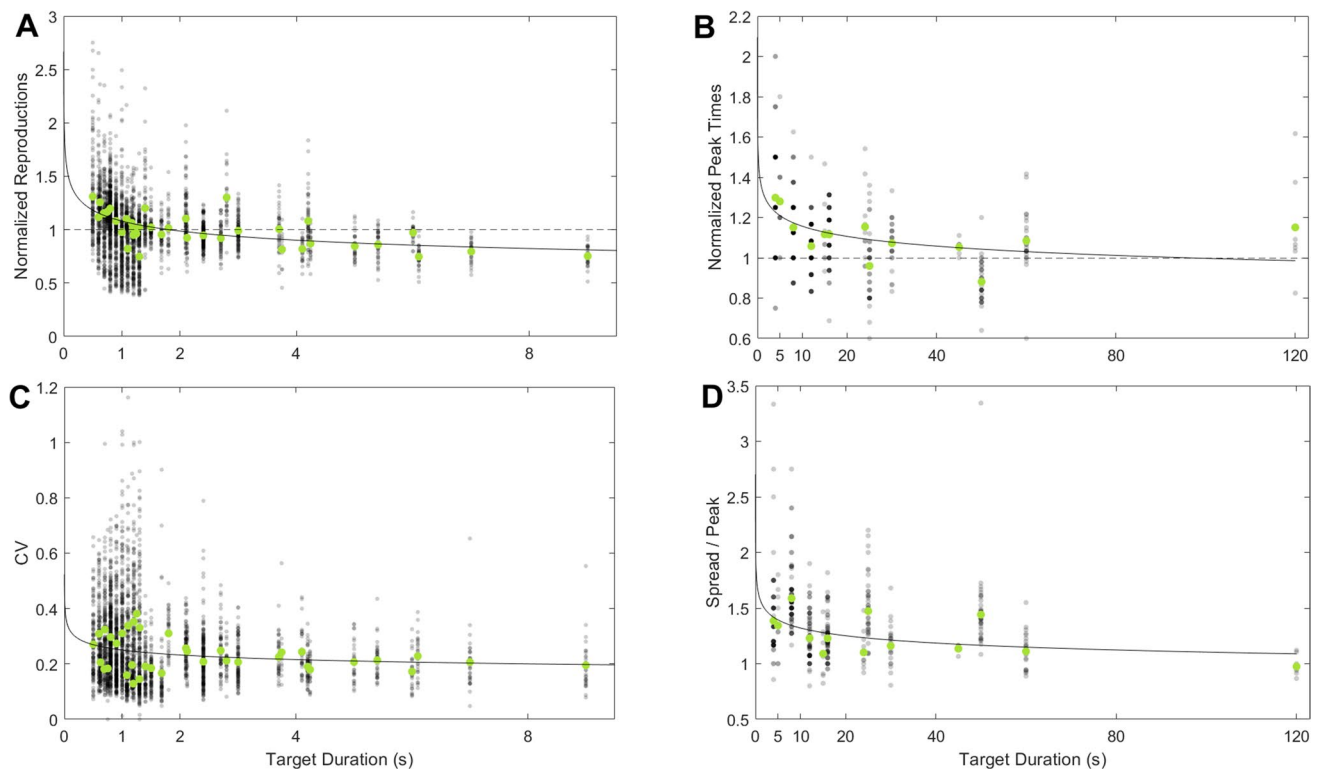


Fig. 1 **A, B** In order to visually study how temporal accuracy changed with target duration, normalized reproduction times (**A**) and peak times (**B**) are plotted as a function of the target duration. Black (darker) circles correspond to individual subjects' data whereas the green (lighter) circles correspond to the average data per target duration. A power function is then fit to the averaged normalized values

(green/lighter circles). **C, D** The same analysis is applied to the CV/normalized spread values to visually judge the scalar property. The power function had a greater effect size (**C**) than the linear fit (power fit adj. $R^2 = .189$, linear fit adj. $R^2 = .171$) in temporal reproduction, but they were approximately the same (power fit and linear fit adj. $R^2 = .227$) for the peak interval procedure (**D**)

PI) using a wide range of target durations. We followed these analyses by attempting to pinpoint the characteristic location of the PSE (i.e., comparing the arithmetic and the geometric means of reference intervals as the primary index of this parameter). Finally, we investigated whether temporal sensitivity of individuals in the TB task changes as a function of the task difficulty, defined by the ratio of short to long reference intervals. The first three questions test the three most fundamental assumptions of the interval timing literature. The fourth question explores whether temporal sensitivity is fixed or can be modulated by experimental demands, such as difficulty.

