RUNNING HEAD: DYNAMIC TIMING VIOLATES SCALAR PROPERTY

Mice Fail to Show the Scalar Property When Timing Sinewaves

Carter W. Daniels1,2, Margaret Zhong2, & Peter D. Balsam1,2

1Department of Psychiatry, Columbia University

2Department of Psychology, Barnard College

Word Count: XXX

Send Correspondence To:

Carter W. Daniels

E-mail: [carter.wa.daniels@gmail.com](mailto:carter.wa.daniels@gmail.com)

Abstract

[INSERT ABSTRACT TEXT HERE]

*Keywords:* Interval Timing, Time Perception, Temporal Tracking, Sinewave, Dynamic Timing

**Mice Fail to Show the Scalar Property When Timing Sinewaves**

**Introduction**

The study of interval timing—timing in the seconds-to-minutes range—has long been explored in the context of static intervals; intervals that do not change appreciably across trials or sessions (XXX, XXX, XXX, XXX, XXX, XXX). These studies have yielded two fundamental regularities of interval timing. Intervals are learned rapidly (XXX, XXX, XXX; but see XXX); and once learned, the perception of such intervals adhere to a strong form of Weber’s law known as the Scalar Property in which estimates of normalized variance is constant across a wide range of intervals (XXX, XXX, XXX, but see XXX). However, it is unclear to what extent these regularities hold in the context of dynamic intervals such as intervals that change trial-to-trial (e.g., XXX). Staddon and colleagues studied timing when intervals change trial-to-trial while outlining potential mechanisms (for reviews see Luzardo, Ludvig, & Rivest, 2013; Staddon, Chelaru, & Higa, 2002a, b; Wynne, Staddon, & Delius, 1996); however, few other researchers have attempted to replicate, let alone expand upon the garnered insights (but see Sanabria & Oldenburg, 2013). In the present paper we aimed to replicate and expand Staddon and colleagues research by studying dynamic interval timing in mice.

Staddon and colleagues largely studied dynamic interval timing via two timing procedures: a response-initiated delay (RID) and fixed-interval schedule of reinforcement (FI). In RID, subjects initiate intervals via a single response (e.g., pigeons peck a key or rats press a lever); after the interval elapses, response-independent reward is delivered. In FI, intervals are typically initiated by the experimenter (e.g., via illuminating a light, extending a lever, turning on an audible tone; but see XXX and XXX for RI variants of FI) and response-dependent reward is delivered following the first response the interval elapses. To keep timing performance comparable between both procedures, we focus on Staddon and colleagues preferred measure of timing performance: the initial pause following reinforcement and just before the first response. This time is referred to as the post-reinforcement pause (PRP) or latency-to-the-first response (LFR). Importantly, prior work indicates both the PRP and LFR are sensitive to the passage of time (XXX, XXX, XXX).

Staddon and colleagues found that the PRP in both pigeons and rats was a useful metric of dynamic interval timing. A brief review of this research suggests some stand-out features of dynamic interval timing worth noting. First, like the tracking of static intervals, tracking of dynamic intervals appears abruptly with few notable changes in timing performance with continued training (XXX, XXX; but see XXX). Indeed, PRPs appear to track intervals with a slight lag (1-2 intervals; XXX, XXX). Second, tracking is sensitive to whether intervals are growing or shortening with animals adjusting PRPs quickly to shortening intervals but slowly to growing intervals; likewise, animals are able to track short impulses (i.e., an unexpected, large change in the interval for one trial) but not long impulses. Interestingly, a parallel can be found in static timing where after being trained on relatively longer intervals mice can be transferred to short intervals with little training but require more training for longer intervals (XXX; XXX; XXX). Third, tracking is sensitive to sequence composition. Although pigeons can rapidly track changing intervals within a session or across days following a variety of sequences (sinewave, square-wave, etc. XXX; XXX; XXX), whether animals are able to track the sequence depends in part on the number of short and long intervals that compose that sequence (XXX, XXX) and, relatedly, how quickly the sequence is moving through the short and long phases of the sequence (XXX, XXX). Fourth, and perhaps most surprising, is that when tracking dynamic intervals the PRP fails to adhere to the Scalar Property (XXX, XXX, XXX).

