

Leveraging synergistic intervals to enhance timing under multiple interval schedules

by

Robert Southern

B.S., Georgia Southern University, 2019

M.S., Georgia Southern University, 2021

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Psychological Sciences
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2025

Abstract

Accurate and precise timing is a fundamental aspect of many behaviors, skills, and occupations. This study tested whether structuring multiple intervals with simple multiplicative relationships enables participants to use rewards from shorter intervals as cues to improve timing on longer intervals. Across two online experiments ($N = 88$), participants tracked five independent fixed-interval schedules simultaneously. In Experiment 1, synergistic intervals were integer multiples of 5 s (5, 10, 20, 30, 40). In Experiment 2, nonsynergistic intervals (5, 13, 17, 21, 38) preserved the overall reinforcement rate but lacked simple multiplicative relationships. Across both experiments, two dominant strategies emerged: Steady Responders, who responded at a near-constant rate across intervals while largely disregarding programmed durations, and Interval Trackers, who showed graded sensitivity to interval length consistent with timing each interval. Synergistic interval structures did not reliably improve accuracy or precision, and scalar-like timing patterns were observed across both conditions and strategies. These findings suggest that in complex multi-interval environments, people often favor low-effort response strategies over exploiting embedded temporal cues, indicating that any benefits of synergistic intervals, if present, are subtle and difficult to detect.

Leveraging synergistic intervals to enhance timing under multiple interval schedules

by

Robert Southern

B.S., Georgia Southern University, 2019

M.S., Georgia Southern University, 2021

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Psychological Sciences
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2025

Approved by:

Michael Young
Major Professor

Copyright

© Robert Southern 2025.

Abstract

Accurate and precise timing is a fundamental aspect of many behaviors, skills, and occupations. This study tested whether structuring multiple intervals with simple multiplicative relationships enables participants to use rewards from shorter intervals as cues to improve timing on longer intervals. Across two online experiments ($N = 88$), participants tracked five independent fixed-interval schedules simultaneously. In Experiment 1, synergistic intervals were integer multiples of 5 s (5, 10, 20, 30, 40). In Experiment 2, nonsynergistic intervals (5, 13, 17, 21, 38) preserved the overall reinforcement rate but lacked simple multiplicative relationships. Across both experiments, two dominant strategies emerged: Steady Responders, who responded at a near-constant rate across intervals while largely disregarding programmed durations, and Interval Trackers, who showed graded sensitivity to interval length consistent with timing each interval. Synergistic interval structures did not reliably improve accuracy or precision, and scalar-like timing patterns were observed across both conditions and strategies. These findings suggest that in complex multi-interval environments, people often favor low-effort response strategies over exploiting embedded temporal cues, indicating that any benefits of synergistic intervals, if present, are subtle and difficult to detect.

Table of Contents

List of Figures	viii
List of Tables	ix
Acknowledgements	x
Dedication	xi
Chapter 1 - Introduction.....	1
Single Interval Timing.....	2
Multiple Interval Timing	5
Timing in Parallel	6
Intervals as Discriminative Stimuli.....	7
Mean Absolute Deviation	12
Hypotheses.....	13
Chapter 2 - Method	15
Experiment 1	15
Sample Size.....	15
Participants.....	15
Materials	15
Stimuli.....	16
Procedures.....	16
Experiment 2.....	19
Participants.....	19
Materials and Stimuli	19
Procedure	20
Determining Nonsynergistic Interval Lengths	20
Chapter 3 - Data Analysis	22
Chapter 4 - Results.....	25
Inter-response Time Analysis	26
Interval Length As Categorical Predictor	26
Interval Length As Continuous Predictor	27
Exploratory Analyses.....	30

K-means Clustering Analysis	30
K-means Clustering Results.....	31
Time To Completion Analysis.....	35
Mean Absolute Deviation	38
Chapter 5 - Discussion	42
Mental vs. Physical Effort	42
Exploration vs. Exploitation Within Interval Schedules	45
Alternate Forms of Optimality.....	46
Consistent Responding	46
Lack of Synergistic Benefits.....	48
Implications for Timing Theories	50
Implications for a General Audience	54
Chapter 6 - References.....	56
Appendix A - Supplemental Descriptive Statistics.....	60
Appendix B - Response Error Graphs	62
Appendix C - Model Assumption Checks	64
Appendix D - Return Plots.....	65
Appendix E - Supplemental Individual Subject Plots.....	69
Appendix F - Total Clicks Plot.....	70
Appendix G - Reward and Response Tables and Figure	71

List of Figures

Figure 1.1. Comparison of Scalar and Nonscalar Timing Errors	10
Figure 2.1 Example Visual Display	18
Figure 4.1. Non-Normal Response-Time Distributions Cannot be Attributed to Learning Effects	26
Figure 4.2 Model Estimated IRTs Across Discrete Interval Lengths	27
Figure 4.3 Results Show Sensitivity to Interval Length but No Meaningful Differences Between Conditions	29
Figure 4.4 K-means Clustering Solution Reveals Two Distinct Behavioral Response Patterns .	32
Figure 4.5 Individual Subject Data Support Distinct Response Profiles Identified by K-means Clustering.....	34
Figure 4.6 Neither Strategy Led Participants to Finish the Task Earlier	36
Figure 4.7 Frequency of Responses Made Throughout the Interval's Duration.....	37
Figure 5.1 Intermediate Rewards on Shorter Interval Schedules are Hypothesized to Reduce Timing Errors on Longer Schedules	52
Appendix Figure B.1 Raw Response Errors Show Distributions with Consistent Early Responses to Each Interval	62
Appendix Figure B.2 Standardized Response Error Plot Support Consistent Early Responding	63
Appendix Figure C.1 Posterior Predictive Checks From Each IRT Model.....	64
Appendix Figure D.1 Full Return Plots for Each Interval	65
Appendix Figure D.2 Steady Responder Return Plots.....	66
Appendix Figure D.3 Interval Tracker Return Plots.....	67
Appendix Figure E.1 Individual Participant Data Confirms Scalar Timing Properties in Interval Trackers Cluster	69
Appendix Figure F.1 Steady Responders Produced Significantly More Clicks Compared to Interval Trackers	70
Appendix Figure G.1 Mean Responses and Rewards by Interval Length, Cluster, and Condition	77

List of Tables

Table 1.1 MAD Values Across Simulated Conditions	13
Table 2.1 Experiment 1 (Synergistic) Interval Lengths.....	19
Table 2.2 Experiment 2 (Nonsynergistic) Interval Lengths.....	20
Table 4.1 Parameters Included in Clustering Analysis Showed Acceptable Degree of Between- subject Variation (SD).	31
Table 4.2 Measures of Scalar Timing using MAD by Condition and Cluster.....	40
Table 5.1 Total Clicks Varied Drastically by Cluster but not Condition.....	46
Appendix Table A.1 Supplemental Descriptive Statistics for Figure 4.1	60
Appendix Table G.1 Experiment 1 (Synergistic Intervals)	71
Appendix Table G.2 Experiment 2 (Nonsynergistic Intervals)	74

Acknowledgements

I would like to thank my family for their continuous love, support, and encouragement throughout this project and my time at Kansas State University. I am especially grateful to Dr. Michael Young for the valuable insights, generous contributions, thoughtful advice, and for sharing your passion and wisdom for data analysis and statistics, which have profoundly shaped my development as a researcher. I'd also like to thank Dr. Kimberly Kirkpatrick for her additional guidance and for providing the inspiration to study timing. I'd like to thank Dr. Travis Smith and Dr. Kelsey Panfil from the Reward, Timing and Decision Lab as well as Dr. Brian Howatt, Kari Payne, Patrick Hancock, and Luke Watson from the Dynamic Decision Making lab and all the undergraduate researchers for countless fond memories and experiences. Finally, thank you to my committee members, Dr. Gary Brase, Dr. Charles Pickens, and Dr. Suzanne Mitchell for their advice and guidance throughout the development of this project.

This research was made possible thanks to the Cognitive and Neurobiological Approaches to Plasticity (CNAP) Center at Kansas State University (NIH, P20GM113109). Data analysis for this project was performed on the Beocat Research Cluster (NSF grants CNS-1006860, EPS-1006860, EPS-0919443, ACI-1440548, CHE-1726332, and NIH grant P20GM113109) with support from CNAP's Neuroinformatics core. Tremendous thanks to Kari Payne, the Cognitive and Neuroscience Network for Experimentation and Computing Technologies (Connect) Lab, and CNAP's Cognitive Neuroscience core for experimental programming support for this project.

Dedication

For my parents, David and Julie. Thank you for the unabating support throughout this process and for the sacrifices you both made so I could pursue my dream. I could not have done this without you.

Chapter 1 - Introduction

Timing is a fundamental cognitive process that influences our perceptions, decisions, and behaviors. Despite the pervasive nature of timing, many questions remain about how we track the passage of time, especially if we must track multiple intervals simultaneously. Consider a chef during a busy shift. This individual must keep track of how long each item must cook, all while multiple items are cooking simultaneously. For example, the fries may take 2 min to cook, but each burger can be fully-grilled in one min. During peak hours, chefs often oversee many more items, meaning they must time even more intervals simultaneously without sacrificing accuracy or precision in their timing estimations.

This study examines whether participants can capitalize on simple multiplicative relationships among five fixed-interval (FI) schedules to boost timing accuracy and precision. In Experiment 1, each interval was an exact integer multiple of the shortest interval (FI-5), to test whether participants use reinforcements on the shorter intervals as temporal cues to improve timing on the longer intervals. In Experiment 2, we eliminated these multiplicative relationships—so that each interval must be timed independently—and determine whether timing variability returns to the hallmark scalar pattern commonly observed in timing behaviors. First, we provide a review of research involving single-interval timing before discussing research investigating timing under multiple-interval schedules. Finally, we discuss how rewards from shorter intervals can function as predictable cues to track longer intervals and review evidence from simulations of plausible timing strategies.

Single Interval Timing

One of the first empirical investigations exploring timing processes in non-human animals involved the use of FI schedules (Skinner, 1938). FI schedules reinforce the first response made after an interval elapses. Skinner observed that response rates increased as the end of the interval approached, reaching a peak at the end of the programmed interval, creating a “scalloping” pattern in responses. This increase in response rate as the end of the interval neared indicates the ability to estimate the passage of time.

Dews (1970) extended this work by testing animals' ability to time through a peak-interval procedure. In this procedure, subjects are first trained to respond after a fixed interval and later tested with non-reinforced ‘peak’ trials in which reinforcement is omitted. During peak trials, response rates typically increase to a maximum around the expected time of reinforcement before declining. This pattern allows researchers to track timing errors, which represent the difference between the anticipated and actual occurrence of an event. For example, if a subject responds to a 10-s interval after 7 s, the raw timing error for this response would be -3 s, since the response occurred 3 s early. Dews observed that while the absolute error increased with longer intervals, the errors remained proportional to the interval duration. This phenomenon, known as the scalar property of timing, laid the groundwork for Gibbon’s Scalar Expectancy Theory (SET).

Gibbon (1977) hypothesized that non-human animals (and by extension, humans) perceive the passage of time through an internal “clock”. According to SET, a pacemaker emits pulses that accumulate over time and are compared to a memory trace of previously reinforced intervals. When the number of accumulated pulses approximates the number of pulses stored in

memory, the organism infers that a similar amount of time has passed and, consequently, is likely to respond.

A critical aspect of SET is the inherent variability of the pacemaker. In this framework, the pulses generated by the pacemaker are modeled according to a Poisson distribution, such that the variance in the pulse count equals its mean. Therefore, as the interval length increases, both the mean and the variance of the pulse count increase proportionally. Although longer intervals yield higher pulse counts, they also include greater variability in these estimates, which addresses why longer intervals are more difficult to track precisely. The consistent ratio between the interval length and the standard deviation of timing errors (also referred to as the coefficient of variation) remains a fundamental feature of nearly all timing theories, suggesting that timing follows scalar properties consistent with Weber's law (see Bizo et al., 2006; Lejeune & Wearden, 2006, for notable exceptions).

In response to SET and other theories of timing that rely upon proposed internal clocks (i.e., attentional-gate models), Killeen and Fetterman (1988) proposed the Behavioral Theory of Timing (BeT). BeT suggests that time is represented by a series of ongoing behavioral states rather than by a single internal clock. As an organism engages in behavior, it naturally transitions through distinct states, and with training, these states become associated with specific time intervals so that the probability of a particular response is tied to the progression of these states over time. These behaviors often vary in length but have stochastic durations, similar to durations drawn from a distribution with a consistent mean and variance (Killeen & Fetterman, 1988). The variability in the duration of every behavioral state continues to compound as the interval increases and more behavioral states occur. When many states occur over a longer

interval, these individual errors add up proportionally to the overall duration, resulting in a consistent coefficient of variation—a hallmark of the scalar property of timing.

In short, both SET and BeT model scalar timing on single intervals, but these theories differ regarding how the passage of time is processed. Both theories rely on an internal pacemaker that emits pulses at a given rate. Unlike SET, BeT suggests this rate is governed by the overall rate of reinforcement (Bizo & White, 1994), predicting that the subjective passage of time should be sensitive to changes in the reinforcement rate, a finding which has been repeatedly supported (Bizo & White, 1994, 1995b, 1995a; Killeen & Fetterman, 1993; MacEwen & Killeen, 1991; Morgan et al., 1993).

Another interesting contrast between the two theories stems from how judgements on timing are made. According to SET, timing judgments must be made retrospectively. One must wait until the number of pulses in the accumulator matches the number of pulses in their memory before the decision to respond can be made. Conversely, BeT posits that a subject's behavioral transitions serve as discriminative stimuli for the status of the interval being timed. This enables the subject to anticipate how soon the interval will end, leading the subject to respond more rapidly (i.e., scallop) as the interval approaches the mean duration. Fetterman and Killeen (2010) demonstrated that pigeons produce more accurate timing when temporal cues were provided prospectively (during the interval) versus retrospectively (during the choice phase) in a temporal discrimination task. Despite both theories predicting scalar properties of timing, BeT provides a more complete account for single-interval timing behaviors compared to SET.

Multiple Interval Timing

Mixed fixed-interval (FI) schedules provide a unique opportunity to distinguish between the competing predictions of SET and BeT. Specifically, SET predicts that coefficient of variation remains constant regardless of interval length or the number of intervals being simultaneously timed. In contrast, BeT proposes that timed behaviors (such as lever pressing or key pecking) arise from a sequence of adjunctive behaviors, where transitions between these behavioral states serve as discriminative stimuli signaling elapsed time. According to BeT, these state transitions are regulated by pulses from an internal Poisson pacemaker which pulses at a fixed rate based on the overall reinforcement rate (Bizo & White, 1994; Morgan et al., 1993).

Importantly, both SET and BeT predict scalar timing under conditions where only a single interval is tracked. However, when multiple intervals are timed simultaneously, BeT predicts that the number of adjunctive behavioral states scale with the interval length—longer intervals should elicit more adjunctive behaviors. However, because the pacemaker rate remains anchored to the overall reinforcement rate (rather than adjusting individually for each interval), BeT predicts that timing across multiple intervals will not exhibit scalar properties. In other words, unlike SET, BeT expects the coefficient of variation to differ across intervals of varying durations when tracked simultaneously.

To further these predictions from SET and BeT, Leak and Gibbon (1995; Experiment 2) tested pigeons' ability to time two intervals simultaneously using mixed FI schedules. A mixed FI schedule combining an FI-40 and an FI-240 means reinforcement is available after either 40 s or 240 s have elapsed since the last reinforcer, depending on which interval is active. Importantly, no discriminative stimuli were provided to indicate which of the intervals was activated. On any given trial, subjects must first time the 40-s interval, resulting in behavioral

response patterns similar to those of peak-interval trials. If no reinforcement is delivered, the subject must continue to time the larger interval and make a response after 240 seconds to receive reinforcement.

Using mixed schedules involving two or three FI schedules, Leak and Gibbon (1995) confirmed SET's predictions. They showed that the relative variability of timing errors remained constant among the tracked intervals, providing clear evidence that scalar properties are preserved when tracking two intervals of different lengths. Although Leak and Gibbon (1995) provided important evidence regarding scalar properties across two intervals, the intervals were always sequentially nested rather than truly parallel. This limitation leaves open the question of how scalar properties may generalize when two or more intervals begin simultaneously but must be independently maintained as required by the hypothetical chef.

Timing in Parallel

Sequential timing involves timing intervals one after the other, with the completion of one timing interval triggering the start of the next. On the other hand, parallel timing requires independently tracking multiple intervals without relying on cues from completed intervals.

Consider again the example of a fast-food grill chef. Suppose the chef must start cooking French fries, which require two min, while simultaneously grilling a burger patty, which takes one min. Rather than independently timing each item in parallel, the chef might simplify the amount of effort required by synchronizing the times each item begins to cook. Rather than having to track the cooking times for each item independently, the chef only needs to track a single interval sequentially (one min for French fries followed by one additional min for the burger). By subtracting the shorter interval from the longer one, the chef can preserve precise

(nonscalar) timing without having to track each interval independently. But is it possible to time two or more intervals at the same time when the intervals are truly independent?

Rule and Curtis (1985) demonstrated that human participants can reliably produce the average duration of two independent intervals presented in parallel suggesting humans can reliably track multiple intervals if all of the intervals have synchronized starting times. van Rijn and Taatgen (2008) expanded this research to investigate how humans time multiple overlapping intervals without synchronized starting times. Participants estimated two intervals, initially both 2 s long, before transitioning to 2 s and 3 s for Interval 1 and 2, respectively, during a follow-up session. These intervals were offset via a delay (referred to as the stimulus offset asynchrony [SOA]), which created an overlap between the two intervals. Results showed that estimates for the second interval, but not the first, were positively correlated with the SOA.

Rather than using separate internal clocks to track each interval, their results suggest humans rely on the strategic use of a sole internal clock to infer the status of multiple intervals simultaneously. This leaves one key question unanswered: if humans use one shared clock strategically, then what types of strategies or heuristics can be used to accurately and precisely time multiple intervals occurring simultaneously?

Intervals as Discriminative Stimuli

Revisiting our previous example, our chef doesn't have to track the cooking times for each item individually. Instead, he can use shorter cooking times to infer how much time has passed on longer intervals. If the fries begin cooking at the same time as the burger, the chef only needs to time a single 1-min interval twice—once to grill the burger and once more to finish cooking the fries. By manipulating the longer 2-min interval into two sequential intervals of 60 s,

the chef has turned a slightly more difficult task (tracking a 2-min interval) into a somewhat less challenging task (tracking a single 1-min interval twice).