Results

Temporal accuracy and the scalar property

Timing behavior/time estimates are usually accurate on average (for review see Lejeune & Wearden, 2006; Wearden & Lejeune, 2008, for conformity to and violation of this property). We tested this assumption based on the TR task.

Figure 1A shows the time reproductions of human participants normalized by the target interval as a function of the target interval. These data showed that time intervals < 1.5 s were over-reproduced whereas longer intervals were slightly under-reproduced. This relation was captured by a power function. Figure 1B shows that a similar power function captures the data from nonhuman animals in the PI task.

Timing uncertainty is frequently reported to increase (scale) with the target interval, referred to in the timing literature as the scalar property. Therefore, it is typically assumed that the CV (ratio of the standard deviation of timed responses to their mean) stays constant across various time scales. This finding also accommodates the applicability of Weber's law (constancy of the discriminability of percepts) to the sense of time. The scalar property in timing, therefore, dictates that the timed responses should superpose when expressed on a relative (normalized) time scale. We tested this assumption using temporal reproduction in humans and peak interval procedure in nonhuman animals. For the TR task, the CVs of reproductions were calculated for each subject and the relationship between CV and target interval was characterized (see Fig. 1C). These data showed that CV was

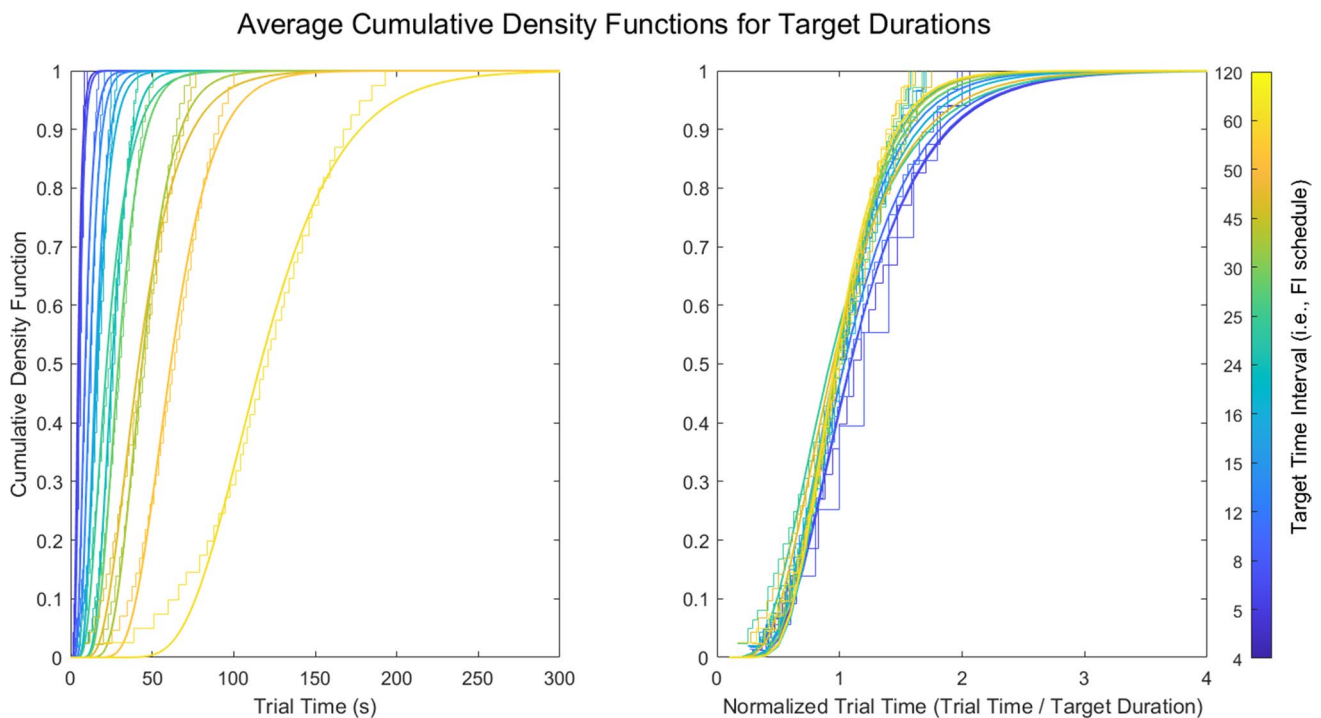


Fig. 2 Cumulative average response curves for all datasets that used the peak interval task. Color scale refers to the target interval (i.e., FI schedule in which subjects were trained)

sharply decreased as a function of time for intervals < 1.5 s, but this decrease was not detectable for longer intervals. Similar to TR data gathered from human participants, the peak response curves gathered from target intervals < 5 s violated superposition due to wider normalized spread at short durations (Fig. 1D). These data suggest that the scalar property holds for long-enough target intervals, which are different for human (> 1.5 s) and nonhuman animal participants (> 5 s). The PI data are plotted in a different form in Fig. 2 to show that not only do the first two statistical moments scale with duration, but the entire response function does as well.