Timing performance that fails to adhere to the Scalar Property is not unprecedented. Indeed, the Scalar Property fails in the sub-second time scale (XXX, XXX, XXX), some strains of mice (XXX, XXX), specific response topographies (XXX, XXX), professionally trained musicians (XXX), and potentially in some animals because of contamination due to high rates of responding (XXX). The mechanisms proposed to explain these failures are as unique as the failures themselves; but more importantly, it is worth mentioning that these failures provide constraints on when the Scalar Property is observed. Staddon and colleagues finding may indicate one of the more interesting constraints on the emergence of the Scalar Property as dynamic intervals may be more akin to the real-world experiences of time than static intervals. Indeed, it is easy to imagine how rewards along with other biologically significant events likely occur with some varying periodicity or seasonality (XXX, XXX).

However, the replicability of Staddon and colleagues is unclear as few outside of their group have studied the Scalar Property in dynamic interval timing. Although Staddon and colleagues have generalized their work across procedures (response-initiated delay: XXX; fixed-interval schedules of reinforcement: XXX) and species (pigeons: XXX; rats: XXX), they relied heavily on the post-reinforcement pause (PRP) as their preferred measure of timing. Previous work indicates the PRP is sensitive to many variables including fluctuations in motivation (XXX, XXX), presence or absence of a preceding reinforcer (XXX, XXX), and contaminated by random responding (XXX, XXX). Such sensitivity suggests the PRP is providing only a rough estimate of timing performance.

Luckily, one of the most common timing procedures—fixed-interval (FI) schedules of reinforcement—allows for assessment of both the PRP and what is commonly referred to as the breakpoint (BP). Briefly, in FI schedules of reinforcement animals are trained such that the first response after the interval has elapsed is reinforced (XXX). Responding in FI typically starts at a relatively low rate and then abruptly transitions to a relatively high rate as the end of the FI approaches (XXX; XXX; but see XXX for a nuanced take). Whereas the post-reinforcement pause is the time to the first response and thus likely part of the low rate of responding, the BP is the time at which animals transition from the low rate to the high rate. In contrast to the PRP, the BP is relatively robust to non-timing variables making it a cleaner estimate of timing (XXX; XXX; XXX). Thus, in the present study we trained mice on a dynamic fixed-interval schedule of reinforcement and assessed their timing performance as expressed in both the PRP and BP.

**Methods**

**Subjects.**

Twenty-four male C57BL/6J mice (Jackson Laboratories, Bar Harbor, ME) served as subjects. All mice were housed on a 12/12-hour light/dark cycle (lights on/off at 7:00 AM/ 7:00 PM). Behavioral training and food restriction was implemented such that their body weight was maintained at 85% of their free food baseline weight, determined by baseline measurements taken three days prior to removing free food. Subjects had free access to water at their home cages for the entire duration of the experiment. All animal handling procedures followed National Institutes for Health guidelines and were approved by the New York State Psychiatric Institute’s Institutional Animal Care and Use Committee.

**Apparatus.**

Experiments were conducted in 24 MED associates (St. Albans, VT, USA) modular test chambers (MED Associates Inc. ENV-307W; 22x18x13cm). The front and back walls and the ceiling of test chambers were made of Plexiglas; the front wall was hinged and acted as a door to the test chamber. The floor consisted of thin metal bars, suspended over a metal catch tray filled with corn bedding. The reinforcer receptacle was a circular opening located around 1cm above the floor and centered on the test panel. The receptacle provided access to a dipper (MED Associates, ENV-302W-S) fitted with a cup (ENV-202C) that holds 0.01cc of a liquid reinforcer (Parrot Brand Evaporated Milk, San Francisco, CA). The receptacle was furnished with a head-entry detector (ENV-254-CB). On either side of the reinforcer receptacle, there were two rectangular holes where metal levers could be presented (ENV-312-3W). A house light (ENV-315W-LED) was located behind the wall opposite to the reinforcer receptacle. Experimental events were arranged via a MED PC interface connected to a PC controlled by MED-PC IV software.

**Procedure.**

All training sessions for each group were conducted once daily, 5 days a week. All mice were trained

Pre-training.

**Data Analysis & Results**

All data and analyses are available at the associated Github (<https://github.com/drcwadaniels/variance_cyclic_timing>) or OSF.io repository (DOI 10.17605/OSF.IO/M6AG5). Here we present only those analyses and information necessary for testing whether sinewave timing adheres to the scalar property.