But consider what happens when the chef is asked to cook one order of burgers and fries for two separate customers. Now the chef is asked to track four independent intervals (one interval per item ordered). But our chef is experienced and knows that time can be saved if both orders are cooked in parallel. The total number of intervals the chef needs to track has increased by two, but the cognitive effort required has not changed—the grill takes 1 min to cook a burger, no matter how many burgers are present.

How these intervals are structured (e.g., their scaffolding or the relative relationships between each of the intervals) can produce predictive cues for the remaining intervals through reinforcement delivery, provided the subject can accurately time the shorter interval(s). By restricting each interval's duration to be a multiple of the shortest interval, we create “synergistic” interval schedules, whereby learning the duration of one schedule (e.g., FI-5) enhances the ability to track longer intervals (e.g., FI-20) through repeated, predictable reinforcement. In synergistic interval schedules, reinforcement from shorter intervals provides two key pieces of feedback to the subject: how much time has passed and how much time remains on a longer interval (Freestone & Church, 2010).

Frequent and clear feedback improves timing estimates (Riemer et al., 2019; Bader & Wiener, 2021), but this feedback can take many different forms. Descriptive feedback specifies the magnitude and direction of error (e.g., “You were 3 s early”), providing precise feedback which is maximally informative. In contrast, FI schedules provide relative feedback through reinforcement delivery. On an FI 5-s schedule, reinforcement delivery signals only that at least 5

s have elapsed since the last reinforced response, whereas an unreinforced response indicates that 5 s have not yet passed. Either way, reinforcers from FI schedules provide no additional information about how precise a timing estimate was.

Synergistic interval schedules have the unique property that feedback provided from shorter intervals facilitates accurate and precise timing on longer intervals. Conversely, multiple interval schedules lack synergy when the durations of each interval cannot be easily leveraged to facilitate precise timing of longer intervals. When tracking two intervals that do not share a simple multiplicative relationship (e.g., FI-7 vs. FI-18), the amount of information signaled by reinforcement from the FI-7 schedule decreases. Reinforcers provided from the shorter interval continue to provide *some* information about the status of the longer interval, but this fractional ratio (1:2.6) means more cognitive effort is required to precisely time the longer interval.

Experiment 1 of the present study examined how humans time multiple intervals simultaneously using a novel paradigm with synergistic interval schedules structured to facilitate the acquisition of each independent schedule. By strategically employing synergized intervals, we predict that participants would learn each of the schedules independently, as evidenced by high accuracy and precision non-scalar timing across all interval lengths.

When comparing shorter versus longer intervals, SET predicts that scalar properties will be preserved, such that the coefficient of variation remains proportional to the interval length (Gibbon, 1977). However, participants may use the shorter, nested intervals to artificially inflate their ability to track longer intervals. If asked to time both an FI-5 and an FI-10 simultaneously, one can infer that the FI-10 will have reinforcement available (i.e., is primed) if two rewards from the FI-5 have been collected previously. By reliably timing shorter intervals, participants

may show timing behaviors that are inconsistent with the predictions of scalar timing.

Specifically, if participants are using shorter intervals as predictive cues for longer intervals, the timing errors observed on longer intervals should be more precise than scalar timing would predict.

Figure 1.1. Comparison of Scalar and Nonscalar Timing Errors

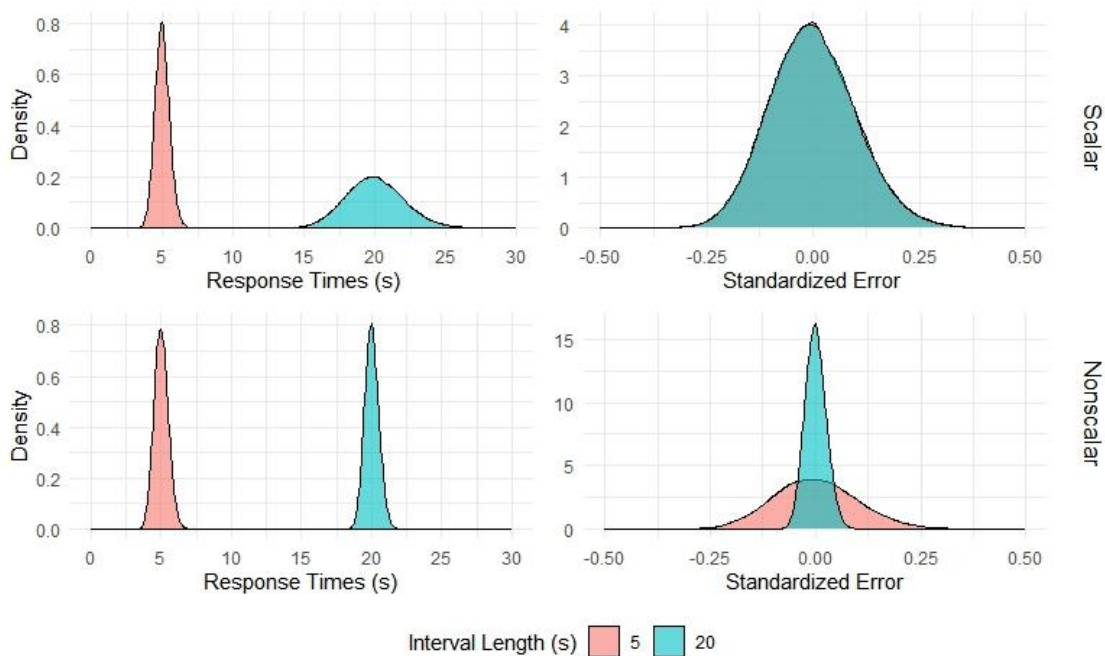


Figure 1.1 compares scalar and one example of nonscalar timing using simulated responses across two interval lengths (5 s and 20 s). The top panels depict scalar timing, where variability increases proportionally with interval length. Here, the spread of the 20-s interval is visibly wider in the raw scale, but when standardized (top right), both intervals show similar error distributions. The bottom row shows results from nonscalar timing, where variability remains constant across interval lengths. In the original scale (bottom left), the distributions for both intervals are nearly identical, and in the standardized scale (bottom right), the shorter

interval shows relatively larger error. This illustrates the core property of scalar timing: proportional variability that preserves the coefficient of variation across durations.

In Experiment 1, we predicted that synergistic interval lengths (e.g., FI-5 and FI-10) improve participants' ability to time all intervals. We also predict that participants on longer intervals would result in absolute timing errors comparable to the absolute timing errors on shorter intervals, directly opposing the predictions made by SET (Allan, 1998; Gibbon, 1977; Leak & Gibbon, 1995). Instead, if participants are using the shorter intervals to predict the status of longer intervals, the intermediate responding on shorter intervals (FI-5) acts as adjunctive behavioral states that are used to estimate how much time has passed on longer intervals (Haight & Killeen, 1991). Therefore, we expect participants to show non-scalar timing errors, consistent with the predictions from BeT, rather than scalar timing properties predicted by SET.

In Experiment 2, we modified the experimental design by manipulating the predictive relationships between intervals, thus making it more challenging for participants to independently time each interval. For example, two identical FI-5 schedules would be maximally predictive of one another, assuming the intervals began at the same moment (van Rijn & Taatgen, 2008). Once a response is reinforced on one schedule, a participant can be confident that the next response to the alternative schedule will provide reinforcement as well. However, by removing the synergistic relationships between the intervals, participants may shift away from the cognitive heuristics or strategies utilized in Experiment 1.

More specifically, if participants are asked to track multiple intervals simultaneously, and the status of the shorter intervals can be used to infer the state of longer intervals, then we expect participants' timing errors to be nonscalar (Experiment 1). However, if the intervals lack synergy, such that the status of one interval cannot be easily inferred based on the status of another, then

we expect timing errors to be scalar, consistent with participants tracking each interval independently (Experiment 2).

Importantly, the overall rate of reinforcement did not change between Experiments 1 and 2. BeT predicts that the pulse rate of the internal pacemaker is dependent upon the overall rate of reinforcement. By controlling for the overall rate of reinforcement, we can isolate the effects of the synergistic intervals and measure how timing accuracy and precision change under synergistic vs. non-synergistic reinforcement schedules.

Mean Absolute Deviation

In traditional timing research, peak trials are included to measure when a response occurs during the interval's full duration. The resulting distributions are approximately normal, and the characteristics of these distributions are used to assess timing accuracy (mean) and precision (variability). We elected not to include peak trials in the current study for both methodological and practical reasons. Peak-trials are time-intensive as each trial must look at a single interval in isolation for much longer than the programmed interval's duration to capture the full response distribution. Given that five intervals were included and each interval would need to be probed multiple times to capture variability, the data we would obtain from including peak-trials would be quite limited. In addition, the entire experiment was programmed to be completed within 45 minutes. Therefore, even including the minimum number of peak trials necessary would sacrifice a substantial amount of data that could be obtained from the main task. Given these constraints, we prioritized continuous performance in the primary task rather than sacrificing large portions of session time for a small set of probe trials.

Given that peak trials were omitted, the resulting response distributions are positively skewed, making traditional metrics inadequate. Instead, we calculate the mean absolute deviation

(MAD) as a measure of how much variability is present in the timing responses. We used this measure of variability to assess whether scalar timing properties were observed as the interval length increased.

MAD is a reliable method to estimate variability in non-normal distributions while providing point estimates that can be compared across participants. Note that scalar timing would predict MADs that are approximately 4x larger for the 20-s interval than the 5-s interval, because the 20-s interval is 4x longer. Meanwhile, nonscalar timing with equal variance would predict that these MAD values are equivalent regardless of interval length when the underlying distribution is normal. Due to the positive skew from gamma distributions at shorter interval lengths, nonscalar timing with unequal variance is evidenced by MAD values that are approximately twice as large for the 20-s compared to the 5-s interval. Table 1.1 shows the MAD values for each condition.

Table 1.1 MAD Values Across Simulated Conditions

Timing Error	Interval Length (s)	Mean Absolute Deviation
Scalar	5	0.49
	20	1.55
Nonscalar	5	0.41
	20	0.81

Hypotheses

In Experiment 1, we hypothesized that, under synergistic interval schedules, participants would exhibit precise, nonscalar timing—specifically, the mean absolute deviations of timing errors for the 20-s interval would be 1.5 ($\pm .1$) times larger than those for the 10-s and 2 ($\pm .1$) times smaller than 5-s intervals:

$$\text{Hypothesis 1A:} \quad MAD_{20s} = 1.5(MAD_{10s}) = 2(MAD_{5s})$$

Alternatively, SET predicts that the MAD of timing errors should scale proportionally with interval length regardless of schedule synergy:

$$\textit{Hypothesis 1B:} \quad MAD_{20s} = 2(MAD_{10s}) = 4(MAD_{5s})$$

In Experiment 2, we predicted that non-synergistic conditions would evoke scalar timing properties, such that the MADs of timing errors should increase in direct proportion to interval length, under conditions where $i < j$ with i and j denoting intervals of differing durations.

$$\textit{Hypothesis 2A:} \quad MAD_j = \frac{j}{i} MAD_i$$

However, by increasing the overall task complexity and/or by adopting suboptimal strategies, participants may show evidence of timing errors larger than those predicted by scalar properties:

$$\textit{Hypothesis 2B:} \quad MAD_j > \frac{j}{i} MAD_i$$

Chapter 2 - Method

Experiment 1

Sample Size

To our knowledge, no previous study has employed a similar design to measure human performance across multiple fixed-interval schedules with synergistic intervals. Therefore, no suitable effect-size estimates were available. Given the within-subject design of the experiments and the large number of observations being obtained from each participant, 45 participants were expected to provide sufficient power to detect any small-to-medium effects. This sample was well-powered to detect the smallest effect size of interest of Cohen's $d = 0.1$, with 80% power, and a significance threshold set at $\alpha = .05$ (Götz et al., 2022; Lakens, 2022).

Participants

A total of 45 U.S. adults ($M_{\text{age}} = 39$, $SD_{\text{age}} = 11$, 21 men, 23 women, 1 non-binary) from CloudResearch's Connect platform participated in an experiment we expected to take 45 minutes. Participants were paid \$7.25. One participant's demographic data, and five participant's behavioral data was missing, leaving a final analyzable sample of $N = 40$. All participants provided informed consent before the experiment began. Participants were assigned a unique, randomly generated ID number that was used to link their Qualtrics data to their behavioral data.

Materials

The experiment was built via the builder function in PsychoPy (Peirce et al., 2019; version 2024.1.2) and was hosted on Pavlovia.org (Open Science Tools, Nottingham, UK). Qualtrics (Provo, UT) was used to obtain informed consent, provide a debriefing, collect demographic data, and provide the link to the online behavioral task.

Stimuli

The behavioral task was completed in a window from the participant's internet browser. The relative sizes of the window and stimuli may have differed depending on participants' display settings. Five equally sized stimuli were arranged equidistantly from the window's midpoint (see Figure 2.1). Five of nine possible stimuli (fruits) were randomly selected at the beginning of the experiment and remained fixed for the experiment's duration. Each fruit's location on the screen and interval length were randomized across participants to minimize any bias introduced by certain arrangements.

Procedures

The median completion time for Experiment 1 was 30 mins (minimum = 23, maximum = 66). Participants opted-in to participate by selecting the experiment from the list of available studies on Connect. Upon enrollment, participants were directed to Qualtrics, gave informed consent, and were provided with a link to access the behavioral task with the following instructions:

“Welcome to the experiment!

We will be investigating how individuals track time under different circumstances.

In this task, you will see 5 fruits arranged in a pattern. Each fruit provides rewards at different rates. Clicking on a fruit earns points. However, those points are available only after time has passed since the same fruit was clicked. The length of time varies across fruits.

Important: To ensure that your natural time tracking ability is measured accurately, please do not use any external aids such as a watch, timer, or smartphone, and refrain from counting in your head.

If you use any external timekeeping aids or counting strategies, your participation will be invalidated.

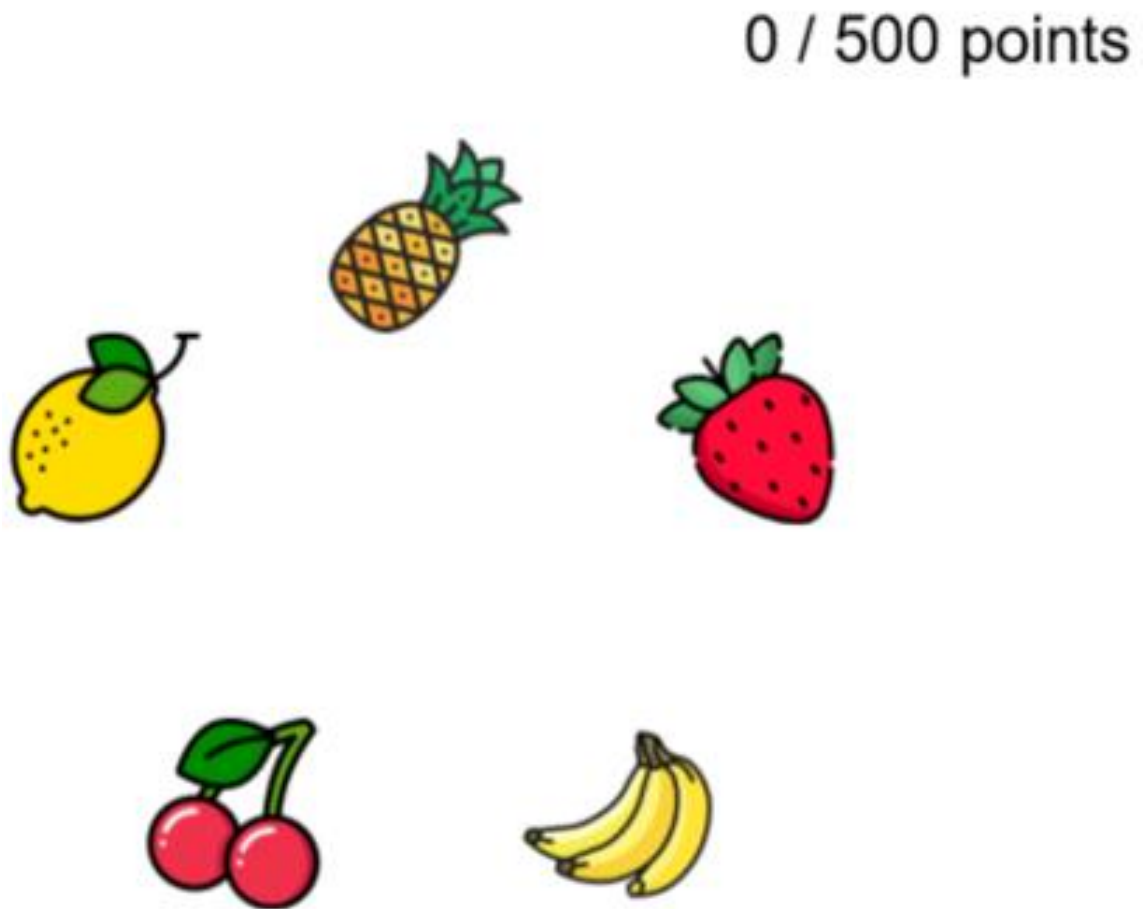
Your goal is to earn as many points as possible.

The experiment will end once you reach 500 points.

Press SPACE to begin the experiment.”

Once the experiment was completed, the participants were told to return to Qualtrics to provide demographic data. After the behavioral task was completed, and all demographic questions were answered, participants were directed to the completion page and were compensated for their participation. Figure 2.1 shows how the stimuli were presented to participants at the beginning of the experiment. Each stimuli implements an independent FI schedule with the interval length varying across fruits. Using a computer mouse or trackpad, participants clicked on each option to earn points.

Figure 2.1 Example Visual Display



Once the experiment began, the first response to each option resulted in reinforcement delivery and began the interval. Responses made before the interval were recorded but not reinforced. Unreinforced responses triggered a brief aversive audio cue to signal that the interval has not elapsed. The first response made after the interval is completed resulted in a reward and the interval was reset. Rewarded responses triggered a separate pleasant audio cue alongside a 1-s animation indicating the participant has earned one point. Over 45 min, the participants were asked to track all five FI schedules independently to maximize the number of rewards obtained. Payoff structures were selected to sample a wide range of interval lengths (Table 2.1).