Location of the psychometric function

The point of subjective equality is the duration that participants judge as equidistant to short and long reference intervals. We tested whether PSEs were best characterized by the arithmetic or the geometric mean of the reference intervals. Figure 3 shows the distribution of PSEs as normalized by the arithmetic mean (top panel) and geometric mean (bottom panel) of the reference intervals. These data show that PSE values estimated from human participants were distributed around the arithmetic rather than the geometric mean of the reference intervals. This observation suggests that subjective time is a linear function of objective time and that bisection decisions might rely on difference rather than ratio rule in humans.

Is temporal sensitivity malleable?

Temporal sensitivity (the precision with which durations are timed) is assumed to be a trait factor, which canonically determines variability in temporal judgments. Contradicting this view, Wearden and Ferrara (1996) have shown that the estimate of Weber's fraction gets smaller when the task is made more difficult by diminishing the ratio between the long and short reference intervals, presumably due to increased attention paid to the temporal portion of the task. The ratio between long and short reference intervals is taken as a measure of task difficulty since according to Weber's law, the discriminability of magnitudes decreases with decreasing ratio between the long and short reference intervals. We tested this hypothesis with the data available in the *Timing Database*. We confirmed that the estimated Weber's fraction decreased when the ratio of the reference durations was decreased ($\beta = -.84$, $SE = .074$, $p < .001$, adjusted $R^2 = .17$).

Discussion

As in many research domains in cognitive and behavioral sciences, research on interval timing is typically conducted in the form of relatively small-scale studies, each composed