**First Impressions Suggests Dynamic Timing Performance is Scalar Invariant.**

After 36 sessions of training, all mice in each training group had met the training criterion of the mean breakpoint on both the minimum, base, and max interval not perceptibly changing for at least 12 sessions. Figure 1 Left Panel shows on the left y-axis the average breakpoint of mice in each training group (Base 20, Base 40, Base 60) and on the right y-axis the programmed FI as a function of trial. Visually, these data indicate that mice learned to track temporal sinewaves; and that the fidelity of this tracking was not obviously subject to any lag relative to the programmed FI. Indeed, breakpoints reached their minimum and maximum on the same trial as the programmed FI. This indicates that performance is likely based on some mixture of memory for the immediately previous and expectation of the immediately forthcoming trial.

In the present paper, however, we are not concerned with the fidelity of dynamic timing. Instead, we are interested in whether performance adheres to the scalar property. As a first pass, we looked at whether the mean breakpoint as a function of trial for each FI superimposes when normalized by the Base FI. Figure 1 Right Panel shows the same data as Figure 1 Left Panel but normalized by the average breakpoint on the Base FI. Inspection of this figure suggest that breakpoints largely superimpose with some divergence around the short FIs comprising the troughs. This degree of superimposition suggests initial evidence that dynamic timing is scalar invariant. To confirm that this overlap was consistent with the scalar property, we asked whether parameter estimates of a sinewave differed between training groups in both the raw and normalized breakpoints. Sinewave parameters were estimated in Python using a custom function in which we passed to an optimizer the both the raw and the normalized data with the following equation,

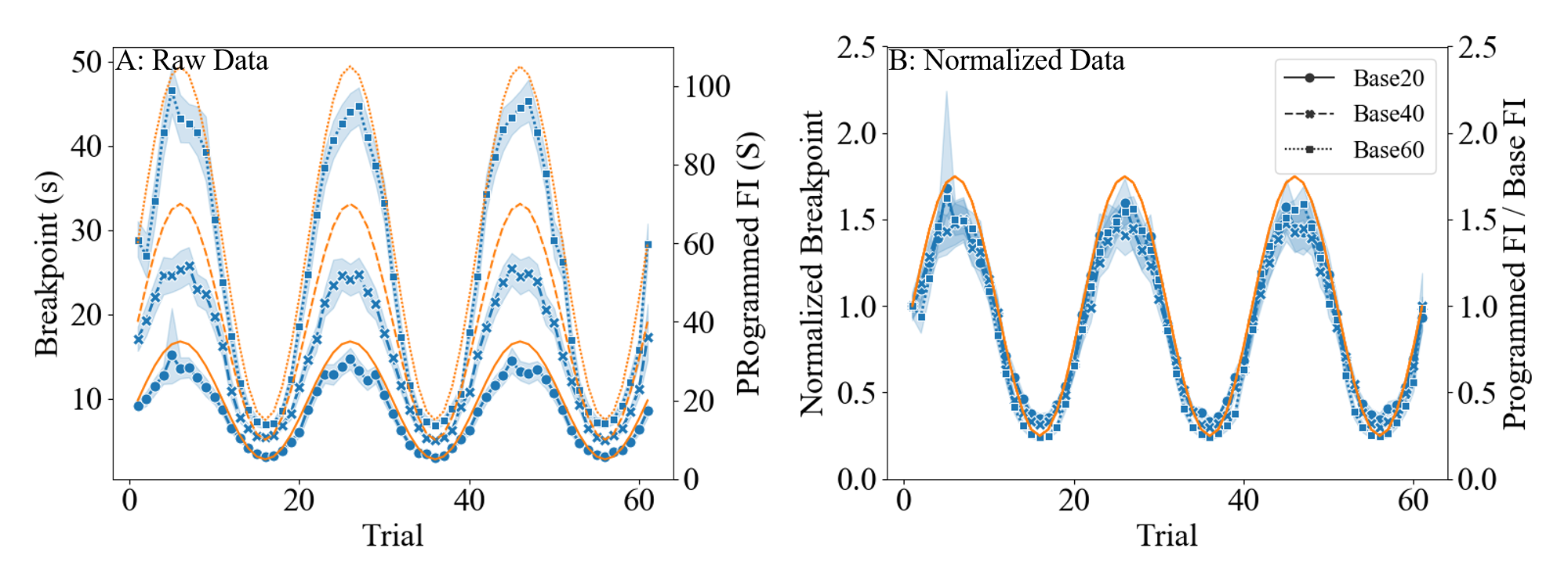
, (1)