Table 2.1 Experiment 1 (Synergistic) Interval Lengths

Option	Interval Length (s)	Reinforcement Ratio
1	5	.200
2	10	.100
3	20	.050
4	30	.033
5	40	.025
Overall		.408

Experiment 2

Experiment 1 leveraged synergistic reinforcement schedules to aid acquisition of the interval lengths, assist in tracking each interval's duration, and to determine whether scalar timing properties were preserved under synergistic schedules. Experiment 2 examined acquisition and timing performance under similar reinforcement schedules that lack synergy. Experiment 2 was run concurrently to Experiment 1 to ensure no selection biases were present during enrollment. Participants were not randomly assigned to either experiment, but participants were only eligible to complete one experiment based on Connects' exclusionary criteria.

Participants

Due to the increased complexity and anticipated difficulty of the timing intervals in Experiment 2, we enrolled a larger sample to ensure we could detect effects comparable to those observed in Experiment 1. We enrolled 60 U.S. adults ($M_{\text{age}} = 39$, $SD_{\text{age}} = 12$, 33 men, 26 women, 1 non-binary), but only received data from 48 participants, leaving a final analyzable sample of $N = 48$.

Materials and Stimuli

All materials and stimuli in Experiment 2 were identical to those used in Experiment 1, with only the interval lengths differing.

Procedure

The median completion time for Experiment 2 was 30 mins (minimum = 23, maximum = 120). The overall timeline, experimental layout, and task characteristics were identical to Experiment 1. The key exception is that the relative relationships between the five independent schedules were modified to reduce the potential for participants to infer the status of longer intervals based on the number of rewards obtained from the shorter intervals. Table 2.2 shows the final set of interval lengths for Experiment 2.

Table 2.2 Experiment 2 (Nonsynergistic) Interval Lengths

Option	Interval Length (s)	Reinforcement Ratio
1	5	.200
2	13	.077
3	17	.059
4	21	.048
5	38	.026
Overall		.410

Determining Nonsynergistic Interval Lengths

We wanted to eliminate any synergistic relationships between each reinforcement schedule, by minimizing the predictability of reinforcement from shorter intervals which could be used to cue when to respond to the longer intervals. We simulated over 10,000 combinations of five randomly-generated unique delay values as candidate interval sets. Then we eliminated sets that did not conform to three key criteria.

First, we wanted to make sure that the overall reinforcement ratio was constant between Experiments 1 and 2 (see Tables 2.1 and 2.2, respectively). This step was necessary to ensure the predictions of BeT are consistent and transferable between experiments because the rate of the pacemaker, BeT claims, is tied directly to the overall rate of reinforcement (Bizo & White, 1995b; Killeen & Fetterman, 1988).

The second criterion was that no interval's duration may be a multiple of another interval as this property is a defining characteristic of synergistic interval schedules. While this requirement did not eliminate the predictiveness of the shorter schedules—one can still infer some information about the status of a 17-s interval by counting how many 5-s intervals have elapsed—it does greatly reduce the amount of information obtained, while requiring more cognitive effort to accurately time each interval.

The third criterion was that in situations where two interval values are near multiples of one another, we preferentially selected for options where the larger interval was greater than the multiple (e.g., 7 s and 15 s), rather than options which fell below the multiple (e.g., 7 s and 13 s). We included this criterion to discourage participants from simply adopting a strategy that is less cognitively taxing while accepting a small amount of error. In the example of two intervals—7 s and 13 s—participants may simply treat the 13-s interval as if it were 14 s, functionally modifying these two schedules into synergistic schedules. Therefore, we opted to select schedules where near multiples were larger to provide an additional cost (decreasing the overall rate of reinforcement) to reduce the likelihood of adopting a similar strategy.

Finally, we eliminated any candidate sets that did not meet all three criteria and settled on a set of 5 delay values (see Table 2.2) that had an equivalent overall rate of reinforcement and no direct multiples. These five values were used as interval lengths in Experiment 2.

Chapter 3 - Data Analysis

Traditional timing studies assess timing accuracy (peak) and precision (spread) of response-times within an interval. Because responding ceases once a reward is obtained on a FI schedule, probe trials are occasionally intermixed where reinforcement is omitted. This technique allows researchers to investigate the full shape of the response distributions, rather than only the pre-reinforcement portion which occurs up until the first response is made after the interval is primed. The resulting response distributions from peak-trials are often approximately normal and show increased variability for longer intervals (scalar timing). In our design, these traditional distribution-based methods are not appropriate as we only obtained response data from the pre-reinforcement portion of the distribution.

Instead, we analyzed the inter-response times (IRTs), operationalized as the duration between the last reinforced response on a given interval and the next response to that same interval (i.e., post-reinforcement pauses). Very short IRTs suggest either misestimation of the interval's duration, or a strategy not based on the FI duration. Very long IRTs most likely represent a lack of attention from the participants. Therefore, any IRTs beyond 80-s (twice the longest interval length) were removed. This criterion removed fewer than 0.5% of all IRTs and these exclusions were evenly distributed across experiments minimizing the risk of bias from differential attrition.

We analyzed IRT data using two separate approaches. The first approach treated Interval Length as a categorical predictor, while our second approach treated Interval Length as continuous. We initially modeled IRTs with categorical interval lengths because we presumed participants might have distinct response patterns for each interval length. By treating Interval Length as categorical, our models do not presume that the spacing among each interval is

equivalent, allowing us to more precisely measure nonlinear changes in the IRTs as interval lengths increase. Therefore, we analyzed data from Experiment 1 (the synergistic design) and 2 (the nonsynergistic design) separately because Interval Lengths were nested within Experiment. In contrast, we analyzed the combined data from Experiments 1 and 2 when we treated interval length as continuous thus allowing us to determine if the sensitivity to interval length varied between experiments.

For Experiment 1, two full factorial multilevel gamma regression models with log-link functions were tested and AIC criterion was used to determine the best fitting model. Model 1 included Time (mean-centered; log+1 transformed), Interval Length (effect-coded), and their interaction as fixed effects, with random intercepts for Participant and the Participant-by-Interval Length interaction, as well as random slopes for Time at both levels. The simpler model, Model 2, shares the same fixed effects but incorporates only random intercepts for Participant and the Participant-by-Interval Length interaction without random slopes. Model selection for Experiment 2 proceeded in a similar fashion.

In our second approach, we analyzed IRT data using a full factorial Bayesian multilevel gamma regression model with a log-link function that included Time (mean-centered; log+1 transformed), Interval Length (mean-centered; divided by 10 to aid convergence) and Condition (effect-coded) as fixed effects, with a maximal random effect structure (Barr et al., 2013) including Participant as the random intercept, with Time, Interval Length, and their interaction as random slopes. Weakly-informed priors ($M = 0$, $SD = 1$) were applied to all fixed effects. The model was estimated using 4 chains (4,000 iterations; 1,000 warmup iterations per chain).

For Bayesian analyses, p-values are not provided unlike traditional null hypothesis significance testing (NHST) models. Therefore, we treated any variable which did not contain 0

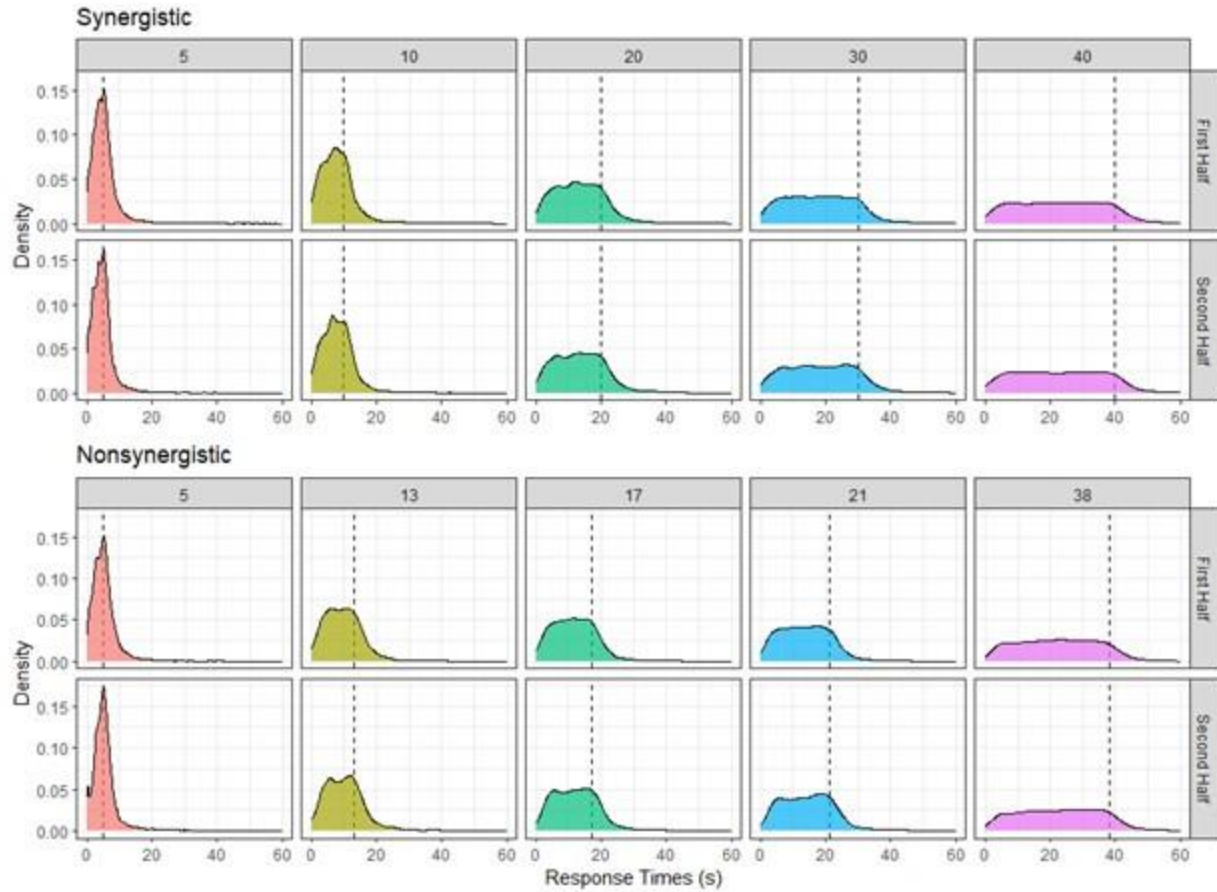
within the 95% credible interval as providing strong evidence of an actual difference—analogue to NHST models where $p < .05$.

Additionally, Bayesian analyses allowed us to treat Interval Length as a continuous predictor, rather than categorical, providing two key benefits. First, we used the estimated behaviors from Experiment 1 to help inform predictions for Experiment 2, and vice versa. This additional flexibility was used to inform estimates for participants behavior on interval lengths that weren't experienced by that individual. In short, even though a synergistic participant (Experiment 1) never experienced a 13-s interval, we were able to estimate their IRT for a 13-s interval using the observed data from both experiments. Second, the additional flexibility also allows us to make between-experiment comparisons, despite the nested interval structures. Therefore, we opted to treat Interval Length as a continuous predictor, as well as introducing Condition (i.e., Experiment) as a between-subject predictor in the overall model to assess for differences across Experiment 1 and 2. Because the two experiments were run concurrently, treating Experiment as Condition will not give rise to selection bias concerns.

Chapter 4 - Results

We began by investigating the response times within the full interval duration. These non-normal response distributions did not change when we compared the responses from the first half of each experiment (0-250 rewards) to responses from the second half (251-500 rewards; see Appendix Table A.1 for more details). These data suggest that the non-normal response distributions are not due to the participants learning the durations of each interval. Instead, we see that an insensitivity to the interval lengths (in which behavior often levels off well before the target key is baited) remains present throughout the duration of both experiments. Figure 4.1 shows the overall response distributions for each experiment. For the raw response error distributions and the standardized distributions, see Appendix Figures B.1 and B.2, respectively.

Figure 4.1. Non-Normal Response-Time Distributions Cannot be Attributed to Learning Effects



Note. The halfway point for both experiments occurs after receiving 250 awards.

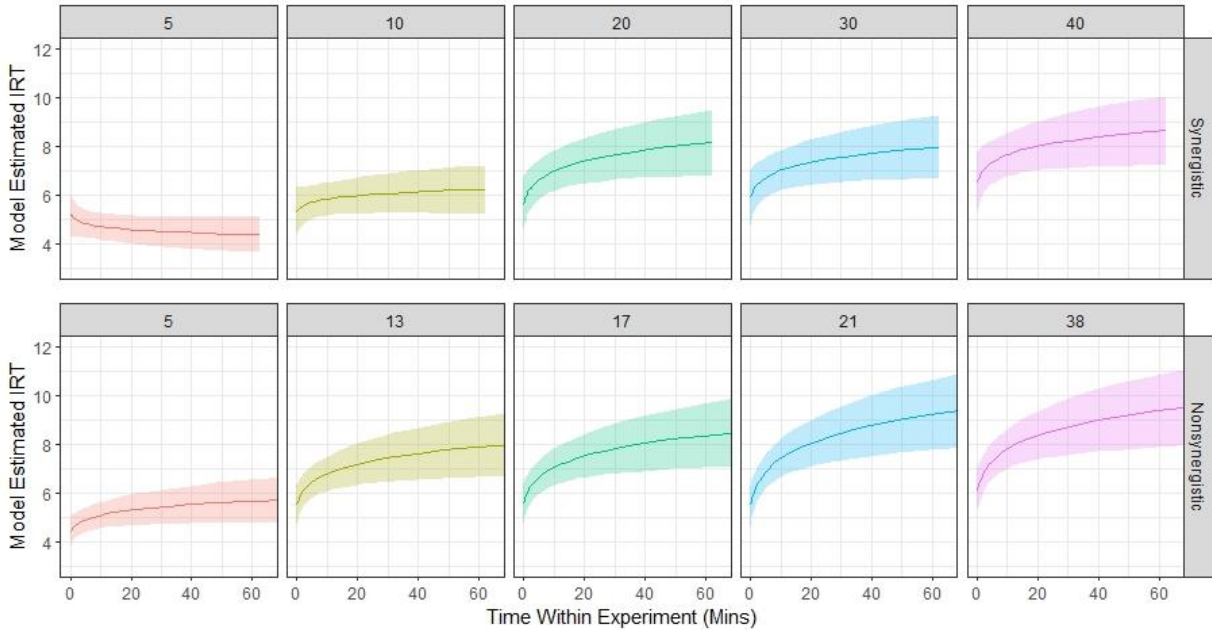
Inter-response Time Analysis

Interval Length As Categorical Predictor

Model 1 was the best fitting model for both Experiment 1 ($\Delta\text{AIC} = 2260$), and Experiment 2 ($\Delta\text{AIC} = 2911$), compared to the simpler models. Upon further investigation, underdispersion was detected in Experiment 1 (dispersion ratio: 0.35, $p = 0.016$) but not Experiment 2 (dispersion ratio: 0.51, $p = 0.152$). Posterior predictive checks for both models showed moderate divergences from the observed data, suggesting that the selected models

struggled to produce an accurate fit (see Appendix Figure C.1). Figure 4.2 shows the results from Experiment 1 (top) and Experiment 2 (bottom).

Figure 4.2 Model Estimated IRTs Across Discrete Interval Lengths



Note. Error bands represent ± 1 standard error.

Although these models provided reasonable estimations of participant behavior, model diagnostics for the best-fitting models in Experiments 1 and 2, along with mild to moderate deviations observed in posterior predictive checks, indicated some violations of model assumptions. To address these issues, we employed our second approach using Bayesian modeling to provide greater flexibility for handling these deviations.

Interval Length As Continuous Predictor

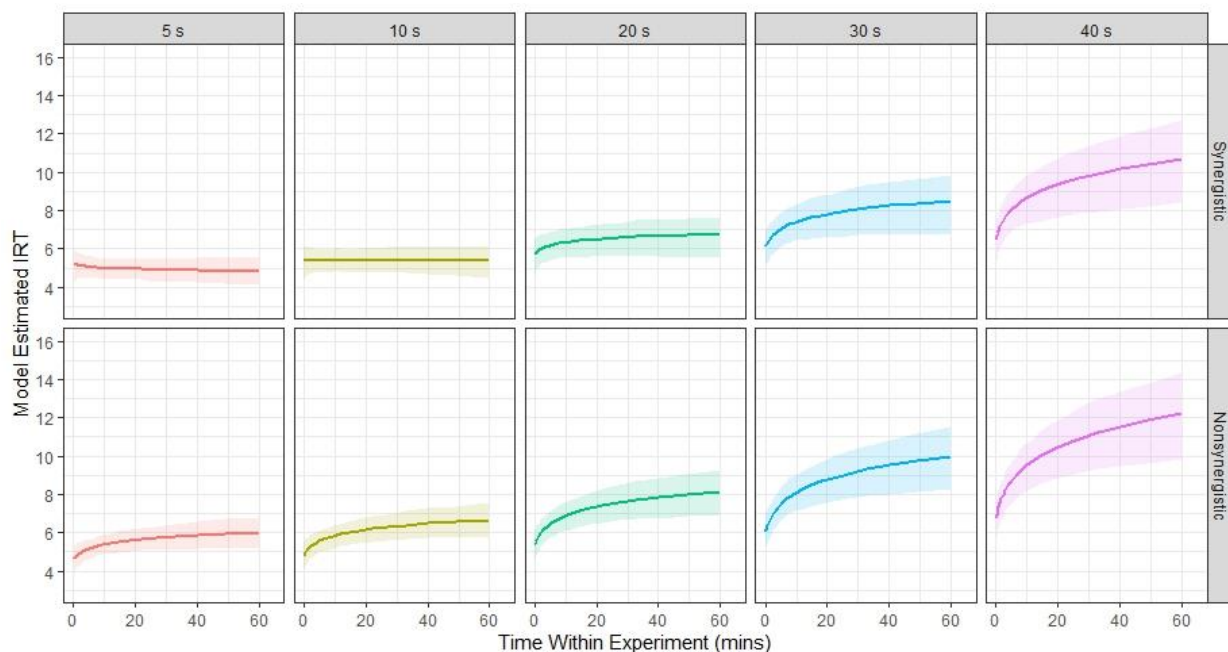
The results from the Bayesian models largely matched the main conclusions from the previous modelling approach. Model diagnostic metrics were superior, with $R_{hats} \leq 1.01$ and posterior predictive checks indicated that any prior deviations are better captured when using a Bayesian framework with Interval Length treated continuously.

Our results suggest that there was no substantive evidence of a main effect of time ($\beta_{\text{Time}} = 0.05$, 95% CrI = [-0.03, 0.13]), indicating that no major change to participants' reaction times occurred as the experiments progressed, when all other variables were held constant at their means. We did observe differences across Interval Lengths ($\beta_{\text{IL}} = 0.17$, 95% CrI = [0.13, 0.20]), such that participants' estimated IRTs increased as the interval length increased.