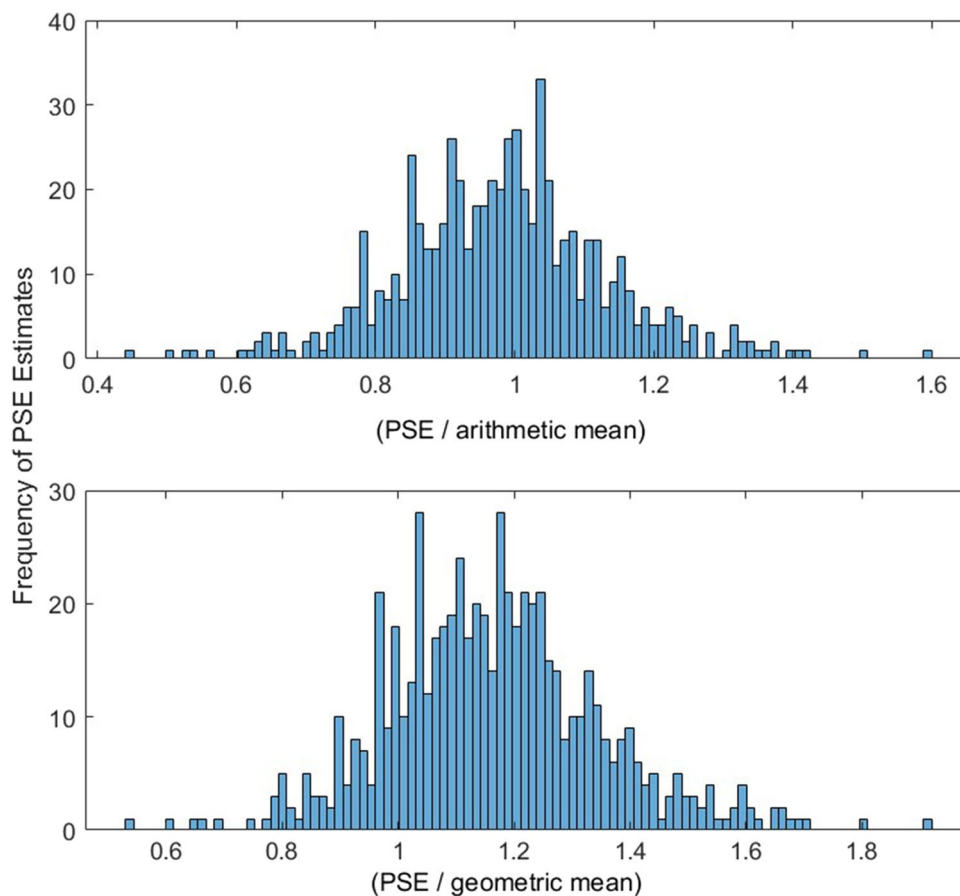


Fig. 3 Distribution of PSE values normalized by the arithmetic mean (top panel) and the geometric mean (bottom panel) of the reference intervals

of a single or at most a few experiments. This is because most studies are designed to test a specific research question and, therefore, use existing behavioral testing procedures with a limited range of task parameters, sometimes for practical and other times for theoretical reasons. Consequently, data gathered from individual studies do not lend themselves to testing research questions that entail the use of a wide range of task parameters. For instance, testing the scalar property of interval timing, as well as the few rules that appear to govern interval timing (e.g., Weber's law, Vierordt's law), should ideally involve testing how variability in time estimates changes as a function of a wide range of target intervals (e.g., ranging from milliseconds to minutes). The *Timing Database* aims to make such sophisticated analyses possible by utilizing online technologies and the abundance of as-of-yet unorganized data in the timing literature, whereby studies using various behavioral testing procedures are standardized, combined, and made available to everyone.

In this paper, we showcased how such large datasets can be used to test fundamental questions regarding temporal

information processing (i.e., temporal accuracy, scalar property, location of PSE, constancy of WF). Our analyses of temporal reproduction data showed that the scalar property holds after target intervals that are longer than 1.5 s in humans. Specifically, the relative timing variability decreased with increasing target intervals (see also Lewis & Miall, 2009). The higher variability that we observed for target intervals shorter than 1.5 s can be attributed to the contribution of the constant sensorimotor noise manifested in timing behavior. A similar pattern was observed in nonhuman animal data but over a different time scale. Similarly, violations of Weber's law have previously been reported with a much narrower range of intervals (with a U-shaped relation; Bizo et al., 2006). Consistent with this observation, our analyses of the nonhuman animal data gathered from the peak interval procedure showed deviations from superposition on the relative time scale for targets that were shorter than approximately 5 s. In support of earlier claims (e.g., Buhusi & Meck, 2005), these findings appear to support the argument that the automatic and cognitively controlled (and even sub- and supra-second intervals for human timing

behavior) might be underlain by disparate neural substrates with different information processing properties (Ivry, 1997).

Our analyses of temporal accuracy provided support for Vierordt's law (i.e., central tendency effect), even in data collected with a between-subjects experimental design. Specifically, durations shorter than 2 s were over-reproduced whereas there was a slight tendency to under-reproduce longer intervals (particularly targets longer than 1.5 s in humans). Note that these findings are difficult to explain by adhering to accounts such as Bayesian inference since the data in question were collected in a between-subjects fashion. Interestingly, a similar finding was recently observed in retrospective time judgments, where the most accurate estimates were observed for the 15-min duration (see also Chaumon et al., 2022). But readers should note that humans are always exposed to different intervals outside the experiment context, and it is possible that we have priors built based on these earlier experiences.