where *B* is the base interval of the FI sinewave, *S* is the ± swing from B, and *P* is the period or number of trials it takes for a full oscillation (base 🡪 peak 🡪 base 🡪 trough; or base 🡪 trough 🡪 base 🡪 peak) to occur and *t* is the trial indexer. Log transformed parameter estimates (B, P, and S) were analyzed using both NHST and Bayesian variants of ANOVA and post-hoc *t*-tests with a holm correction*.*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Table 1. Mean (±SEM) Sinewave Parameter Estimates | | | | | | |
|  | Raw Data | | | Normalized Data | | |
| Base FI | B | S | P | B | S | P |
| 20 | 8.48 (0.42) | 5.62 (0.46) | 20 (0.02) | 0.93 (0.04) | 0.61 (0.04) | 20 (0.02) |
| 40 | 15.14 (0.80) | 10.38 (0.80) | 20 (0.02) | 0.89 (0.02) | 0.60 (0.03) | 20 (0.02) |
| 60 | 25.03 (0.68) | 19.89 (0.89) | 19.89 (0.89) | 0.88 (0.03) | 0.69 (0.02) | 19.89 (0.01) |

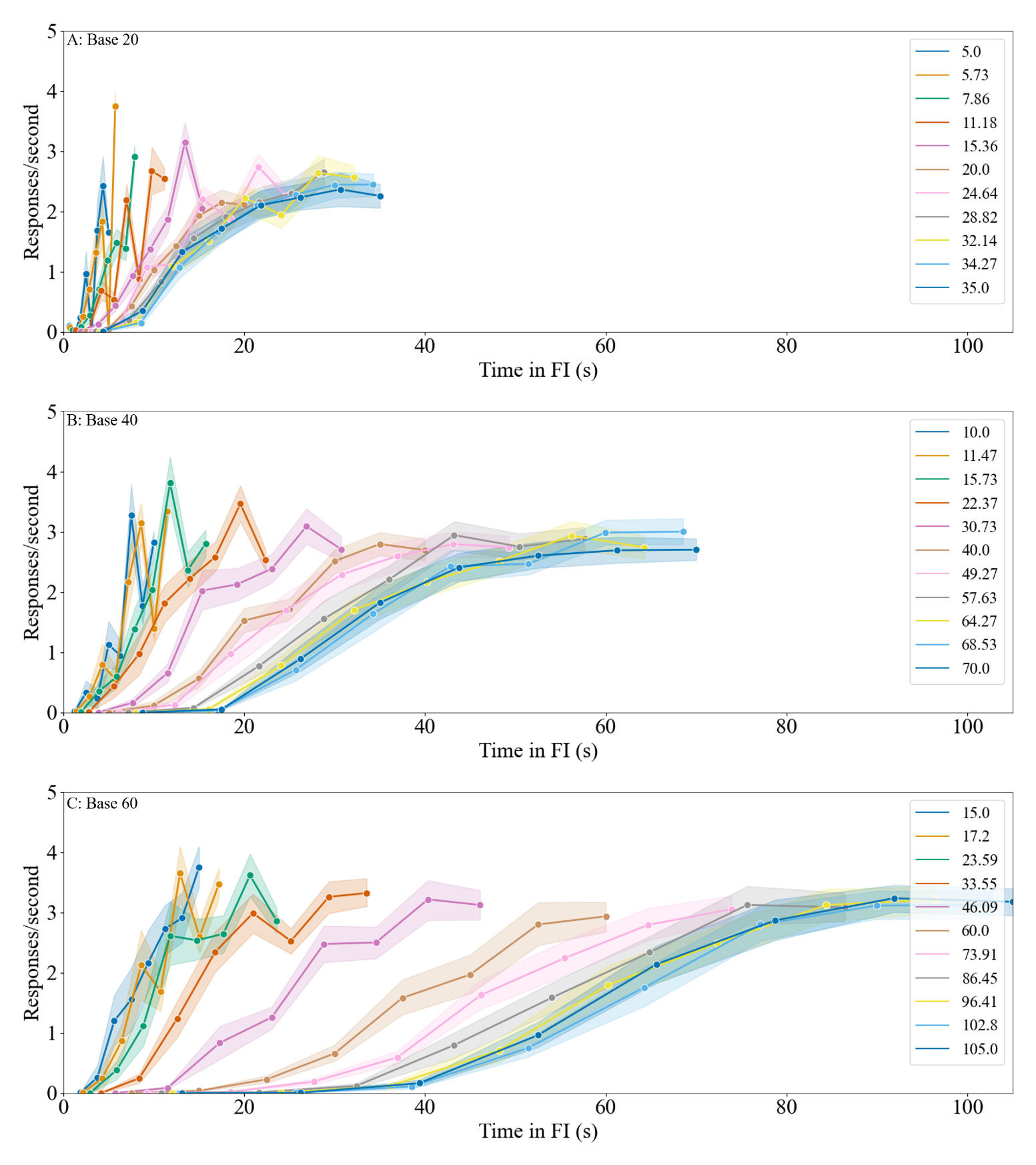
*Note.* These estimates were calculated on the linear rather than on the log scale following by being exponentiated back because we found that the obtained effects and/or lack thereof did not depend on the scale data was analyzed.