We also observed a Time $\times$ Interval Length interaction ($\beta_{\text{Time} \times \text{IL}} = 0.03$, 95%CrI = [0.01, 0.05]), suggesting that IRTs increased at a higher rate for longer intervals. To accurately represent the interaction between Time and Interval Length, we extracted model-based predictions from five evenly-spaced points along the Interval Length continuum. We chose these values because they sampled the full range of the observed values. These estimates are plotted using the posterior predictions which incorporate data from *all participants* regardless of Condition. Figure 4.3 shows participants increase their IRTs as Interval Length increases, but fall well below the true value for the longer intervals

Finally, we observed no actual differences in Mean IRTs across the two conditions ($\beta_{\text{Cond}} = -0.05$, CrI = [-0.20, 0.10]). Conditions also did not differ in their sensitivity across Time ($\beta_{\text{Cond} \times \text{Time}} = -0.04$, CrI = [-0.11, 0.04]), Interval Length ($\beta_{\text{Cond} \times \text{IL}} = 0.00$, CrI = [-0.04, 0.03]), or in their Time $\times$ Interval Length interaction ($\beta_{\text{Cond} \times \text{Time} \times \text{IL}} = 0.01$, CrI = [-0.01, 0.03]). Therefore, the evidence suggests that the interval's structure (synergistic vs. nonsynergistic) had no noticeable impact on participants' IRTs.

Figure 4.3 Results Show Sensitivity to Interval Length but No Meaningful Differences Between Conditions



Note. Error bands represent 68% credible intervals.

The results from both methods of analyzing IRTs revealed an unanticipated result. Participants seemed to show weak sensitivity to interval length and strong underestimation of the longer durations, regardless of experimental design (synergistic vs. nonsynergistic). Instead, participants appear to be tracking much shorter intervals—or pressing each option at a rate consistent with shorter intervals—with average IRTs around ~10 s for any interval beyond 10 s in duration.

To further probe these results, we began some exploratory investigations by examining the individual participants' data. When we plotted the model predictions (e.g., the fitted values) from the Bayesian model for each participant, we noticed two key details. First, our 88 participants showed multiple types of response patterns (e.g., strategies). Second, one of these strategies involved steady, consistent responding to every interval. It appeared that many of our participants found that steady, consistent responding to each interval was less effortful than

tracking the durations of each interval independently and responding only when the end of the interval approached (our anticipated response pattern based on timing experiments involving fewer fixed intervals; Leak & Gibbon, 1995). To confirm whether certain participants' behaviors matched closely with the behaviors from other participants, we used k-means clustering to group participants based on similar response patterns. This approach provides an empirical way to assess how much variability we see in individual responses and is useful to determine whether our participants appear to be using similar strategies to complete the experiment.

Exploratory Analyses

K-means Clustering Analysis

To assess the degree of similarity between each participant's response patterns, we used subject-specific coefficients (intercepts and slopes) from the Bayesian model as the basis for our analysis. Rather than relying on post hoc self-reports of strategy, which are often biased and poorly aligned with actual behavior (Nisbett & Wilson, 1977), we used these model-derived parameters to identify distinct response strategies empirically from their performance measures.

For each subject, a vector was constructed from the coefficients of four key parameters, which served as the inputs to the clustering analysis. Table 4.1 shows the extracted parameter estimates, as well as the amount of variability (between-subject variation, presented as standard deviations of the estimates) that was present in the model. This data was then scaled to ensure that the clustering assignments were not biased based on larger absolute values for certain parameters. Given that no between-condition differences were observed in the Bayesian model, Condition and any related interaction parameters were expressly omitted from the clustering analysis.

Table 4.1 Parameters Included in Clustering Analysis Showed Acceptable Degree of Between-subject Variation (SD).

Parameter	Std. Deviation Estimate	95% Credible Interval	
		Lower	Upper
Intercept	0.70	0.60	0.81
Slope (Time)	0.37	0.31	0.43
Slope (Interval Length)	0.16	0.14	0.19
Slope (Time $\times$ Interval Length)	0.07	0.06	0.09

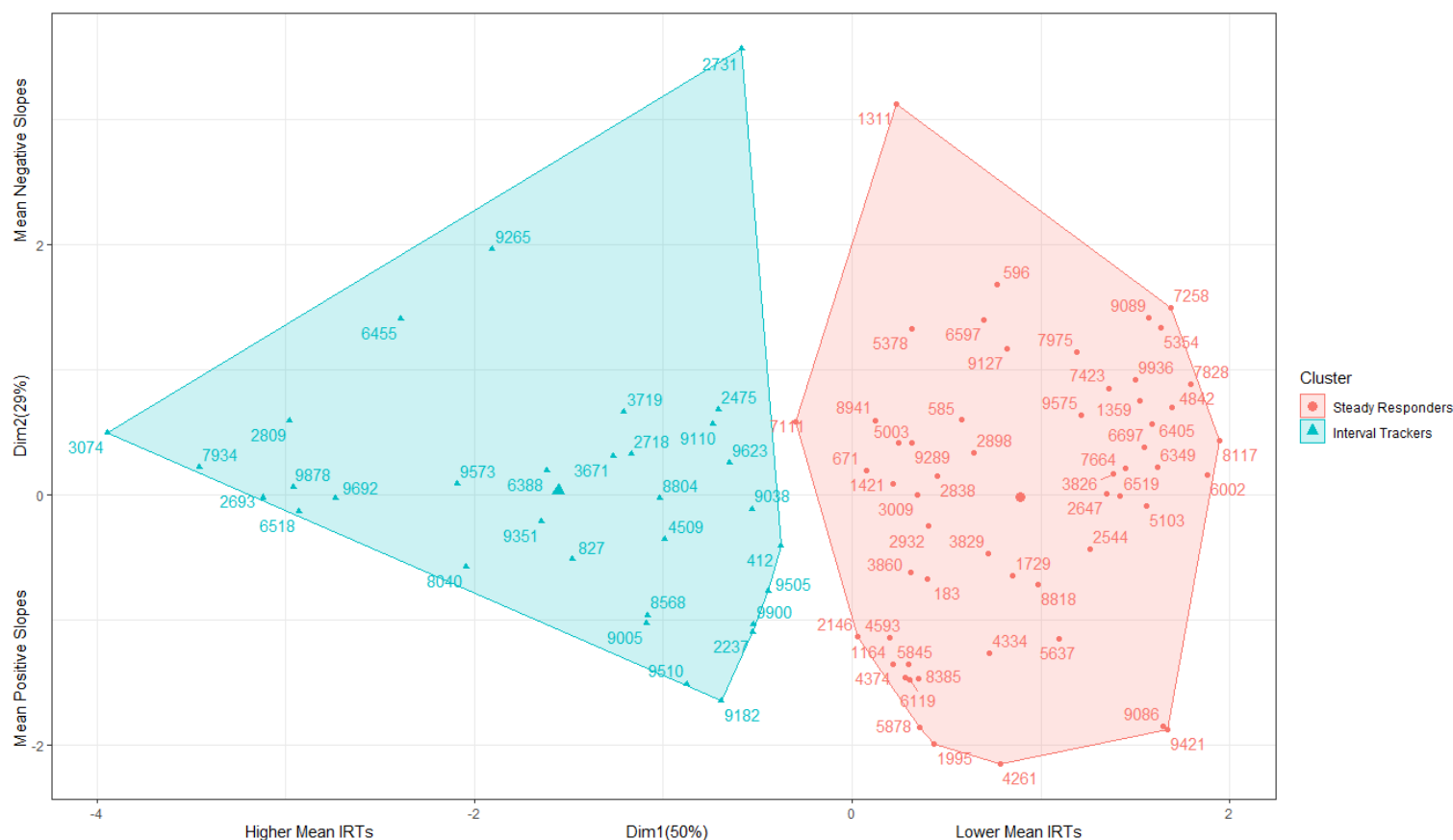
K-means Clustering Results

The optimal number of clusters was determined using the Silhouette method, which calculates a silhouette score for each data point for every cluster solution—reflecting how well each data point fits within an assigned cluster. The optimal number of clusters is the solution that maximizes the mean silhouette score (Rousseeuw, 1987).

The results indicated that the 5-cluster solution was the optimal number of clusters for our data (mean silhouette score = 0.331). However, we determined that the 2-cluster solution (mean silhouette score = 0.327) provided more interpretable cluster groupings without sacrificing cluster discriminability. Therefore, we will present the 2-cluster solution for all further analyses.

Figure 4.4 shows the cluster groupings across two principal component dimensions. Dimension 1 accounted for 50% of the total variability and corresponded closely to mean IRTs. Dimension 2 primarily distinguishes participants based upon their mean slopes and accounted for 29% of the total variability. These two dimensions showed great discriminability among participants and were used to qualify distinct strategies (e.g., response patterns) across all subjects.

Figure 4.4 K-means Clustering Solution Reveals Two Distinct Behavioral Response Patterns

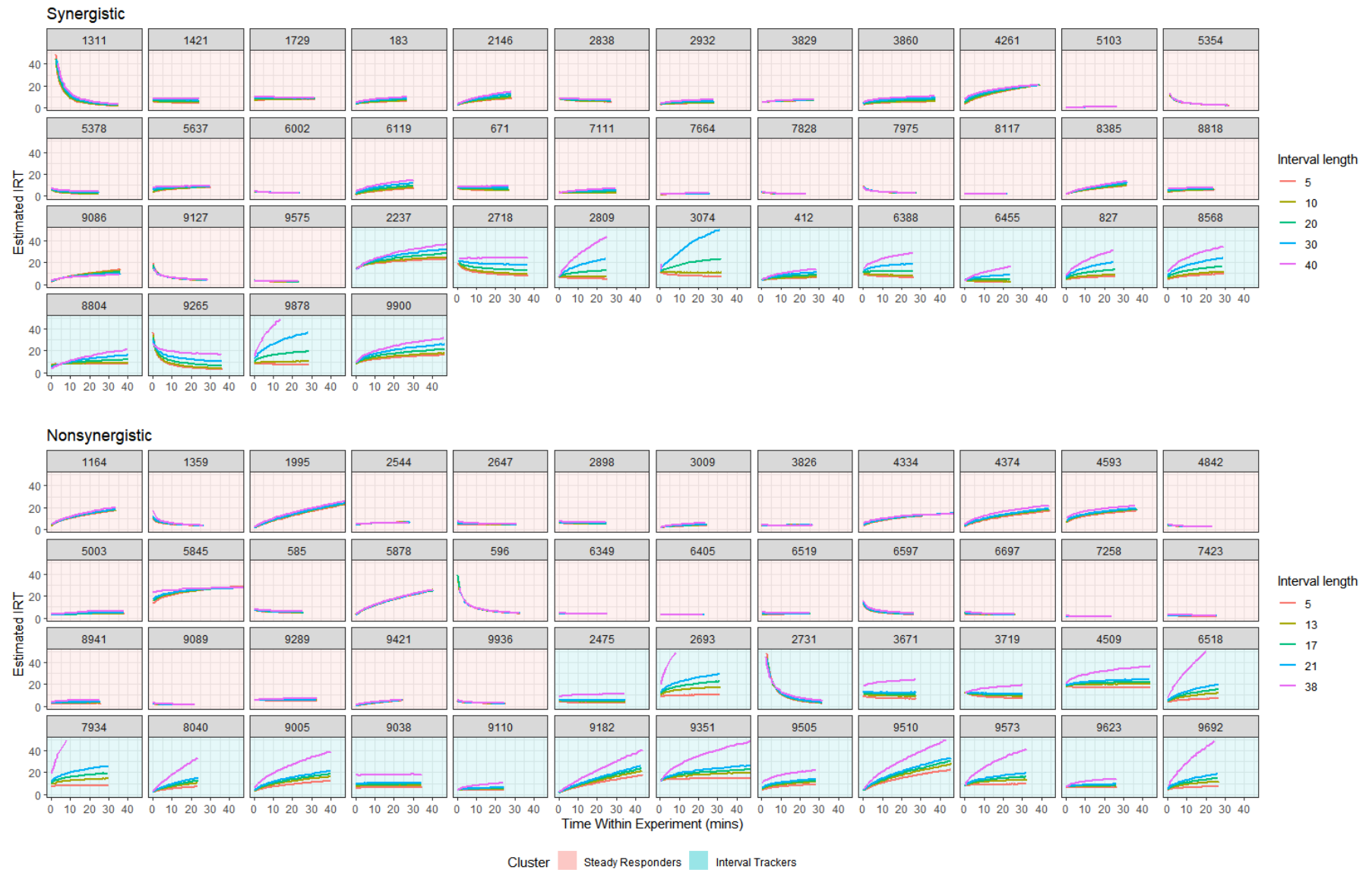


Cluster 1 included 56 participants (64%) and largely consisted of subjects with relatively stable IRTs across interval lengths. Participants in this cluster appeared to prefer consistent steady rates of responding to all intervals—resulting in similar mean IRT values regardless of Interval Length. In short, participants in Cluster 1 opted to respond to every interval frequently, which resulted in relatively short IRTs, and little-to-no discriminability among the different interval durations. Therefore, we labeled these participants as “Steady Responders”.

Cluster 2 included 32 participants (36%) and primarily included participants with longer mean IRTs that typically increased across the duration of the experiment, and with IRTs that increased as the length of the interval increased. These participants largely showed sensitivity to

each interval length, meaning that participants typically responded most often to the shortest interval, followed by the next shortest interval, and so on. We labeled these participants as “Interval Trackers”. Figure 4.5 shows each participant’s response patterns clustered similar to what we anticipated from the individual subject plots. According to the data, Steady Responders primarily elected to respond frequently to every interval at a consistent rate. This strategy was likely less mentally taxing, as they elected not to independently track the interval durations and responded to each one at the same rate, but was likely more physically taxing, as they made many more erroneous (unreinforced) responses to each interval before that interval was primed.

Figure 4.5 Individual Subject Data Support Distinct Response Profiles Identified by K-means Clustering



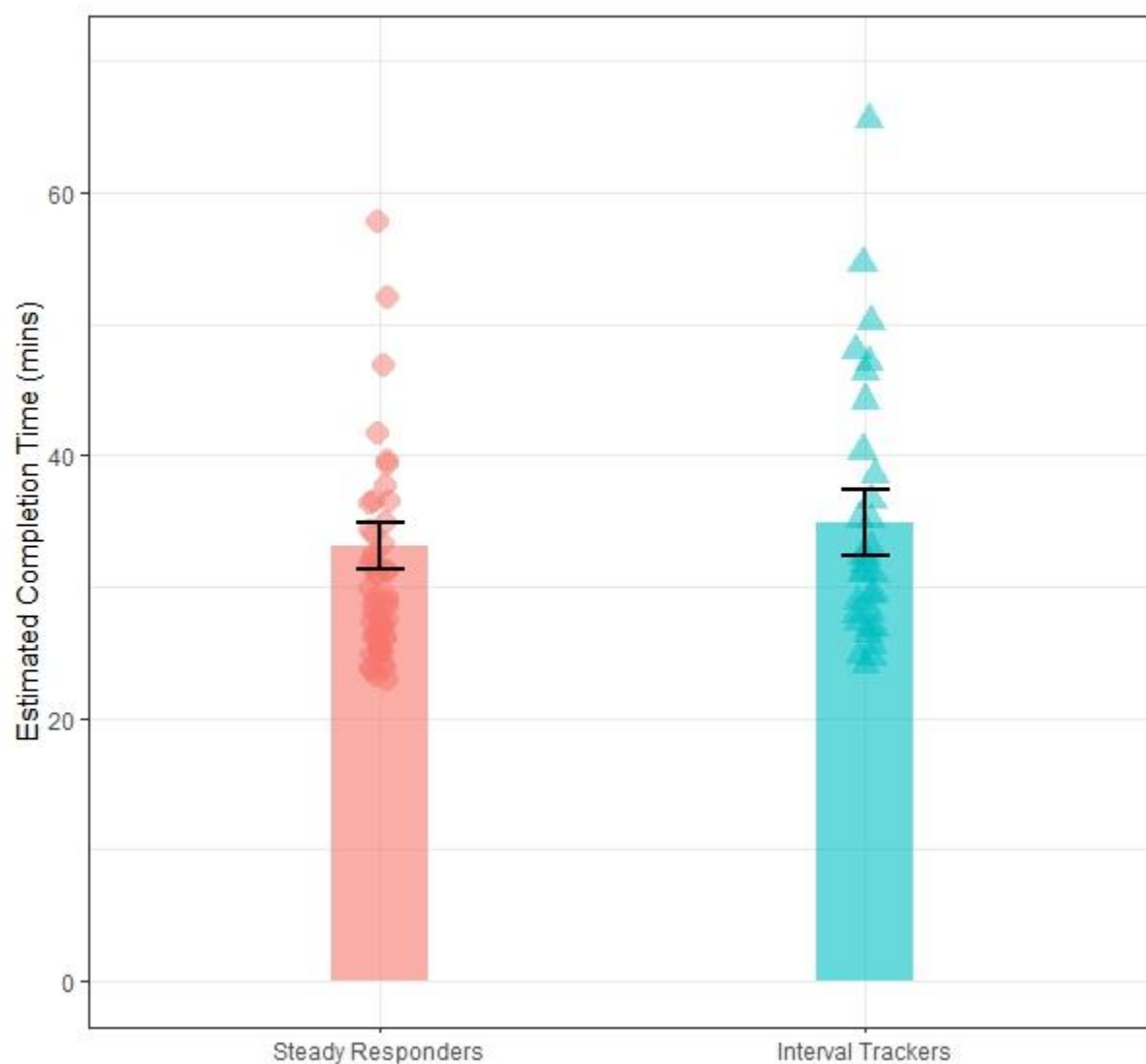
Interval Trackers cluster performed largely as we presumed participants would behave in our task. These participants showed sensitivity to the interval lengths by choosing to wait longer in between responses on the same key as the interval lengths increased. Compared to Steady Responders, these participants appeared to time each interval and primarily responded to each interval only when they perceived the interval's duration neared.

Time To Completion Analysis

We wanted to know whether one strategy was objectively optimal compared to another. While there are many ways to define optimality in this task, we concluded that Connect participants were strongly incentivized to complete our experiment in the shortest time possible while still completing all requirements. Therefore, we investigated whether one cluster's strategy resulted in lower completion times compared to the other cluster. Completion time was measured as the time the final reward was earned, which completed the task. Since each click was recorded, we could precisely identify how long each participant took to complete the experiment.