We also analyzed the location of the point of subjective equality in the temporal bisection tasks with respect to the geometric and arithmetic mean of the reference intervals. This is an important question since although PSE at the geometric mean can be accounted for by both logarithmic (with subtraction that refers to ratios in the linear scale) and linear time scales (with a ratio rule), PSE at the arithmetic mean can most readily be explained by a linear time scale (and not by logarithmic scale; unless exponentiating). Thus, the location of PSE has been a point of interest for many theoretical debates (Kopeck & Brody, 2010; Wearden & Ferrara, 1996). Our analysis on a large set of data showed that the location of PSE can be best accounted for by the arithmetic and not the geometric mean of the reference intervals. Briefly, our findings regarding PSEs support the linear time scale with a “difference rule” in human participants, and, thereby, weigh in strongly on this long-term debate (Crystal, 2006). Finally, we tested whether temporal sensitivity stayed constant or varied with task difficulty, presumably since task difficulty could modulate the amount of attention paid to the timing task. We found that timing sensitivity indeed increased with increasing task difficulty, suggesting that asymptotic estimates of timing sensitivity can be gathered only in tasks that are sufficiently challenging. These findings are consistent with an earlier report by Ferrara et al. (1997). Through the analyses outlined above, we demonstrated the potential uses of the *Timing Database* in addressing many different questions afforded by the richness of a large-scale database, and how these data can provide insights into answers for the most fundamental questions that relate to interval timing. These analyses will be repeated, and results will be shared on the *Timing Database* regularly as new data are admitted.

The *Timing Database* will continue accepting new datasets that not only use the procedures already listed here, but

also those collected with novel methods (including neural data), and is in that sense highly inclusive in scope. This dynamic and future-proof nature of the *Timing Database* will add to its potential use in studying even more complex questions and making truly new discoveries regarding timing behavior, which may very well surpass our current vision at its inception. To this end, authors of both new and previously published manuscripts in interval timing will be encouraged to submit their datasets in the prescribed format. Future dataset contributions will be required to abide by the data formatting standards in order to preserve the compatibility of the datasets contributed at different times. We will ensure that the new contributions will adhere to the prescribed standards by applying a script for quality checking (in Python and MATLAB), which has already been used with the currently available data. This code is also made available for future contributors as a supplement to this document, who are encouraged to use it before sending their data to us for quick and accurate inclusion in the database. In the case of updated data structure formats, all earlier datasets can be easily restructured in a high throughput fashion using the same code. If a correction or an update is made to the already contributed datasets, the contributors will need to include this information in the readme file, whereas the original version of the dataset will be preserved on the server and will be made available to researchers upon request (e.g., possibly to reevaluate the results of studies that may have used the earlier version of the data). Since each contribution is structured as a stand-alone entry to the database, the contribution of new datasets will not affect the structure and fidelity of earlier contributions. The current coverage of the data was based on the availability of data and responsiveness of researchers, but additional approaches including implicit versus explicit measures and comparative data are also welcome. There are no structural limitations on the number of data categories that can be integrated into the *Timing Database*.

A Web of Science search for all fields with the keywords “Interval Timing” or “Time Perception” has revealed on average around 250 papers annually between 2010 and 2020. These statistics alone point to the scale-up potential of the *Timing Database*. Authors will also be encouraged to share unpublished data on the database, which will help overcome the effects of publication biases. A further strength of the *Timing Database* is that it contains data collected not only from human participants (with varied demographics that makes it possible to address the developmental trajectory of timing ability) but also from other animals. This nature of the database allows researchers to make cross-species comparisons and analysis, and possibly shed light on the phylogenetic treatment of timing behavior. The sustainability of the *Timing Database* is ensured by the Timing Research Forum, the

official scientific community which already possesses a dedicated section for this undertaking (<http://timingforum.org/timing-database/>).

With this database, we also aim to contribute to the FAIR principles in the interval timing research domain. To this end, the database reinforces the findability of the raw data by collecting the corpus of data in one place that is well-known to the community. The accessibility, interoperability, and reusability of the data are reinforced by providing them in a standardized format and the online compiling tool that accompanies the database. Furthermore, downloading the same task from multiple labs/research groups, and then fitting it in a single multilevel model, allows us to more accurately estimate the robustness of the finding across labs and measure the variance in any given task across labs.