Mean (± SEM) Eq. 1 parameter estimates in Table 1 suggest that whereas *P* is constant across groups at approximately 20 thus reflecting the programmed sinewave, *B* and *S* scale with Base FI and that such scaling is largely eliminated by normalization. Consistent with these impressions, we found substantial effects of Base FI for raw B [*F*(2,21) = 158.92, *p* < 0.0001, *ln*BFi0 = 22.75] and S [*F*(2,21) = 84.26, *p* < 0.0001, *ln*BFi0 = 17.67] that was largely absent in normalized B [*F*(2,21) = 0.36, *p* = 0.702, *ln*BFi0 = -1.07] and S [*F*(2,21) = 4.14, *p* = 0.031, *ln*BFi0 = 0.21]; Post-hoc t-tests indicated that each Base FI was significantly different from the other Base FI s in raw [all *p*s < 0.0001; all *ln*BFi0 ≥ 5.27] but not normalized B and S [all *p*s ≥ 0.14; all *ln*BFi0 ≤ 1.25]. No significant effects of Base FI on raw [*F*(2,21) = 0.68, *p* = 0.52, *ln*BFi0 = -1.14] or normalized [*F*(2,21) = 0.68, *p* = 0.52, *ln*BFi0 = -1.14] P were detected.

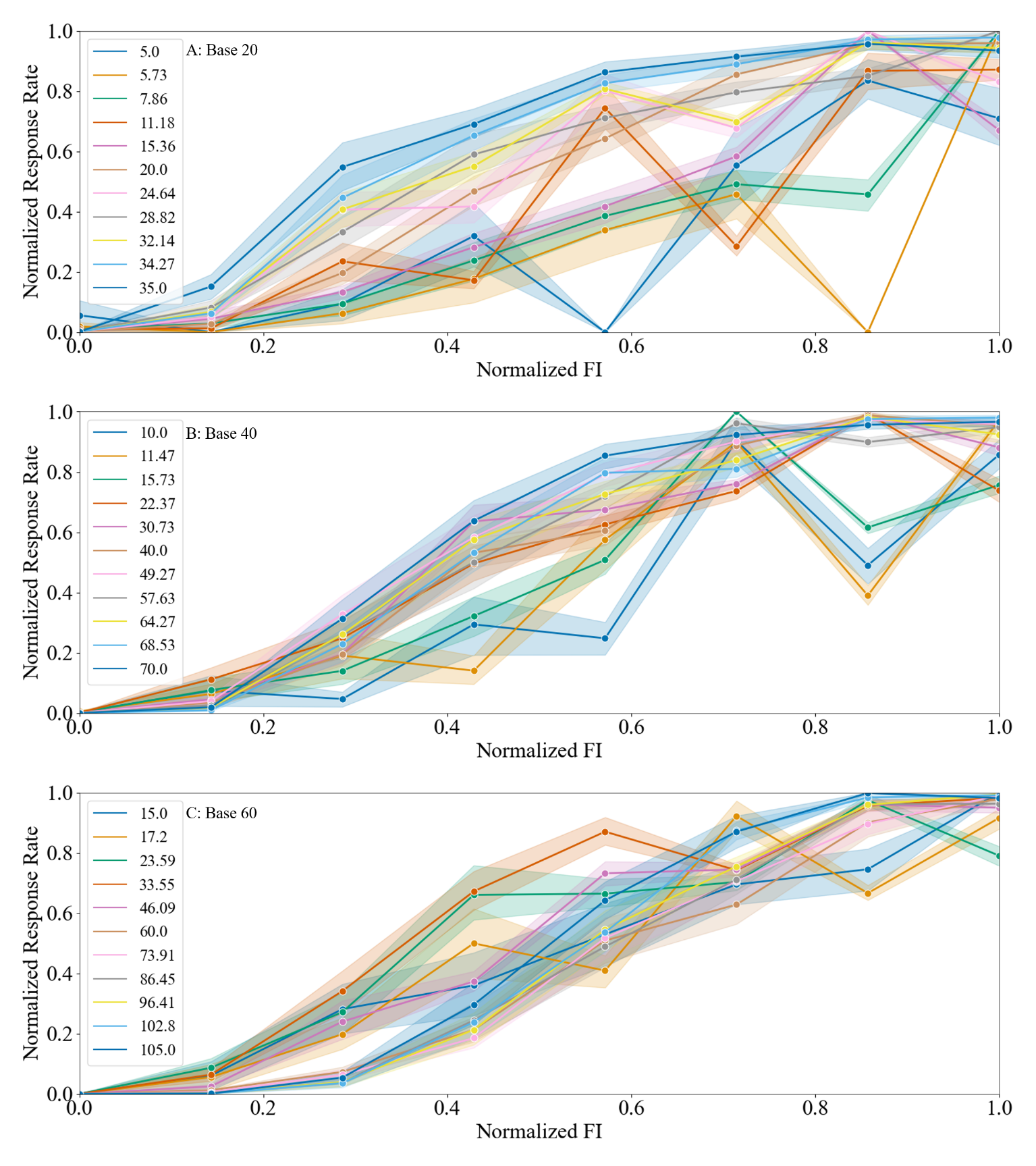
**Figure 1.** Mean raw (Panel A) BPs (blue) and programmed FIs (orange) and normalized (Panel B) BPs and programmed FIs as a function of trial into a session (averaged over 12 sessions) for each training group (Base 20: solid line, circles; Base 40: dashed line, xs; Base 60: dotted line, squares).