Like IRTs, completion times were positively skewed and were analyzed using a gamma regression with a log-link function. Cluster (Interval Trackers vs. Steady Responders) was effect-coded and included as the only predictor. No random effects were included as each subject only provided one observation. The omnibus ANOVA test revealed no differences between clusters ($\chi^2 = 0.32, p = 0.57$) suggesting that on average, participants in one cluster did not finish the experiment faster compared to the other cluster (see Figure 4.6).

Figure 4.6 Neither Strategy Led Participants to Finish the Task Earlier

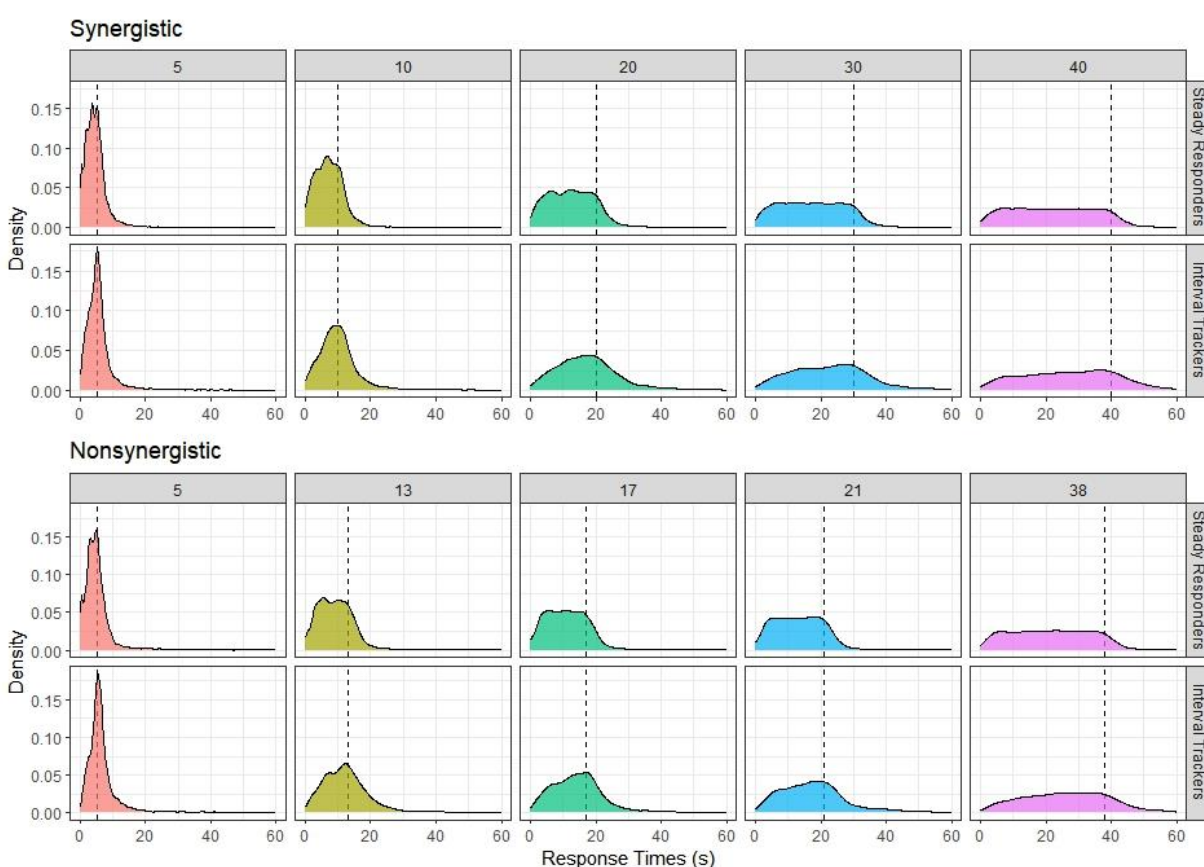


Note. Two participants in the Steady Responders cluster had completion times of 79 and 120 minutes. These outliers are not shown in the figure but were included in the analysis, and their omission does not alter the conclusions.

Figure 4.7 supports these conclusions by showing the distributions of all responses made within a given interval (separately from the inter-response times analyzed above). We expected to observe more pronounced effects of scalar timing for the Interval Trackers cluster, in addition to the typical scalloping behavior that is routinely observed under FI schedules. For the Steady

Responders, we did not anticipate seeing much evidence of scalar timing due to their preferred strategy to respond to each interval consistently, rather than actively time each interval's duration. However, when response persistence was investigated, Steady Responders showed greater persistence (i.e., were more likely to respond to the same key if the first response was not reinforced) compared to Interval Trackers (see Appendix D for more information on timing persistence).

Figure 4.7 Frequency of Responses Made Throughout the Interval's Duration



Overall, only 39% of participants tracked the interval lengths and adjusted their IRTs accordingly. Based on these results, as well as the results described above, we elected to measure scalar responding not based on the actual (programmed) interval duration, but instead on the interval duration that the participants responded to. For example, if participants responded with a

mean IRT of 10-s on a 20-s interval then we would treat this interval as 10-s for the purposes of assessing scalar properties using mean absolute deviation. This is necessary to ensure that the hypothesized effects of synergistic interval schedules are calculated based on the tracked intervals, rather than the programmed interval lengths.

Examining individual response patterns helped us to identify four key insights: 1) many of our participants adopted a strategy of pressing steadily to each interval, rather than timing each one. 2) This common response pattern likely explains the unintuitive results that we saw in the primary analyses (see Figures 4.2 and 4.3). 3) Clustering analysis agreed with these conclusions, noting two distinct clusters in our data. 4) In order to properly test our hypotheses regarding whether scalar timing was more pronounced in Experiment 2 compared to Experiment 1, we needed to investigate the mean absolute deviations (MAD) relative to the timed-intervals (participant's mean IRTs for each interval) rather than the programmed intervals, since those differed starkly from one another. To further confirm or reject the original hypotheses, we examined whether MAD values increased at a rate consistent with scalar timing.

Mean Absolute Deviation

We hypothesized that the synergistic schedules would enhance timing abilities across interval lengths, with the greatest benefits (shortest response errors) observed on the longest intervals. Specifically, synergistic schedules should reduce response errors, since rewards obtained from shorter intervals could serve as predictive cues for when to respond to longer intervals. In contrast, the non-synergistic schedules were designed such that rewards from shorter intervals provided little information about when to respond at longer intervals. Based on this design, we predicted that participants in the synergistic condition would exhibit response errors

that were shorter than scalar timing would predict, whereas participants in the non-synergistic condition would show scalar timing properties.

In the MAD analysis, we saw a large degree of individual variation among participants' IRTs, including how much the IRTs changed over the course of the experiments. To assess for scalar timing while accounting for changes in mean IRTs across the experiment's duration, we first measured the residuals for each IRT. Residuals were derived using the Bayesian model fits (estimated IRT - observed IRT) to assess variation around the modeled mean. We took the absolute value of those residuals to estimate the MAD of each interval. In other words, we used the fitted values as a baseline to control for variability caused by our independent variables. If scalar timing properties are present as predicted by SET, the MAD residual ratio should increase proportional to the IRT Ratio (e.g., if the interval doubled, the MAD should double). If nonscalar timing occurs as predicted by BeT, we should see the MAD residual ratio remain constant as the IRT ratio increases. Alternatively, if scalar timing properties are observed, the ratio between the MAD residual ratio and IRT ratio should remain equal to 1. Our results largely indicate that participants in both conditions demonstrated responses errors consistent with scalar timing properties, regardless of clustering. Appendix Figure E.1 visualizes individual participants' response distributions within selected intervals. Table 4.2 shows the MAD values by condition, cluster, and interval length.

Table 4.2 Measures of Scalar Timing using MAD by Condition and Cluster

Cluster	Interval (s)	IRT		MAD		IRT Ratio to MAD Ratio
		Mean (s)	Ratio	Residual	Residual Ratio	
Synergistic						
Steady Responders	5	4.46	1.00	2.25	1.00	1.00
Steady Responders	10	5.47	1.23	2.71	1.20	1.03
Steady Responders	20	6.22	1.39	3.35	1.49	0.93
Steady Responders	30	6.32	1.42	3.70	1.64	0.87
Steady Responders	40	6.51	1.46	3.99	1.77	0.82
Interval Trackers	5	7.13	1.00	3.65	1.00	1.00
Interval Trackers	10	11.11	1.56	5.01	1.37	1.14
Interval Trackers	20	15.78	2.21	7.30	2.00	1.10
Interval Trackers	30	17.37	2.44	7.99	2.19	1.11
Interval Trackers	40	20.37	2.86	11.35	3.11	0.92

Nonsynergistic

Steady Responders	5	5.25	1.00	2.71	1.00	1.00
Steady Responders	13	6.92	1.32	3.48	1.28	1.03
Steady Responders	17	7.35	1.40	3.69	1.36	1.03
Steady Responders	21	7.63	1.45	3.79	1.40	1.04
Steady Responders	38	8.01	1.53	4.35	1.61	0.95
Interval Trackers	5	7.05	1.00	3.50	1.00	1.00
Interval Trackers	13	13.21	1.87	5.61	1.60	1.17
Interval Trackers	17	14.11	2.00	5.71	1.63	1.23
Interval Trackers	21	16.02	2.27	6.75	1.93	1.18
Interval Trackers	38	18.53	2.63	12.44	3.55	0.74

Note. MAD = mean absolute deviation; IRT = interresponse times. Mean IRT and MAD metrics were computed within each condition $\times$ cluster $\times$ interval. Ratios use the 5-s interval as the reference. IRT to MAD ratios of 1 indicate scalar timing.

Chapter 5 - Discussion

We wanted to examine whether timing accuracy and precision could be modified by manipulating the structure of multiple interval schedules to either enhance (synergistic) or interfere with (nonsynergistic) natural timing abilities. We hypothesized that the participants would track each interval and respond roughly proportional to the interval's programmed length. When we examined individual subject data, it became clear why we observed IRT's much lower than we hypothesized. We then used the MAD ratio to estimate whether timing errors increased as the length of the interval increased.

We found that scalar timing properties were present in both synergistic and nonsynergistic designs, suggesting that the synergistic interval structures failed to aid participant's ability to track multiple intervals simultaneously. In addition, we found that many of our participants elected not to time any of the intervals at all. Rather, these Steady Responders elected to consistently respond to every interval equally. While this strategy did not translate to faster completion times—an implicit goal for online participants—we did see the majority of participants adopt this strategy. Below, we will discuss some plausible explanations as to why our participants elected to forfeit tracking of the intervals, why we may have failed to observe benefits of synergy, what our results suggest for the contemporary timing theories, and what implications our results may have for everyday life.

Mental vs. Physical Effort

One proposed explanation for why we saw many participants opt to avoid timing can be explained by an inherent tradeoff between mental and physical effort in the current task. Like many timing tasks, our task requires two types of effort: mental effort to track each interval, and the physical effort required to respond and receive reinforcement. While effort has been difficult

to operationalize (for further details, see: Thomson & Oppenheimer, 2022), we simply wish to contrast these two types of effort within the confines of our experiments, to shed light on these two distinct strategies employed by the participants.

We predicted that the synergistic design of Experiment 1 would greatly reduce the mental demands necessary to time all five intervals. We hypothesized that participants would use the reward obtained from shorter intervals to infer when to respond to the longer intervals. We also expected the nonsynergistic design to be much more mentally effortful, as the rewards from shorter intervals were much less predictive in Experiment 2.

We can make some inferences regarding the general behavioral patterns we observed. The results of the clustering analysis suggest a strong tradeoff between the mental and physical effort required to complete this task, with many participants opting to take the physically demanding, mentally effortless strategy of consistent responding. The fact that we saw roughly the same proportion of participants in the Steady Responder's cluster for the synergistic design (67%) and the nonsynergistic design (60%) appears to suggest that participants did not perceive the synergistic condition to be less mentally effortful when timing each interval. However, that explanation may be misleading, as participants may not have thoroughly tested more than one strategy before adopting their preferred method.

Consider synergistic participant 9597, who was classified as a Steady Responder (Figure 4.5). According to the model estimates, this participant showed very little difference between the IRTs at the beginning versus the end of the experiment. Instead, this participant appears to have discovered early on that rapidly responding was a viable strategy, and rarely, if ever, deviated from that strategy. For this participant, it would be wrong to infer that this participant selected their strategy only after testing multiple approaches. Nevertheless, we can infer that while this

strategy required some degree of physical effort, the effort required was either tolerable enough in order to finish the task as quickly as possible or was not too costly to warrant searching for a less effortful strategy.

Compare the previous participant to synergistic participant 9265, classified as an Interval Tracker. This participant's IRTs across each interval duration started off quite long and relatively consistent (lower sensitivity to Interval Length), but their IRTs decreased over the course of the experiment, while discriminability to the interval lengths improved. This type of response pattern would be expected as the participant learns each interval's duration. However, this participant's data also indicates a noticeable shift (e.g., a change in strategy) early on. While initially waiting for long durations for each interval before clicking, the participant quickly abandoned that strategy for a strategy better suited to exploit each of the intervals to maximize the overall rate of reinforcement. In this case, participant 9265 adopted a strategy requiring less physical effort, at the cost of requiring more mental effort.

Overall, these tradeoffs between mental and physical effort may have played a significant role in interval timing behavior, and as more intervals must be tracked simultaneously the cognitive effort demands increase exponentially while the physical demands increase only modestly. It's important to note that our results do not suggest that many of our participants showed diminished timing abilities. Instead, it would be more accurate to infer that these participants chose not to invest the cognitive effort required to time all five intervals.

Our results seem to suggest that either 1) the benefits of the synergistic design weren't obvious enough for participants to recognize and subsequently exploit. or 2) the mental demands to track five independent intervals, regardless of synergy, was just too high. The fact that we still observed scalar timing in the synergistic Interval Trackers strongly suggests that these benefits

were much weaker than originally anticipated. While effort trade-offs may account for which strategy was adopted, we can examine the exploration vs. exploitation tradeoff to help clarify how these different strategies emerged over time.

Exploration vs. Exploitation Within Interval Schedules

Over the course of any timing or decision-making experiment, an organism may display behaviors consistent with both exploration and exploitation. Exploration often involves behaviors designed to seek out information, such as occasionally sampling new food patches to see if a better patch exists. These behaviors may not be optimal in the short-term, but over time, these exploratory behaviors ensure the organism approximates optimal behavior by maximizing the global reward rate (Cohen et al., 2007). This tradeoff between exploration and exploitation is also highly relevant to interval timing tasks, especially when there are multiple intervals present.

Simple FI schedules with one interval require very little need for exploration once the interval is learned. But many timing situations involve multiple intervals. Consider two concurrent FI schedules (e.g., a 30-s FI and a 60-s FI), that must be tracked simultaneously. An exploitation-based strategy may involve the organism directing most of their behavior to the FI-30 schedule, but failure to direct any attention to the other interval forfeits potential reinforcers. Therefore, humans and many other species employ common heuristics, such as directed exploration, to ensure leaner FI schedules are still sampled, just less frequently (Kleinman et al., 2016; Wilson et al., 2014). This is the exact strategy we anticipated seeing—and did see—in our Interval Trackers. This method of directed exploration may provide one plausible explanation for why participants elected to respond to longer intervals much earlier than we would expect. Next, we compare these strategies by examining the alternative behaviors participants adopted to reduce effort.

Alternate Forms of Optimality

Optimality can be defined in at least two ways for our experiment. Participants can choose to minimize the time required to reach 500 points—thereby completing the experiment, or they can minimize the total amount of clicks made (fewest erroneous responses). Our results suggest that neither strategy completed the experiment faster than the other, but we can see large differences between the total numbers of clicks produced (Table 5.1; Appendix Figure F.1).

Table 5.1 Total Clicks Varied Drastically by Cluster but not Condition

Cluster	Median Total Clicks	Minimum	Maximum	N
Synergistic				
Steady Responders	2066	938	3949	27
Interval Trackers	1031	571	2127	13
Nonsynergistic				
Steady Responders	1954	721	5056	29
Interval Trackers	1036	639	2238	19

This suggests that steadily responding was not a better strategy but was largely preferred by 64% of participants. This strategy appears to require more physical effort, but less mental straining, compared to the Interval Trackers' primary strategy.

Consistent Responding

We likely saw so many Steady Responders due to the lack of any response costs, or penalties for incorrect responding. Response costs are commonly delivered as a brief time-out period or are implemented strategically to simulate naturalistic demands (e.g., changeover delays to simulate travel time in foraging tasks). Changeover delays are often used in timing research to discourage or penalize alternating between two concurrent intervals. However, in our design, we opted not to include changeover delays as they would interfere with the natural rhythmic

responding patterns hypothesized. Because neither changeover delays, peak trials, nor timeout periods were included in the current design, participants were free to respond to any interval at any point during the experiment. The only additional ‘cost’ incurred from an early response was a brief auditory cue to inform the participant their click registered but occurred too early to be rewarded.

We wanted to avoid using changeover delays in our task as their presence discourages exploratory behaviors. With five independent intervals, we expected a fair degree of exploratory behaviors to occur early in the experiment. But as Figure 4.5 indicates, participants' response patterns quickly plateaued within the first 10 minutes. Most participants reached response plateaus within the first 10 minutes of either experiment. This result suggests that the steady responders showed few, if any, attempts to explore the individual intervals before settling into their desired strategy. In contrast, interval trackers seem to show differences in the individual slopes for each of the interval lengths, suggesting that these participants took more time to switch from exploration to exploitation, presumably once the approximate interval lengths were learned. Simply put, steady responders quickly switched from exploration to exploitation once they learned that consistent responding was an effective strategy. Interval trackers on the other hand took longer to cease their explorations of the intervals, before they began exploiting the intervals at rates proportional to their durations.

In most real-world situations, there is a direct cost for mistiming a given interval. The severity of these costs vary depending upon many factors, but generally, the larger the timing error, the more costly the consequences are. For example, in a busy kitchen, overcooking a dish by 30 seconds may be acceptable, but overcooking by 3 minutes can render the dish inedible,

wasting ingredients, time, and effort. In situations like these, timing errors carry clear monetary costs, in addition to the extra effort required to remake an order.

In our task, the cost for timing misestimations was intentionally minimal. No penalties were applied to early responses meaning participants could immediately respond again (see Appendix D). On the other hand, late responses resulted only in lost reinforcement opportunities, and even consistently poor timing estimations would translate to, at most, a modest increase (on the order of a few minutes) in total task duration. This low-cost structure likely reduced pressure to fine-tune timing performance while also making strategies of consistently high responding rates especially attractive.