Due to the progress made in data storage technologies over the past two decades, researchers today have access to unprecedented amounts of information (i.e., big data). Big data has the potential to bring together similar domains of study within the cognitive sciences to elucidate historically and empirically fundamental questions which may have previously eluded scrutiny due to lack of sufficiently comprehensive data. For instance, although to a lesser extent in terms of the scope of the data, Gibbon and Balsam's (1981) made an important discovery decades ago by analyzing raw data from many laboratories: Their analysis showed that the learning rate in Pavlovian conditioning (1/ reinforcements to acquisition) was an approximately scalar function of a protocol's informativeness, which is the ratio (C/T) between the average inter-reinforcement interval (C for cycle duration) and the conditioned stimulus–unconditioned stimulus (CS – US) interval (T for trial duration). Gallistel and Latham (2022) review recent work confirming the generality of this law.

The idea for the inception of the *Timing Database* was conceived as a natural step within the current zeitgeist in order to effectively resolve the issues related to isolated empirical efforts and lack of data sharing in the domain of interval timing. Specifically with regard to research on timing and time perception, there are numerous theories—and related hypotheses—still awaiting scrutiny, with the potential to be challenged, revised, refined, or even downright rejected through the powerful lens that big data provides us. With this collaborative work, we aim to close any gaps which may be preventing this from happening due to lack of convenience or proper organization of knowledge within the scientific community. Thereby, we hope to inspire imagination, speed up collaboration, and possibly exponentiate the advancement of new and creative research within this highly dynamic field. Finally, our database is designed to meet interest in extending the database to more data than duration perception and thereby attract researchers collecting data in time structure (simultaneity judgments) and not time perception exclusively.

Methods

Temporal bisection

A two-parameter cumulative Weibull distribution was fit to the obtained probabilities of long responses to estimate PSE and WF. Two outliers (out of 623 individuals) were excluded from further analysis. In order to evaluate the location of the PSE, PSE estimates were evaluated with respect to the arithmetic and logarithmic mean of the reference intervals. Furthermore, WF values were regressed on the task difficulty.

Temporal reproduction

The dataset consists of reproduced durations per target duration/subject. Time reproductions were averaged per target duration and subject and normalized by the corresponding target duration. The CV of reproduced durations of each subject and target durations were calculated by dividing the standard deviation by the mean of these reproduced durations. Further, the reproductions that are below 33% and above 300% were excluded from the data analyses (3.16% of instances).

Peak interval

The dataset consists of trial-based time stamps of anticipatory responses per target duration/subject. Response times longer than three times the target duration (five times for target durations < 9 s) were discarded and the last 10 sessions were considered as corresponding to steady-state responding. The response times were binned into 1-second intervals. For the response curve analysis, binned data were averaged across trials per subject and smoothed by a factor of $19/30 \times$ target duration (with the exception of target durations < 9 s, which are smoothed by a factor equivalent to the target duration). The peak (maximum value of the smoothed response curve - Balci et al., 2009b) and the spread of the response curves were estimated based on response rates. The peak values larger than 2.5 times the target duration and normalized spread values larger than 5 were excluded from the analyses (2.5% of all data).

Acknowledgements Some of the datasets included in the *Timing Database* were gathered from publicly available sources, which were reformatted according to the *Timing Database* format.

Author contributions T.A. wrangled the data, designed the user interface. T.A., H.K., Y.D., E.G., and F.B. analyzed the data. All authors contributed to the final version of the manuscript and contributed their datasets.

Funding This work is funded by NSERC Discovery Grant, RGPIN/3334-2021 to FB.

Data availability The *Timing Database* will continue accepting new datasets. Clear point-by-point instructions for new submissions, along with user-friendly quality control scripts in Python and MATLAB are provided on the OSF page of the database. Datasets will be formatted by the contributors (either during the data collection or after the experiment) and the required .csv and readme.txt files will be sent to timing-database@gmail.com. The database manager assigned by the *Timing Research Forum* will once again graphically check the received data and add it to the database. An online graphical user interface accompanies the database to facilitate the searching, compiling, visualizing, and downloading of the data (<https://timingdatabase.shinyapps.io/DownloadUI/>). Authors of the papers on interval timing will be encouraged by the community to submit the related datasets to the *Timing Database* to meet the Findability, Accessibility, Interoperability, and Reusability (FAIR) principles for their datasets. Researchers using the database in their work will be reminded to cite the *Timing Database* paper along with the original papers that are associated with the data they use.