**Dynamic Timing Performance is NOT Scalar Invariant when Analyzed as a Function of FI**

To conclude that dynamic timing performance adheres to the scalar property, we need to perform a strong test of the scalar property. In addition to BPs superimposing, it needs to be the case that response rate as a function of time in the FI also superimposes. First, we confirmed that response rates in individual FIs, despite being programmed to increment and decrement following a sine wave, follow the typical sigmoidal shape found in animals trained on static FIs. Figure 2 shows response rates as a function of time in the FI for each FI presented to each training group (bin epoch = 0.125\*FI). Visual inspection suggests that indeed, response rates start at a relatively low rate and then are negatively accelerated until reaching a plateau just before the time of reinforcement. Figure 3A-C shows the same data but with response rates normalized by the max response rate and time normalized by the programmed FI. Visual inspection suggests lack of superimposition among normalized response rates. Base 20 and Base 40 mice respond at higher rates early in FIs around the peak of the sinewave compared to FIs around the trough of the sinewave. In contrast, Base 60 mice respond at higher rates early in FIs around the trough of the sinewave compared to FIs around the peak of the sinewave. This suggests that dynamic interval timing may not be scalar invariant.



**Figure 2.** Mean raw response rates as a function of Time in the FI (s) for Base20 (Top Panel), Base40 (Middle Panel), and Base60 (Bottom Panel) mice. Response rate functions were constructed by binning responses into bins equal to 12.5% of each FI and then dividing by the number of corresponding seconds.