Lack of Synergistic Benefits

Participants may have failed to capitalize on the benefits of synergistic schedules because the demands of the task were just too challenging. While we carefully selected intervals to be maximally synergistic, learning and maintaining the relative positions of five intervals may have exceeded what many participants were willing to track. Perhaps future research should investigate the benefits of synergy using a fewer number of intervals. By reducing the total number of simultaneous intervals to three, participants may be more likely to allocate the mental effort necessary to time each one.

One central advantage of synergy rests in how these intervals are scaffolded. We can likely improve the effectiveness of synergy by slowly titrating or introducing new intervals only after participants have successfully acquired the previous intervals. For example, participants could begin with two intervals (FI-5 and FI-10) and once responding stabilizes, a third interval would be added. If this third interval is synergistic (e.g., FI-20), then participants should quickly acquire the new interval and integrate their responses to the interval with relative ease,

potentially reducing timing errors relative to the other intervals (scalar timing). However, if a nonsynergistic interval is introduced (FI-17), acquisition should be slower and scalar timing errors should be present.

This titration approach also enables us to directly test the limits of synergy. Not only can we assess whether these benefits are even detectable, we may establish when these benefits begin attenuating. For example, we may see benefits of synergy when only two or three synergistic intervals are active, but once a fourth interval is introduced, these benefits may become negligible. This boundary point may shed light on the individual differences in the ability to coordinate multiple intervals, but can isolate a natural point of failure, at which point the mental effort is too great, or our abilities too limited, leading participants to adopt less effortful strategies, similar to our Steady Responders.

Mental effort may also be reduced by widening the spacing of the intervals, without needing to decrease the total number of active intervals. Racey et al., (2011) employed a similar approach, showing that humans can time as many as eight independent synergistic variable interval (VI) schedules when their means increased by multiples of three (3, 9, 27, ... up to 6,561 s; $\pm 50\%$ variability). In that study, participants showed relatively good sensitivity to intervals below 30-s (see Racey et al., Fig. 6) showing systematic decreases in the probability of responding to a given key as the VI interval length increased. Notably, most participants—though not all—appeared to track and allocate responses reliably to roughly four of the eight programmed intervals, with responding dropping sharply for the leaner VI schedules (≥ 243 s; see Racey et al., Fig. 9). Appendix G shows similar decreases in response frequencies for Interval Trackers, but not steady responders.

Racey and colleagues' (2011) results imply that participants may elect to ignore intervals that are too long—only returning to them occasionally. In essence, if too many intervals are present, or the interval durations grow too long relative to shorter (richer) intervals, participants may opt not to track them at all. In the current task, the Steady Responders appeared to reduce mental effort via a different strategy. Rather than selectively ignoring one or two of the longer intervals, they opted to distribute responses evenly across all interval lengths.

One limitation of the current study is that it does not account for the widespread use of external cues to support timing estimates in everyday life. Environmental signals, such as visual changes in food throughout the cooking process or aromas like the smell of burning, provide information about approximately how long the item has been cooked for, and whether the optimal cooking time has been exceeded. This study did not include additional cues to aid timing of multiple intervals, but future research may benefit from investigating how other types of cues, independently from reinforcement, assist timing in multi-interval environments.

Implications for Timing Theories

SET is a seminal theory of timing whose predictions have been experimentally supported over many decades. Our results also support the predictions made by SET under novel experimental conditions. Consistent with SET's predictions, scalar timing properties were observed to some degree for every interval length, regardless of how the intervals were structured (see Table 4.2). When combined with the results from Leak and Gibbon (1995), our results suggest that SET accurately accounts for timing behaviors in multiple interval schedules, whether these intervals are sequentially-nested, or occurring simultaneously.

The results from our synergistic design largely do not support the predictions from BeT (Killeen & Fetterman, 1988, 1993). In our design, we hypothesized that responding to the 5-s

interval would act as an adjunctive behavior while the participant waited to respond to longer intervals. Therefore, each response to the shortest FI interval creates one adjunctive behavior in a sequence, which then informs when to respond to a longer interval. A novel aspect of this design, which included multiple intervals running simultaneously, means the participant can get concrete feedback based on their timing errors on every interval.

When a participant responds to an interval too early, that response earns no reward. Even so, the participant can use the time between that first early response and the next response to the same interval that does earn a reward as a rough gauge of how early they were. This method to estimate timing errors is distinctly separate from the IRTs analyzed above. We defined IRTs as the post-reinforcement pause (the time between a reinforced response and the next response to the same interval). Here, if their IRT is less than the interval's duration, no reinforcement is delivered. The participant may subsequently infer that they responded too early and adjust behavior accordingly.

There are two important limitations. First, this method is only useful when a response is early. If a participant responds late—after the interval is baited—they receive a reward, but there is no simple way to estimate how late that response was. Second, even when they respond early, this method can only be used to approximate the timing error, compared to more explicit feedback (e.g., “1.35 seconds early”).

Nevertheless, this feedback serves two roles. Participants can use this feedback to refine their timing estimates while also tracking how much time has passed. Like adjunctive behavioral states in BeT's theory of timing, this feedback serves as a counter that signals when to respond on longer intervals. Importantly, if reinforcement was removed entirely from this experiment, BeT and SET would both suggest that scalar timing would be observed (Gibbon, 1977, 1992;

Killeen & Fetterman, 1988). Instead, participants can use feedback to refine timing estimates for each interval, resulting in reduced variability of timing errors for the longer intervals (nonscalar timing). Figure 5.1 shows how the presence (solid lines) or absence (dashed lines) of a reinforcer can influence the overall variability of timing errors (shaded).

Figure 5.1 Intermediate Rewards on Shorter Interval Schedules are Hypothesized to Reduce Timing Errors on Longer Schedules



Note. FI = fixed interval. Red shading shows timing uncertainty expands as more time elapses (scalar timing). With no feedback, the uncertainty after 2 minutes is the same whether it's one FI-2 min or two back-to-back FI-1 min periods. When rewards are delivered after responses beyond 1 minute (solid line), the clock is effectively reset, so the uncertainty at 2 minutes should equal that of a single FI-1 min, not two.

As stated previously, we failed to demonstrate any beneficial effects of synergy. Despite providing feedback on each interval, participants did not appear to leverage the rewards from the shorter intervals to reduce timing errors on the longer schedules. Participants either did not (or could not) use short-interval reinforcement as a reliable counter for longer intervals, or that the added cognitive cost of doing so outweighed any potential benefits. It's also possible that

participants were not using responses to the FI-5 as adjunctive behaviors, opting instead to respond equally to each interval (Steady Responders) or perhaps participants may have chosen to engage in other adjunctive behaviors, which did not involve responding to any other interval.

Multiple factors influence the rate and form of adjunctive behavior. Adjunctive responding is most robust under periodic, time-based schedules of reinforcement and is especially likely when the adjunctive behavior is phylogenetically relevant to the primary reinforcer (e.g., schedule-induced drinking when food reinforcers are provided; López-Tolsa & Pellón, 2025).

Behaviors can either collaborate or compete with one another in a behavioral sequence, for example as one behavior occurs more often, rates of other behaviors must decline. But when behaviors function as discriminative stimuli for subsequent behaviors in a sequence, rates of adjunctive behaviors increase (Arday & Rellón, 2004; Pellón & Killeen, 2015; López-Tolsa & Pellón, 2025).

Across both human and nonhuman studies, the relation between FI length and adjunctive behavior typically follows a quadratic form, where rates increase as the interval lengthens up to a point before decline once intervals extend beyond approximately 120 s (Falk, 1966; Wallace et al., 1975; Allen & Kenshalo Jr., 1976; Gray Granger et al., 1984). In the current study, our intervals ranged from 5-s to approximately 40-s, which is well within the typical period where adjunctive behaviors occur more frequently, and should therefore support the emergence of robust rates for adjunctive behaviors, particularly for the longest intervals.

When interpreting these results, it is important to acknowledge that these theories may be contextually dependent rather than mutually exclusive. For instance, when allowed to use behaviors as a proxy for timing estimates, most people will. The most common example of using

behaviors as proxies for timing involves counting (e.g., “one one-thousand, two one-thousand...”). The time it takes to say “one thousand” is approximately one second. Therefore, by counting how many times this behavior occurs, a reasonable estimation of timing can be produced. When use of these behavioral proxies are restricted, for example, when participants are unable or instructed not to count, such was the case in the current study, then timing estimates may rely much more on internal, less precise estimations of time passing.

Nevertheless, our results showed strong support for the predictions made by SET and suggest that BeT fails to accurately account for experimental findings involving multiple interval schedules (Leak & Gibbon, 1995).

Implications for a General Audience

Timing is a ubiquitous skill that is present in countless daily activities. The ability to keep track of multiple independent intervals at the same time is a natural part of the job for most Chefs, Air traffic controllers, nurses and many other occupations. The fact that humans successfully perform these jobs despite their inherent timing demands provides clear evidence that humans are capable of this ability.

Timing is also effortful. In many fast-food kitchens, for example, alarms and other external timers are built into equipment to help reduce the mental demands of their employees. Even when only a single interval must be timed, relying on external devices to provide assistance is crucial for employees who want to reduce the risk of under- or overcooking meals, as well as companies who want to reduce waste while ensuring every item is prepared consistently regardless of chef or location.

We hypothesized that we could improve timing performance by structuring intervals so that each one was a multiple of another, making it easier to track even longer intervals.

Nevertheless, we found no measurable improvement in our synergistic condition. Rather, many participants appeared to adopt a low-effort strategy of responding at a relatively constant rate instead of actively timing each interval.

Our results do not rule out the benefits of synergy. Instead, when those benefits were not easily recognizable, participants tended to come up with alternative ways to reduce effort, even if that tradeoff involved more physical effort for substantially less mental effort. But this tradeoff our Steady Responders adopted is not always viable in real-world settings. In the kitchen, a chef may choose not to time how long a burger has been cooking but instead gauge how cooked the burger is by repeatedly cutting into it or probing it with a thermometer— a technically feasible, but inelegant workaround that a novice might use before quickly abandoning it. Clearly, humans have adopted many different heuristics to reduce mental effort, such as setting timers, or by using spatial organization to offload additional effort and memory requirements, such as placing the burgers that should be cooked to medium on one side of the grill, and medium rare burgers on the other.

Accurately and precisely tracking multiple intervals at once benefits many aspects of daily life. Whether it's a chef ensuring food is perfectly prepared, clinicians managing multiple patients' dosing schedules, or air-traffic controllers ensuring that each and every flight is safe, and on schedule, this ability is mentally draining. By continuing to investigate how we can improve our environments to facilitate our ability to accurately track each dish, patient, or flight independently, we can improve job quality and safety while reducing errors and accidents.

Chapter 6 - References

- Allan, L. G. (1998). The influence of the scalar timing model on human timing research. *Behavioural Processes*, 44(2), 101–117. [https://doi.org/10.1016/S0376-6357\(98\)00043-6](https://doi.org/10.1016/S0376-6357(98)00043-6)
- Allen, J. D., & Kenshalo Jr., D. R. (1976). Schedule-induced drinking as a function of interreinforcement interval in the Rhesus monkey. *Journal of the Experimental Analysis of Behavior*, 26(2), 257–267. <https://doi.org/10.1901/jeab.1976.26-257>
- Arday, J., & Rellón, R. (2004). Effects of withholding the opportunity to press the operant lever on the maintenance of schedule-induced drinking in rats. *Mexican Journal of Behavior Analysis*, 30(1), 79–91. <https://doi.org/10.5514/rmac.v30.i1.25211>
- Bader, F., & Wiener, M. (2021). Awareness of errors and feedback in human time estimation. *Learning & Memory*, 28(5), 171–177. <https://doi.org/10.1101/lm.053108.120>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- 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>
- Bizo, L. A., & White, K. G. (1994). The behavioral theory of timing: Reinforcer rate determines pacemaker rate. *Journal of the Experimental Analysis of Behavior*, 61(1), 19–33. <https://doi.org/10.1901/jeab.1994.61-19>
- Bizo, L. A., & White, K. G. (1995a). Biasing the pacemaker in the behavioral theory of timing. *Journal of the Experimental Analysis of Behavior*, 64(2), 225–235. <https://doi.org/10.1901/jeab.1995.64-225>
- Bizo, L. A., & White, K. G. (1995b). Reinforcement context and pacemaker rate in the behavioral theory of timing. *Animal Learning & Behavior*, 23(4), 376–382. <https://doi.org/10.3758/bf03198937>
- Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philos Trans R Soc Lond B Biol Sci*, 362(1481), 933–942. <https://doi.org/10.1098/rstb.2007.2098>
- Dews, P. B. (1970). The theory of fixed-interval responding. In W. N. Schoenfeld (Ed.), *The theory of reinforcement schedules* (pp. 43–61). Appleton-Century-Crofts.
- Falk, J. L. (1966). Schedule-induced polydipsia as a function of fixed interval length. *Journal of the Experimental Analysis of Behavior*, 9(1), 37–39. <https://doi.org/10.1901/jeab.1966.9-37>

- Fetterman, J. G., & Killeen, P. R. (2010). Prospective and retrospective timing by pigeons. *Learning & Behavior*, 38(2), 119–125. <https://doi.org/10.3758/LB.38.2.119>
- Freestone, D. M., & Church, R. M. (2010). The importance of the reinforcer as a time marker. *Behavioural Processes*, 84(1), 500–505. <https://doi.org/10.1016/j.beproc.2010.01.011>
- Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, 84, 279–325. <https://doi.org/10.1037/0033-295X.84.3.279>
- Gibbon, J. (1992). Ubiquity of scalar timing with a Poisson clock. *Journal of Mathematical Psychology*, 36(2), 283–293. [https://doi.org/10.1016/0022-2496\(92\)90041-5](https://doi.org/10.1016/0022-2496(92)90041-5)
- Götz, F. M., Gosling, S. D., & Rentfrow, P. J. (2022). Small Effects: The Indispensable Foundation for a Cumulative Psychological Science. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 17(1), 205–215. <https://doi.org/10.1177/1745691620984483>
- Gray Granger, R., Porter, J. H., & Christoph, N. L. (1984). Schedule-induced behavior in children as a function of interreinforcement interval length. *Physiology & Behavior*, 33(1), 153–157. [https://doi.org/10.1016/0031-9384\(84\)90025-8](https://doi.org/10.1016/0031-9384(84)90025-8)
- Haight, P. A., & Killeen, P. R. (1991). Adjunctive behavior in multiple schedules of reinforcement. *Animal Learning & Behavior*, 19(3), 257–263.
- Killeen, P. R., & Fetterman, J. G. (1988). A behavioral theory of timing. *Psychological Review*, 95(2), 274–295.
- Killeen, P. R., & Fetterman, J. G. (1993). The behavioral theory of timing: Transition analyses. *Journal of the Experimental Analysis of Behavior*, 59(2), 411–422. <https://doi.org/10.1901/jeab.1993.59-411>
- Kleinman, M. R., Sohn, H., & Lee, D. (2016). A two-stage model of concurrent interval timing in monkeys. *Journal of Neurophysiology*, 116(3), 1068–1081. <https://doi.org/10.1152/jn.00375.2016>
- Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, 8(1), 33267. <https://doi.org/10.1525/collabra.33267>
- Leak, T. M., & Gibbon, J. (1995). Simultaneous timing of multiple intervals: Implications of the scalar property. *Journal of Experimental Psychology: Animal Behavior Processes*, 21(1), 3–19. <https://doi.org/10.1037/0097-7403.21.1.3>
- Lejeune, H., & Wearden, J. H. (2006). Scalar properties in animal timing: Conformity and violations. *The Quarterly Journal of Experimental Psychology*, 59(11), 1875–1908. <https://doi.org/10.1080/17470210600784649>

- López-Tolsa, G. E., & Pellón, R. (2025). Dynamic interactions between induction and reinforcement in the organization of behavior. *Perspectives on Behavior Science*, 48(2), 315–339. <https://doi.org/10.1007/s40614-025-00453-5>
- MacEwen, D., & Killeen, P. R. (1991). The effects of rate and amount of reinforcement on the speed of the pacemaker in pigeons' timing behavior. *Animal Learning & Behavior*, 19, 164–170.
- Morgan, L., Killeen, P. R., & Fetterman, J. G. (1993). Changing rates of reinforcement perturbs the flow of time. *Behavioural Processes*, 30, 259–272.
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195–203. <https://doi.org/10.3758/s13428-018-01193-y>
- Pellón, R., & Killeen, P. R. (2015). Responses compete and collaborate, shaping each others' distributions: Commentary on Boakes, Patterson, Kendig, and Harris (2015). *Journal of Experimental Psychology: Animal Learning and Cognition*, 41(4), 444–451. <https://doi.org/10.1037/xan0000067>
- Racey, D., Young, M. E., Garlick, D., Ngoc-Minh Pham, J., & Blaisdell, A. P. (2011). Pigeon and human performance in a multi-armed bandit task in response to changes in variable interval schedules. *Learning & Behavior*, 39(3), 245–258. <https://doi.org/10.3758/s13420-011-0025-7>
- Riemer, M., Kubik, V., & Wolbers, T. (2019). The effect of feedback on temporal error monitoring and timing behavior. *Behavioural Brain Research*, 369, 111929. <https://doi.org/10.1016/j.bbr.2019.111929>
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Rule, S. J., & Curtis, D. W. (1985). Ordinal properties of perceived average duration: Simultaneous and sequential presentations. *Journal of Experimental Psychology: Human Perception and Performance*, 11(4), 509–516. <https://doi.org/10.1037/0096-1523.11.4.509>
- Skinner, B. F. (1938). *The Behavior of Organisms*. Appleton-Century-Crofts.
- Thomson, K. S., & Oppenheimer, D. M. (2022). The “Effort Elephant” in the Room: What Is Effort, Anyway? *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 17(6), 1633–1652. <https://doi.org/10.1177/17456916211064896>
- van Rijn, H., & Taatgen, N. A. (2008). Timing of multiple overlapping intervals: How many clocks do we have? *Acta Psychologica*, 129(3), 365–375. <https://doi.org/10.1016/j.actpsy.2008.09.002>

- Wallace, M., Singer, G., Wayner, M. J., & Cook, P. (1975). Adjunctive behavior in humans during game playing. *Physiology & Behavior*, *14*(5), 651–654.
[https://doi.org/10.1016/0031-9384\(75\)90194-8](https://doi.org/10.1016/0031-9384(75)90194-8)
- Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., & Cohen, J. D. (2014). Humans use directed and random exploration to solve the explore–exploit dilemma. *Journal of Experimental Psychology: General*, *143*(6), 2074–2081.
<https://doi.org/10.1037/a0038199>