Code availability Codes to the data analysis are available on the OSF website (<https://osf.io/vrwjz/>).

Declarations

Ethics approval All data included in this database are approved by the local IRB and IACUC.

Consent to participate Consent to participate was received from all subjects.

Consent for publication N/A

Competing interests The authors declare no competing interests.

References

- Allman, M. J., Teki, S., Griffiths, T. D., & Meck, W. H. (2014). Properties of the internal clock: First- and second-order principles of subjective time. *Annual Review of Psychology*, 65(January), 743–771. <https://doi.org/10.1146/annurev-psych-010213-115117>
- Balci, F., Freestone, D., & Gallistel, C. R. (2009a). Risk assessment in man and mouse. *Proceedings of the National Academy of Sciences of the United States of America*, 106(7), 2459–2463. <https://doi.org/10.1073/pnas.0812709106>
- Balci, F., Gallistel, C. R., Allen, B. D., Frank, K. M., Gibson, J. M., & Brunner, D. (2009b). Acquisition of peak responding: What is learned? *Behavioural Processes*, 80(1), 67–75. <https://doi.org/10.1016/j.beproc.2008.09.010>
- Balci, F., & Simen, P. (2014). Decision processes in temporal discrimination. In *Acta Psychologica* (Vol. 149, pp. 157–168). <https://doi.org/10.1016/j.actpsy.2014.03.005>
- Bizo, L. A., Chu, J. Y. M., Sanabria, F., & Killeen, P. R. (2006). The failure of Weber's law in time perception and production. *Behavioural Processes*, 71(2–3), 201–210. <https://doi.org/10.1016/j.beproc.2005.11.006>
- Buhusi, C. V., & Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nature Reviews Neuroscience*, 6(10), 755–765. <https://doi.org/10.1038/nrn1764>
- Çavdaroğlu, B., Zeki, M., & Balci, F. (2014). Time-based reward maximization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1637). <https://doi.org/10.1098/rstb.2012.0461>
- Chaumon, M., Rioux, P. A., Herbst, S. K., Spiouas, I., Kübel, S. L., Gallego Hiroyasu, E. M., & van Wassenhove, V. (2022). The Blursday database as a resource to study subjective temporalities during COVID-19. *Nature human Behaviour*, 1–13.
- Crystal, J. D. (2006). Sensitivity to time: Implications for the representation of time. In E. A. Wasserman & T. R. Zentall (Eds.), *Comparative cognition: Experimental explorations of animal intelligence* (pp. 270–284). Oxford University Press.
- Ferrara, A., Lejeune, H., & Wearden, J. H. (1997). Changing sensitivity to duration in human scalar timing: An experiment, a review, and some possible explanations. *Quarterly Journal of Experimental Psychology Section B: Comparative and Physiological Psychology*, 50(3), 217–237.
- Gallistel, C. R., & Gibbon, J. (2000). *Time, Rate, and Conditioning*. 107(2), 289–344.
- Gibbon, J., Malapani, C., Dale, C. L., & Gallistel, C. R. (1997). Toward a neurobiology of temporal cognition: Advances and challenges. *Current Opinion in Neurobiology*, 7(2), 170–184.
- Gallistel, C. R., & Latham, P. E. (2022). Bringing Bayes and Shannon to the Study of Behavioral and Neurobiological Timing. *Frontiers in Behavioral Neuroscience*.
- Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, 84(3), 279–325. <https://doi.org/10.1037/0033-295X.84.3.279>
- Gibbon, J., & Balsam, P. D. (1981). Spreading associations in time. In C. M. Locurto, H. S. Terrace, & J. Gibbon (Eds.), *Autoshaping and conditioning theory* (pp. 219–253). Academic.
- Grondin, S. (2010). Unequal weber fractions for the categorization of brief temporal intervals. *Attention, Perception, & Psychophysics*, 72, 1422–1430. <https://doi.org/10.3758/APP.72.5.1422>
- Grondin, S. (2012). Violation of the scalar property for time perception between 1 and 2 seconds: Evidence from interval discrimination, reproduction, and categorization. *Journal of Experimental Psychology: Human Perception and Performance*, 38(4), 880–890. <https://doi.org/10.1037/a0027188>
- Grondin, S. (2014). About the (non)scalar property for time perception. *Advances in Experimental Medicine and Biology*, 829, 17–32. https://doi.org/10.1007/978-1-4939-1782-2_2
- Harlow, L. L., & Oswald, F. L. (2016). Big data in psychology: Introduction to the special issue. *Psychological Methods*, 21(4), 447–457. <https://doi.org/10.1037/met0000120>
- Ivry, R. (1997). Cerebellar timing systems. *International Review of Neurobiology*, 41, 555–573. [https://doi.org/10.1016/s0074-7742\(08\)60370-0](https://doi.org/10.1016/s0074-7742(08)60370-0)
- Kopec, C. D., & Brody, C. D. (2010). Human performance on the temporal bisection task. *Brain and Cognition*, 74(3), 262–272. <https://doi.org/10.1016/j.bandc.2010.08.006>
- Lejeune, H., & Wearden, J. H. (2006). Scalar properties in animal timing: Conformity and violations. *Quarterly Journal of Experimental Psychology*, 59(11), 1875–1908. <https://doi.org/10.1080/17470210600784649>
- Lejeune, H., & Wearden, J. H. (2009). Vierordt's *the experimental study of the time sense* (1868) and its legacy [review of the book *the experimental study of the time sense*, by K. Vierordt]. *European Journal of Cognitive Psychology*, 21(6), 941–960. <https://doi.org/10.1080/09541440802453006>
- Lewis, P. A., & Miall, R. C. (2009). The precision of temporal judgement: Milliseconds, many minutes, and beyond. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1525), 1897–1905. <https://doi.org/10.1098/rstb.2009.0020>
- Meck, W. H. (1983). Selective adjustment of the speed of internal clock and memory processes. *Journal of Experimental Psychology: Animal Behavior Processes*, 9(2), 171–201. <https://doi.org/10.1037/0097-7403.9.2.171>
- Meck, W. H. (Ed.). (2003). *Functional and neural mechanisms of interval timing*. CRC Press/Routledge/Taylor & Francis Group. <https://doi.org/10.1201/9780203009574>