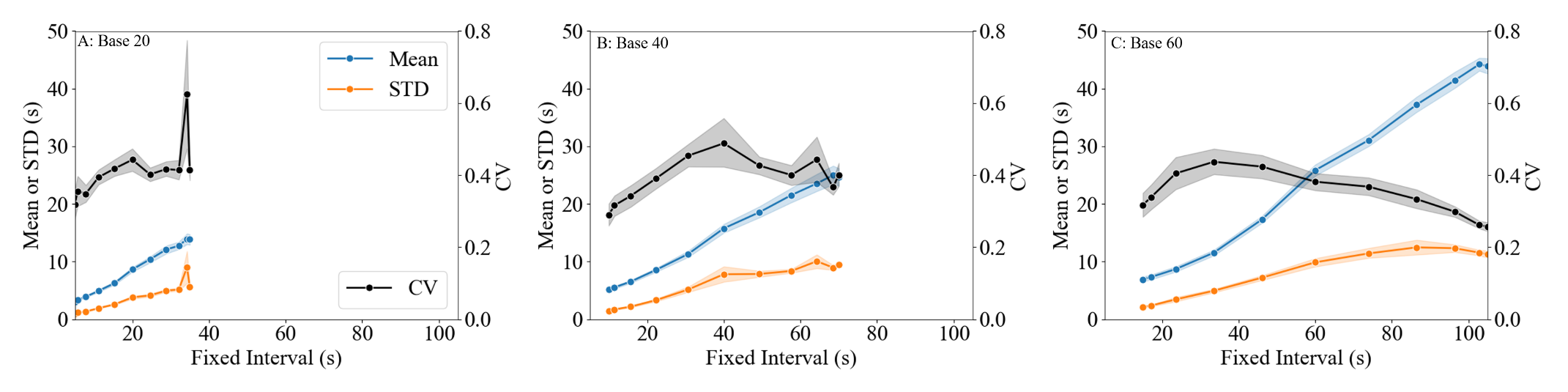
**Figure 3.** Normalized raw response rates as a function of Time in the FI (s) for Base20 (Top Panel), Base40 (Middle Panel), and Base60 (Bottom Panel) mice. Response rate functions were constructed by binning responses into bins equal to 12.5% of each FI and then dividing by the number of corresponding seconds. Response rates/second were then normalized by the max obtained response rate/second and time was normalized by FI length.

To quantify the impression that FI response functions are not scalar invariant, we asked whether parameter estimates of a scaled cumulative Weibull function were sensitive to the Base FI. Cumulative Weibull parameters were estimated in Python using a custom function in which we passed to an optimizer the both the raw and the normalized data with the following equation,

, (2)

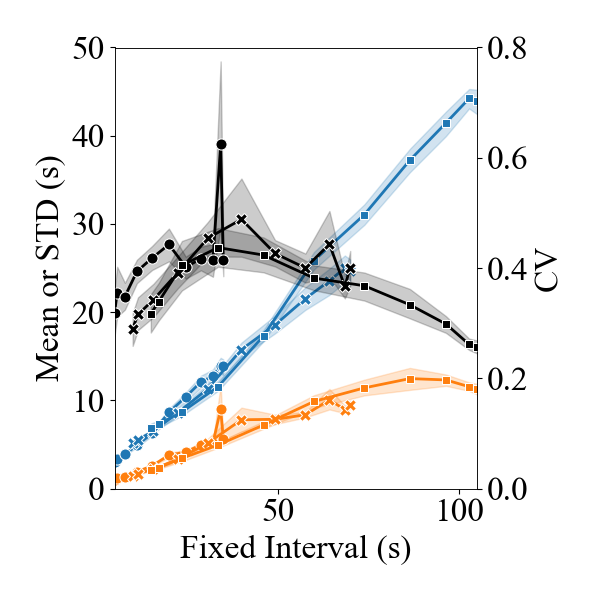
where A is the asymptote of responding, L is the midpoint, k characterizes the slope, and T indicates time into the interval. Log transformed parameter estimates (*A*,*L*,*k*) were analyzed using both NHST and Bayesian variants of a linear mixed effects model containing a random intercept across subjects and fixed effects of Base FI and FI.

Although the lack of superimposition of normalized response rates as a function of normalized time are compelling, it could be an artifact of the bin epoch used in parsing response rates. Indeed, the bin epoch is relatively small for short intervals (< 20 s) which comprise more of the Base 20 and Base 40 sinewaves than the Base 60 sinewaves making those response functions noisier and more difficult to interpret. To provide a higher quality assessment of the scalar property, we focused our analysis on BP distributions as a function of FI. Figure 4A-C shows the mean, standard deviation, and coefficient of variation plotted as a function of FI for each training group. These data suggests that while mean breakpoints scale linearly with the FI, standard deviations appear to scale linearly with a steep slope for FIs below the Base and then linearly again with a shallower slope for FIs above the Base (potentially approximating a power function), resulting in a CV that approximates an inverted-U.