Appendix A - Supplemental Descriptive Statistics

Appendix Table A.1 Supplemental Descriptive Statistics for Figure 4.1

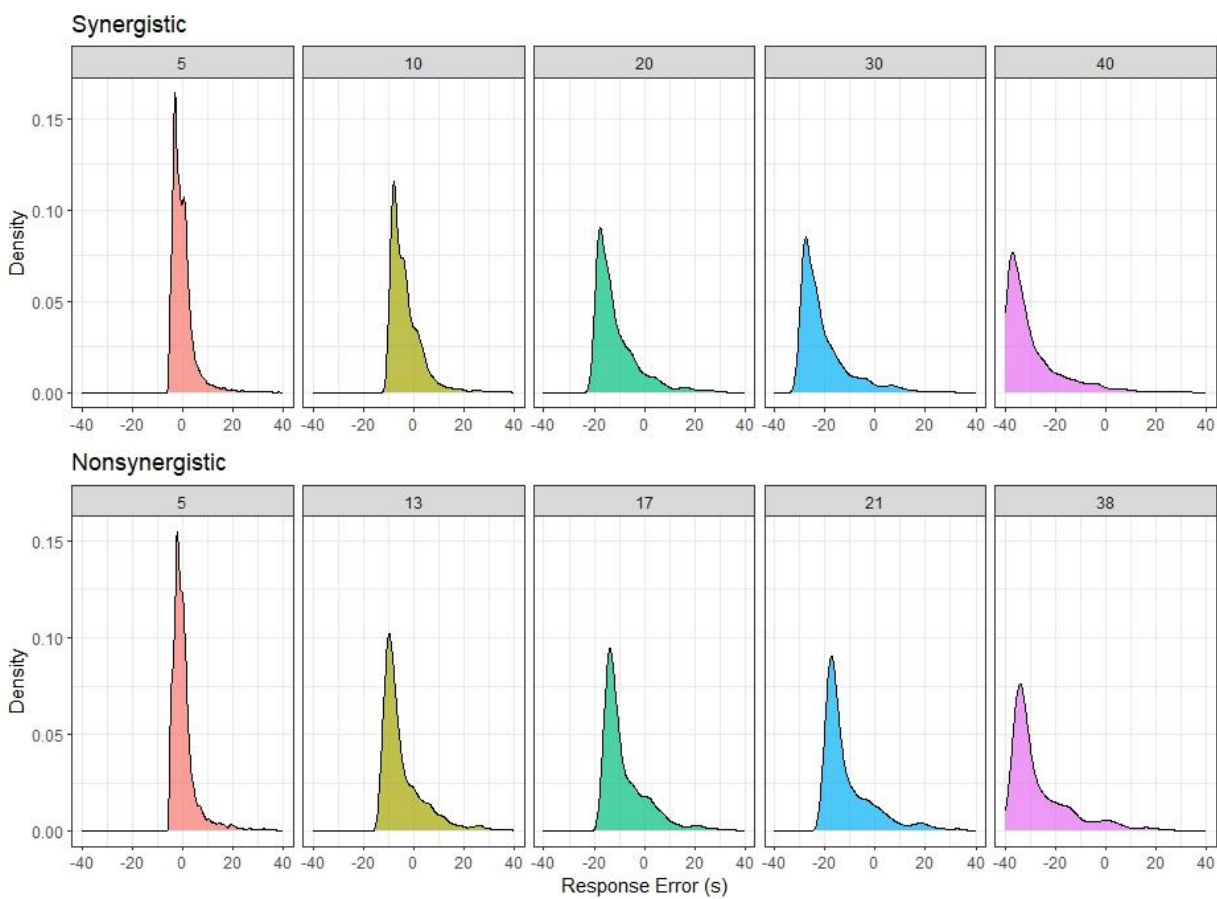
Condition	Interval Length (s)	Phase	Median Response Time (s)	Minimum	Maximum
Synergistic	5	First Half	4.7	0.1	73.7
	5	Second Half	4.4	0.1	73.7
	10	First Half	7.5	0.1	73.3
	10	Second Half	7.7	0.1	82.1
	20	First Half	13.0	0.1	74.2
	20	Second Half	13.3	0.1	81.9
	30	First Half	18.0	0.2	93.5
	30	Second Half	18.5	0.1	96.4
	40	First Half	23.3	0.1	100.8
	40	Second Half	23.2	0.1	98.6

Nonsynergistic	5	First Half	4.8	0.1	83.7
	5	Second Half	5.0	0.1	78.1
	13	First Half	9.5	0.1	80.0
	13	Second Half	9.9	0.1	79.8
	17	First Half	11.6	0.1	89.8
	17	Second Half	12.1	0.1	91.4
	21	First Half	13.8	0.1	94.2
	21	Second Half	14.7	0.1	93.7
	38	First Half	22.7	0.1	98.7
	38	Second Half	23.2	0.1	114.9

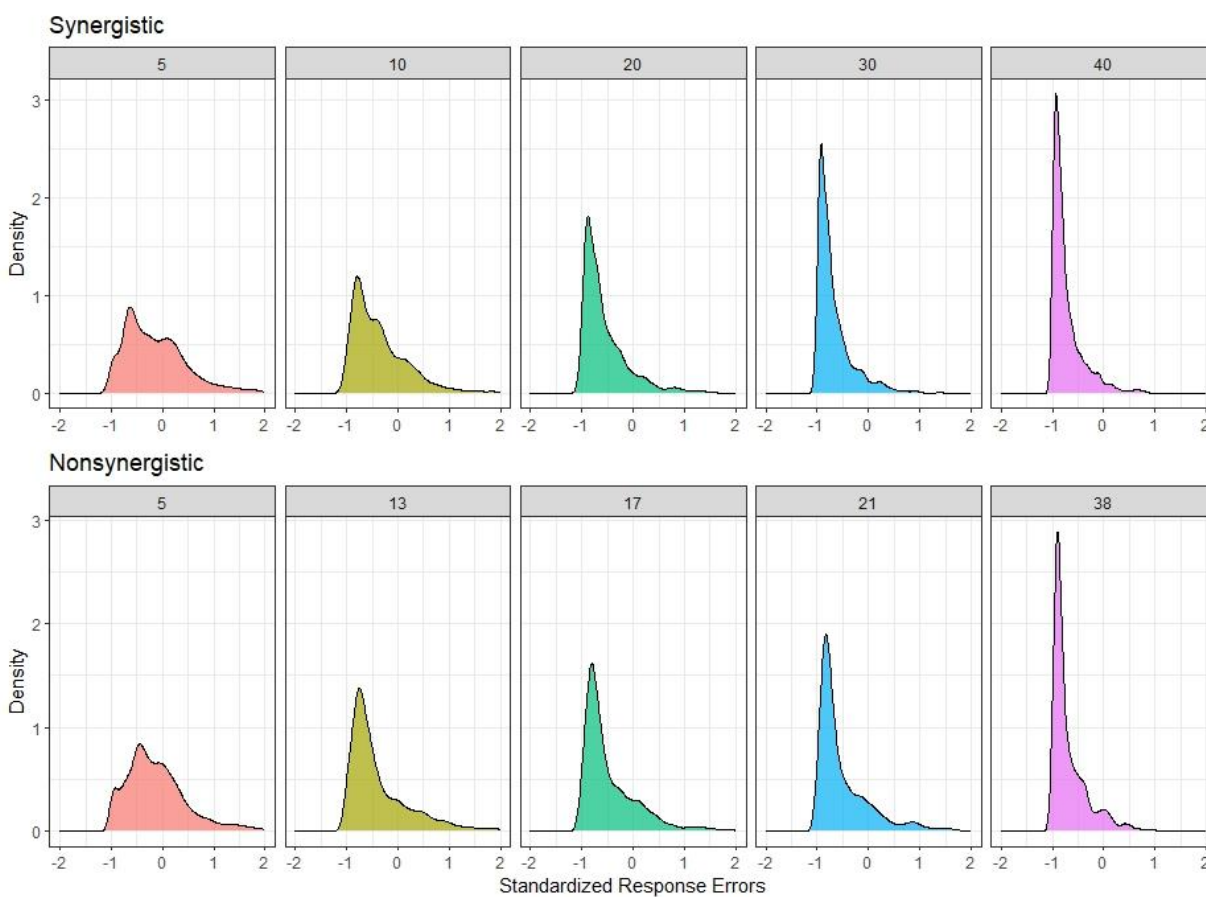
Note. These results show the time within each interval that a response occurred, rather than the interresponse times. Therefore, response times >80-s were retained to show the complete response profile.

Appendix B - Response Error Graphs

Appendix Figure B.1 Raw Response Errors Show Distributions with Consistent Early Responses to Each Interval



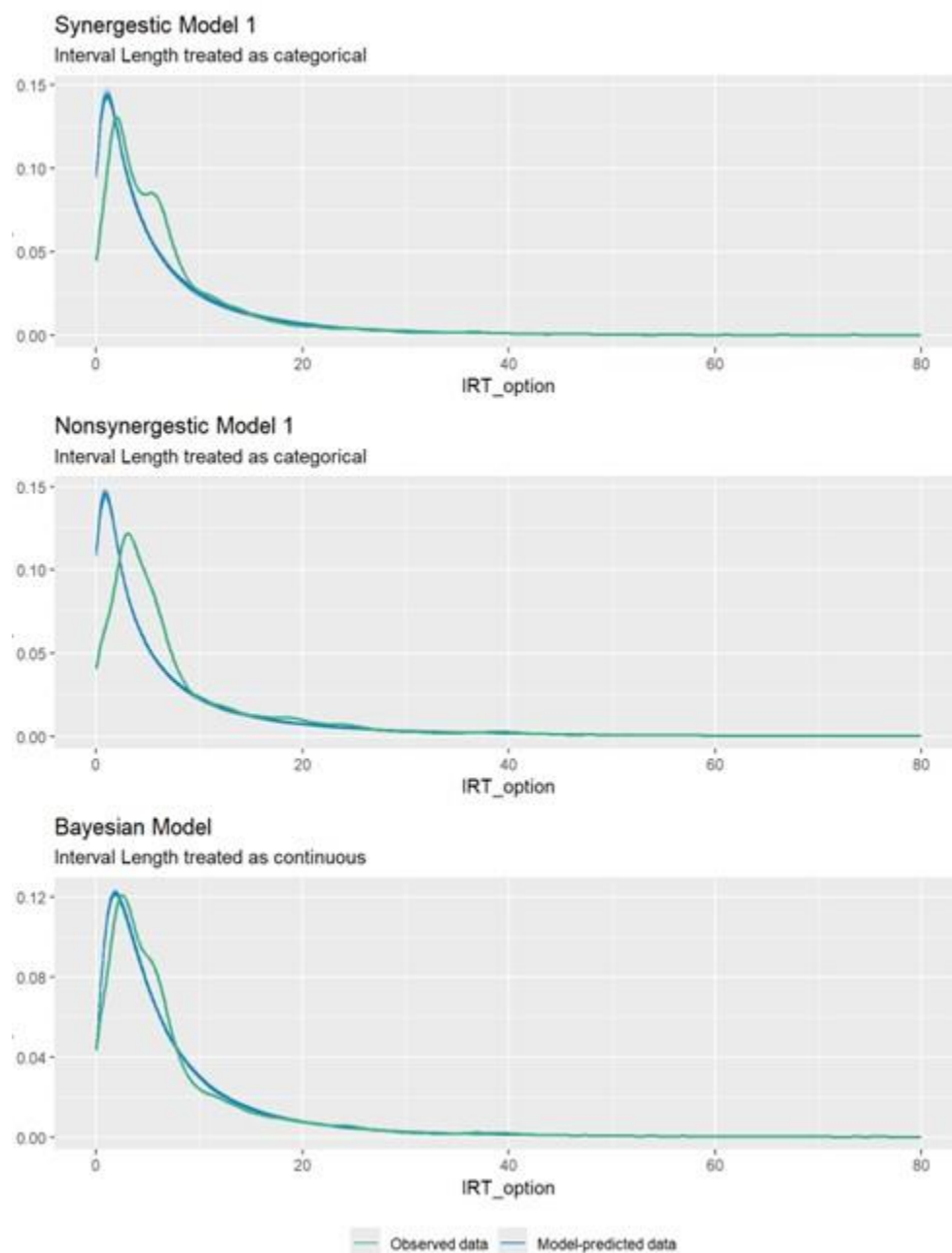
Appendix Figure B.2 Standardized Response Error Plot Support Consistent Early Responding



Note. Standardized response errors were calculated as the raw response error divided by the interval length.

Appendix C - Model Assumption Checks

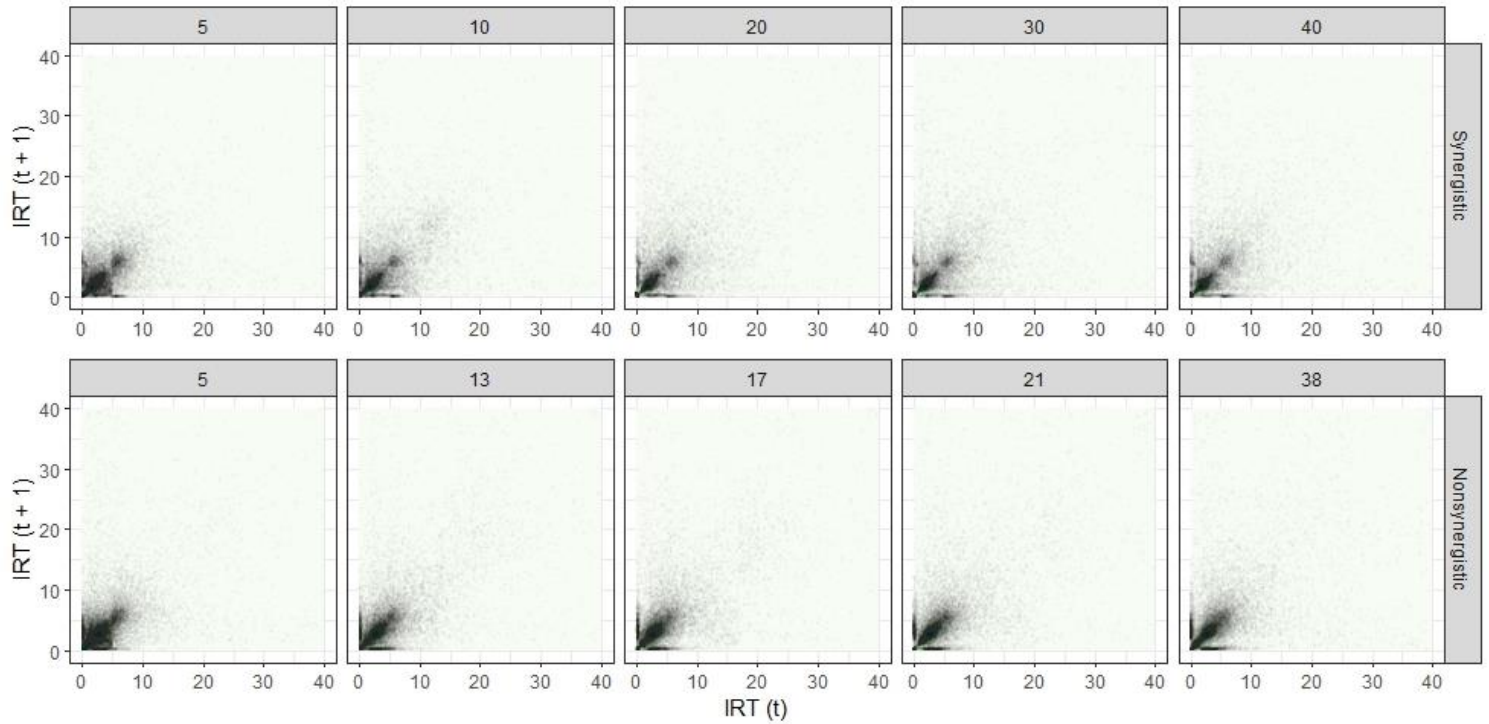
Appendix Figure C.1 Posterior Predictive Checks From Each IRT Model



Note. These graphs illustrate superior model fits to the observed data when Bayesian analyses are used and Interval Length is treated continuously.