- Merchant, H., & de Lafuente, V. (2014). Introduction to the neurobiology of interval timing. *Advances in Experimental Medicine and Biology*, 829, 1–13. https://doi.org/10.1007/978-1-4939-1782-2_1
- Mioni, G., Stablum, F., McClintock, S. M., & Grondin, S. (2014). Different methods for reproducing time, different results. *Attention, Perception, and Psychophysics*, 76(3), 675–681. <https://doi.org/10.3758/s13414-014-0625-3>
- Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Du Sert, P. N., & N., Simonsohn, U., Wagenmakers, E. J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1(1), 1–9. <https://doi.org/10.1038/s41562-016-0021>
- Staddon, J. E. R., & Higa, J. J. (1999). Time and memory: Towards a pacemaker-free theory of interval timing. *Journal of the Experimental Analysis of Behavior*, 71(2), 215–251. <https://doi.org/10.1901/jeab.1999.71-215>
- Sternberg, S., Knoll, R. L. (1973). *The perception of temporal order: Fundamental issues and a general model*. Bell Laboratories, Murray Hill, New Jersey.
- Treisman, M. (1963). Temporal discrimination and the indifference interval. *Psychological Monographs: General and Applied*, 77(13), 1–31.
- Vatakis, A., Balci, F., Di Luca, M., & Correa, Á. (2018). Timing and Time Perception: Procedures, Measures, & Applications. In *Timing and time perception: Procedures, measures, & applications*. <https://doi.org/10.1163/9789004280205>
- Wearden, J. (2016). The psychology of time perception. In *The Psychology of Time Perception*. <https://doi.org/10.1057/978-1-137-40883-9>
- Wearden, J. H., & Ferrara, A. (1996). Stimulus range effects in temporal bisection by humans. *Quarterly Journal of Experimental Psychology Section B: Comparative and Physiological Psychology*, 49(1), 24–44. <https://doi.org/10.1080/713932615>
- Wearden, J. H., & Lejeune, H. (2008). Scalar properties in human timing: Conformity and violations. *Quarterly Journal of Experimental Psychology*, 61(4), 569–587. <https://doi.org/10.1080/17470210701282576>

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.