**Figure 4.** Mean (left y-axis; blue), STD (left y-axis; orange), and CV (right y-axis; black) as a function of FI for Base 20 (Panel A), Base 40 (Panel B), and Base 60 mice (Panel C). Legend can be found in Panel A.

To confirm these impressions, we performed a model selection exercise in which we let the corrected Bayesian Information Criterion (BICc) determine the FI polynomial order needed to describe the mean, standard deviation and CV of breakpoints while accounting for group- and subject-specific effects. Specifically, with fit a mixed effects model with a random intercept and/or slope for each subject nested within each group and then tested which fixed effects were need to build the best fitting model: the FI polynomial order and Group. As shown in Table 3, after determining which random effects were supported by each moment of BPs, the SD supported a 3rd order polynomial, the CV supported a 2nd order polynomial, and the mean supported a 1st order polynomial. Interestingly, both the CV and mean supported adding a fixed effect of Group but this effect was only significant for the mean, not the CV. Taken together, these data support our impression that much of the deviation from the scalar property comes in BP moments that characterize variability.



**Figure 5.** Mean (left y-axis; blue), STD (left y-axis; orange), and CV (right y-axis; black) as a function of FI for Base 20 (circles), Base 40 (xs), and Base 60 mice (squares). Legend for colors can be found in Fig. 4A.

Table 3. Best Fitting Models for Each BP Moment as Determined by the corrected Bayesian Information Criterion (BICc)

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Mu\* | SD± | CV± |
| P1 | **1047.61** | 1077.01 | -398.03 |
| P1 + Group | **1038.49** | 1075.22 | -390.06 |
| P2 | 1022.211 | 987.39 | **-474.13** |
| P2 + Group | - | 986.56 | **-465.13** |
| P3 | - | **979.71** | -475.812 |
| P3 + Group | - | 978.62 | - |
| P4 | - | 977.072 | - |

*Note. The random effect structure of each model was determined prior to determining the fixed effects (FI polynomial order and Group) supported by the data. \*Indicates BP moments that supported the full random effects design with a random intercept and FI slope for subjects nested within groups. ±Indicates the BP moments that supported a random effect of intercept but not FI slope for subjects nested within groups. The stopping point in exercise, model either 1failed to converge or was 2not sufficiently different from the previous best fitting model to continue.*

**Trial-by-Trial Performance Reveals the CV is a Function of Location within and Range of the Temporal Sine Wave**

The previous analysis indicates that despite the mean breakpoint scaling linearly with time, neither raw (SD) or normalized variance (CV) estimates scale linearly. The SD appears to be best approximated by an 3rd order polynomial; in contrast, the CV appears to be best approximated by an 2nd order-polynomial. The visualization of the polynomial fits to the data suggest that dynamic timing performance is less variable around the trough and peak of temporal sinewaves and more variable as the sinewave ascends from the trough and descends from the peak. To test this hypothesis, we investigated mean and variance dynamics as a function of trial within the session. Figure 6A-C shows the mean, SD, and CV as a function of trial for each training group; and Figure 7 shows the same data as a function of trial within the sinewave averaged across sinewave periods. These data clearly show that while the mean and SD oscillate according to the sinewave, the CV takes on within each period of the sinewave a high-low-high-low function where high states correspond to the ascending and descending limbs of the sinewave and the low states correspond to the peak and troughs. Moreover, as the range from the smallest FI to the largest FI of the sinewave increases so does the difference between the high and low states of the CV.

To test these observations… [make Figure 7 then decide how to analyze]

Taken together these data confirm that dynamic interval timing is not scalar invariant.

**CV Is Sensitive to Change Rates**

Upon observation of the CV dynamics, we hypothesized that the CV may be sensitive to the rate at which FIs in the sinewave change such that when FIs are changing quickly (i.e., when ascending to the peak or descending to the trough) then the CV is relatively high, and when FIs are changing slowly (i.e., in the peak or in the trough) then the CV is relatively small. To test this hypothesis, we took the first differences of the FIs in the sinewave and asked whether those differences could be multiplied by a scale factor to fit the CV data.