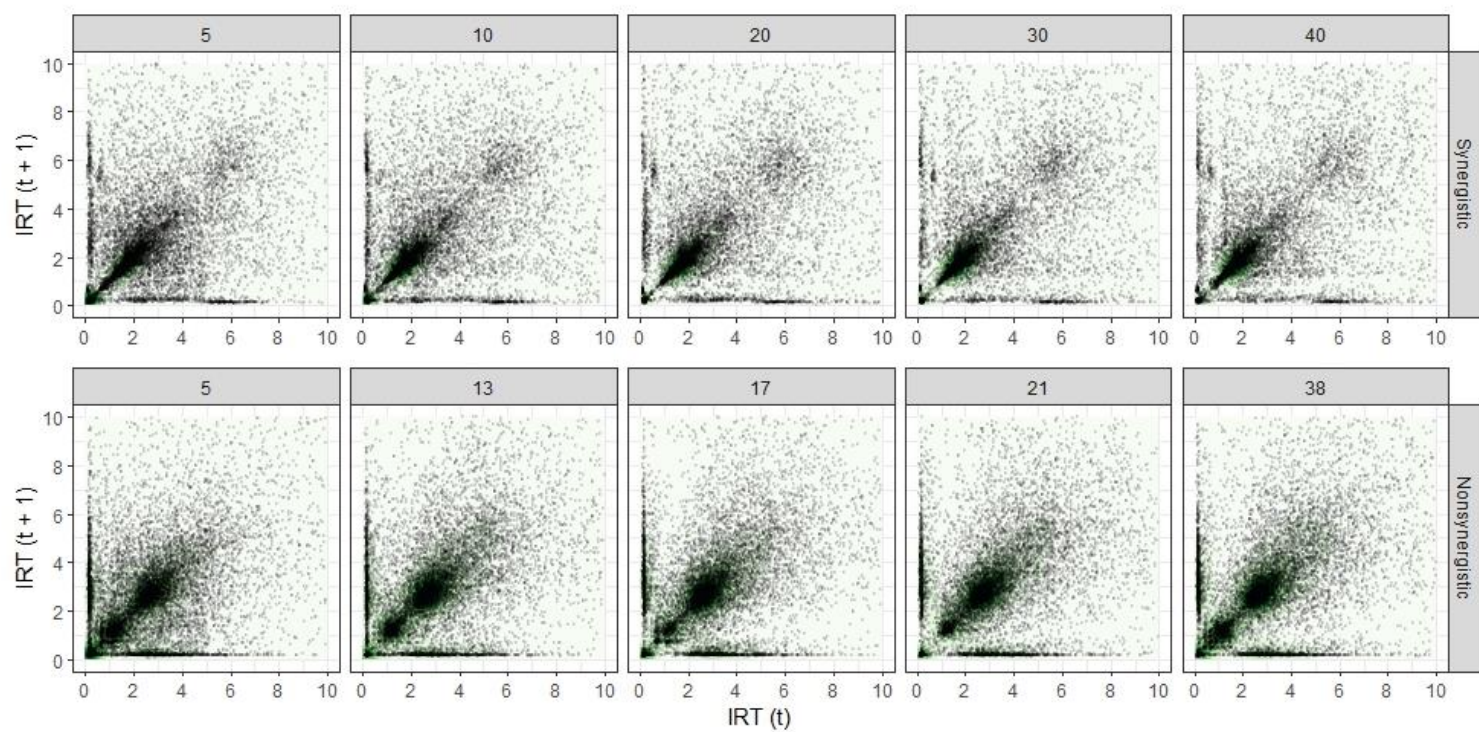
Appendix D - Return Plots

Appendix Figure D.1 Full Return Plots for Each Interval

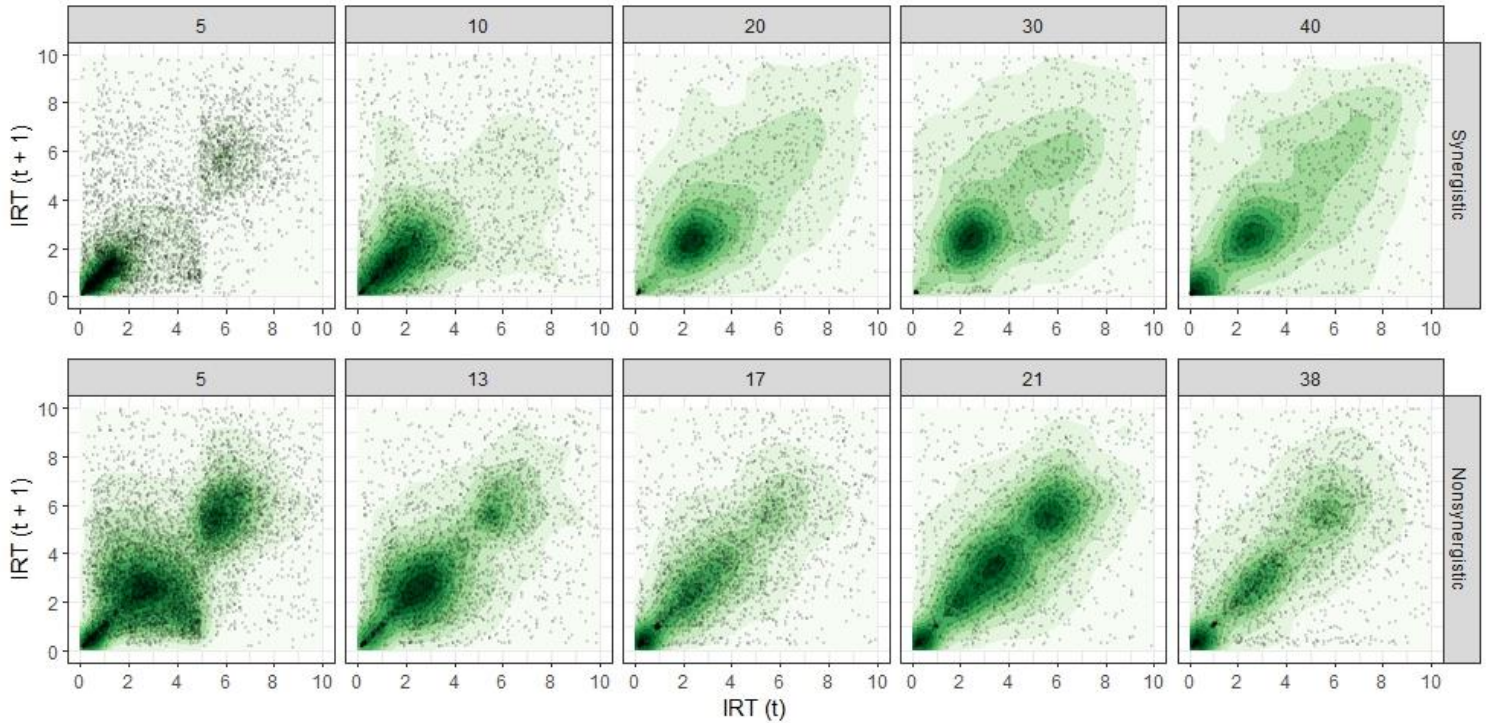


Note. Return map of consecutive inter-response times (IRTs) by target interval. Each panel shows the IRT at time t plotted against the following IRT ($t + 1$) to the same interval. Each point represents one IRT pair. Darker shading indicate higher densities of IRT pairs.

Appendix Figure D.2 Steady Responder Return Plots



Appendix Figure D.3 Interval Tracker Return Plots



These graphs plot the relationship between IRT pairs, by plotting an IRT at time t (x-axis) against the next IRT to that same option at time $t + 1$ (y-axis). This plot contains all IRT data, not just the time between a rewarded response, and the next response to that same interval (i.e., post-reinforcement pauses). IRTs beyond 80-s were removed as these IRTs most likely represent a lack of attention from the participant. Here we plot all IRTs below 80-s to see how consistent participants tend to respond to a given interval.

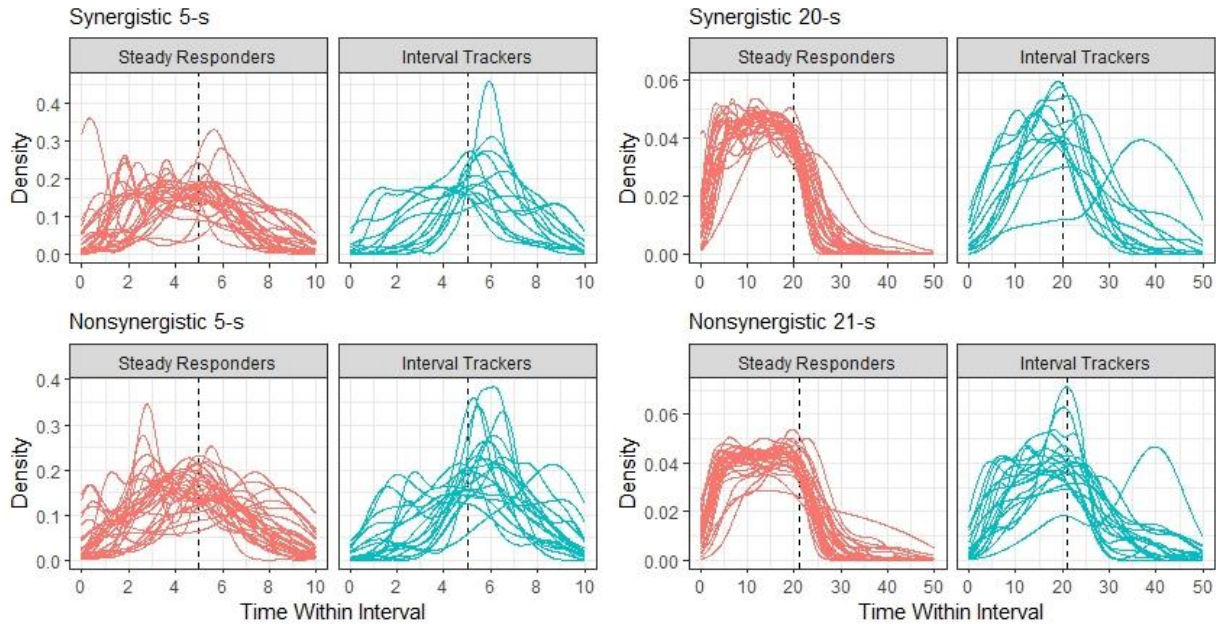
Data points clustered along the diagonal line indicate consistent IRT durations (e.g., a 3-s IRT, followed by a 3-s IRT). Data points clustered along the x-axis (when $y = 0$), suggest that participants quickly returned to that same key (e.g., a 5-s IRT is followed by a 0.1-s IRT). This response pattern indicates persistence, where the participant is likely aware that the interval will pay off soon, so they respond rapidly (presumably until a reward is delivered). Data points

clustered along the y-axis (when $x = 0$) indicate that a reward was likely earned, and the participant is waiting longer before returning (e.g., a 0.1-s IRT is followed by a 5-s IRT).

Appendix Figures D.2 and D.3 show that consistent timing, where IRT pairs share the same duration (data points along the diagonal), is rather common across all intervals regardless of Experiment or Cluster. When results are compared across Cluster, we can see that Steady Responders show much more varied durations among IRT pairs, where data points cluster along either axis. In fact, if interval trackers were timing each interval with some moderate degree of accuracy, we anticipated that these individuals would be more likely to show persistence and return to the intervals soon after if a reward was not obtained. Steady Responders, on the other hand, were anticipated to have much more consistent durations of IRT pairs, and subsequently less persistence. Our results indicate the opposite conclusion that patterns of persistence were much more common amongst Steady Responders rather than Interval Trackers.

Appendix E - Supplemental Individual Subject Plots

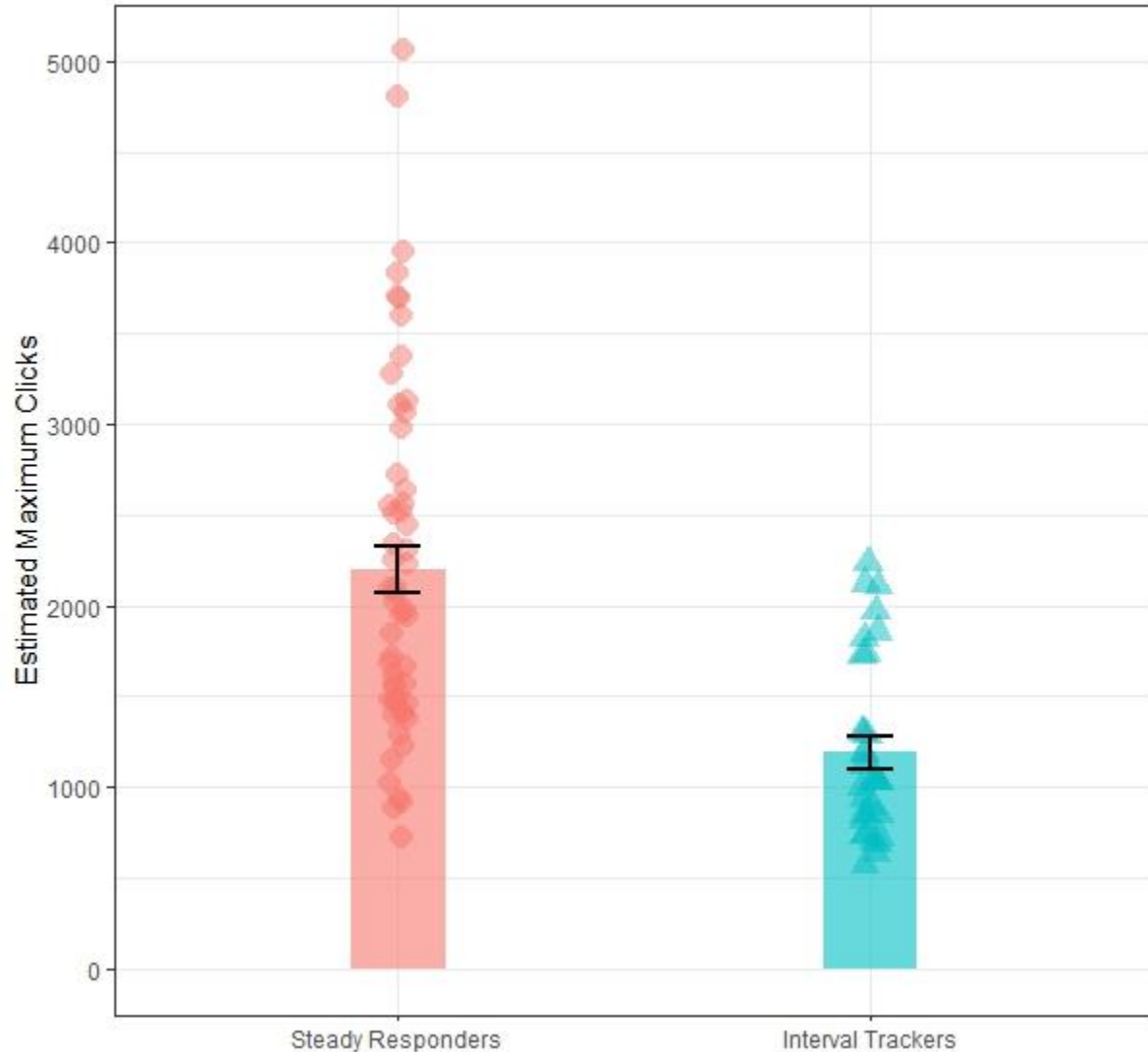
Appendix Figure E.1 Individual Participant Data Confirms Scalar Timing Properties in Interval Trackers Cluster



Note. Two example intervals are shown above. Each line represents an individual participant. These intervals were selected as their durations matched closely across the two experiments.

Appendix F - Total Clicks Plot

Appendix Figure F.1 Steady Responders Produced Significantly More Clicks Compared to Interval Trackers



The distribution of maximum clicks was positively skewed and was modeled using a gamma regression with a log link function, with Cluster (effect-coded) as the only predictor. The omnibus ANOVA test revealed a significant difference in total click counts between Clusters ($\chi^2 = 37, p < .0001$), suggesting that the Steady Responders produced significantly more clicks within the experiment.

Appendix G - Reward and Response Tables and Figure

Appendix Table G.1 Experiment 1 (Synergistic Intervals)

Cluster	Participant	5-s		10-s		20-s		30-s		40-s	
		Resp	Rwds	Resp	Rwds	Resp	Rwds	Resp	Rwds	Resp	Rwds
Steady Responders	0183	443	221	306	127	271	68	270	47	277	37
Steady Responders	0671	313	218	318	129	238	69	202	47	221	37
Steady Responders	1311	638	196	774	124	622	77	640	55	696	43
Steady Responders	1421	511	229	336	125	323	66	253	45	268	35
Steady Responders	1729	261	191	341	134	292	78	271	55	240	42
Steady Responders	2146	375	216	280	127	321	71	211	48	220	38
Steady Responders	2838	316	220	279	125	256	69	264	47	284	39
Steady Responders	2932	530	223	354	120	291	68	355	51	450	38
Steady Responders	3829	271	223	232	119	253	71	238	49	235	38
Steady Responders	3860	533	195	545	124	528	76	574	54	533	41
Steady Responders	4261	226	178	175	130	170	83	185	60	182	49
Steady Responders	5103	811	225	910	122	689	68	768	48	657	37
Steady Responders	5354	597	210	617	125	623	70	600	50	619	40
Steady Responders	5378	619	225	531	125	589	69	340	46	365	35
Steady Responders	5637	547	205	587	136	286	69	312	50	334	40

Steady Responders	6002	508	226	504	127	487	66	519	46	535	35
Steady Responders	6119	486	207	379	126	378	76	376	53	226	38
Steady Responders	7111	551	210	490	126	461	71	445	53	362	40
Steady Responders	7664	799	219	814	126	789	70	782	49	765	36
Steady Responders	7828	668	229	666	126	674	66	632	45	638	34
Steady Responders	7975	1,663	231	554	115	399	67	531	50	551	37
Steady Responders	8117	736	235	731	123	763	65	744	44	736	33
Steady Responders	8385	457	196	449	129	457	77	429	56	436	42
Steady Responders	8818	409	238	288	123	245	61	285	44	259	34
Steady Responders	9086	322	179	313	130	333	82	337	62	322	47
Steady Responders	9127	374	209	346	128	335	72	332	51	330	40
Steady Responders	9575	623	229	617	126	607	66	549	45	578	34
Interval Trackers	0412	427	219	318	120	345	71	360	51	286	39
Interval Trackers	0827	336	233	283	123	160	63	174	46	166	35
Interval Trackers	2237	170	155	133	109	156	95	132	71	134	58
Interval Trackers	2718	271	204	164	124	135	75	125	55	120	42
Interval Trackers	2809	619	241	247	120	149	61	158	45	135	33
Interval Trackers	3074	295	272	150	134	38	31	47	31	24	15
Interval Trackers	6388	297	233	170	117	166	67	149	47	159	36
Interval Trackers	6455	800	236	575	125	264	63	248	42	240	34

Interval Trackers	8568	276	220	191	131	168	70	118	44	99	35
Interval Trackers	8804	478	184	440	126	318	76	301	58	421	46
Interval Trackers	9265	949	242	335	106	191	62	220	42	154	37
Interval Trackers	9878	284	245	147	119	99	61	82	39	73	35
Interval Trackers	9900	218	171	232	128	206	83	202	66	172	51

Note. Resp = Responses. Rwd = Rewards.

Appendix Table G.2 Experiment 2 (Nonsynergistic Intervals)

Cluster	Participant	5-s		13-s		17-s		21-s		38-s	
		Resp	Rwds	Resp	Rwds	Resp	Rwds	Resp	Rwds	Resp	Rwds
Steady Responders	0585	346	217	303	99	308	79	305	67	302	38
Steady Responders	0596	302	192	263	102	259	85	266	72	277	44
Steady Responders	1164	196	165	218	111	243	97	235	80	258	47
Steady Responders	1359	413	210	376	100	413	82	369	68	383	40
Steady Responders	1995	179	89	173	64	152	58	184	55	182	37
Steady Responders	2544	287	202	302	105	306	83	301	69	296	41
Steady Responders	2647	352	202	359	102	317	82	324	70	315	43
Steady Responders	2898	345	224	311	99	285	77	245	62	281	38
Steady Responders	3009	698	225	537	97	547	78	466	64	387	36
Steady Responders	3826	445	213	409	102	418	80	423	66	555	39
Steady Responders	4334	327	165	262	95	276	88	296	77	284	50
Steady Responders	4374	286	162	294	111	318	93	267	75	309	52
Steady Responders	4593	186	160	182	115	167	93	167	82	185	50
Steady Responders	4842	500	226	509	99	504	76	504	63	499	36
Steady Responders	5003	722	233	419	93	295	70	328	63	324	36
Steady Responders	5845	136	123	123	90	108	73	138	77	135	56

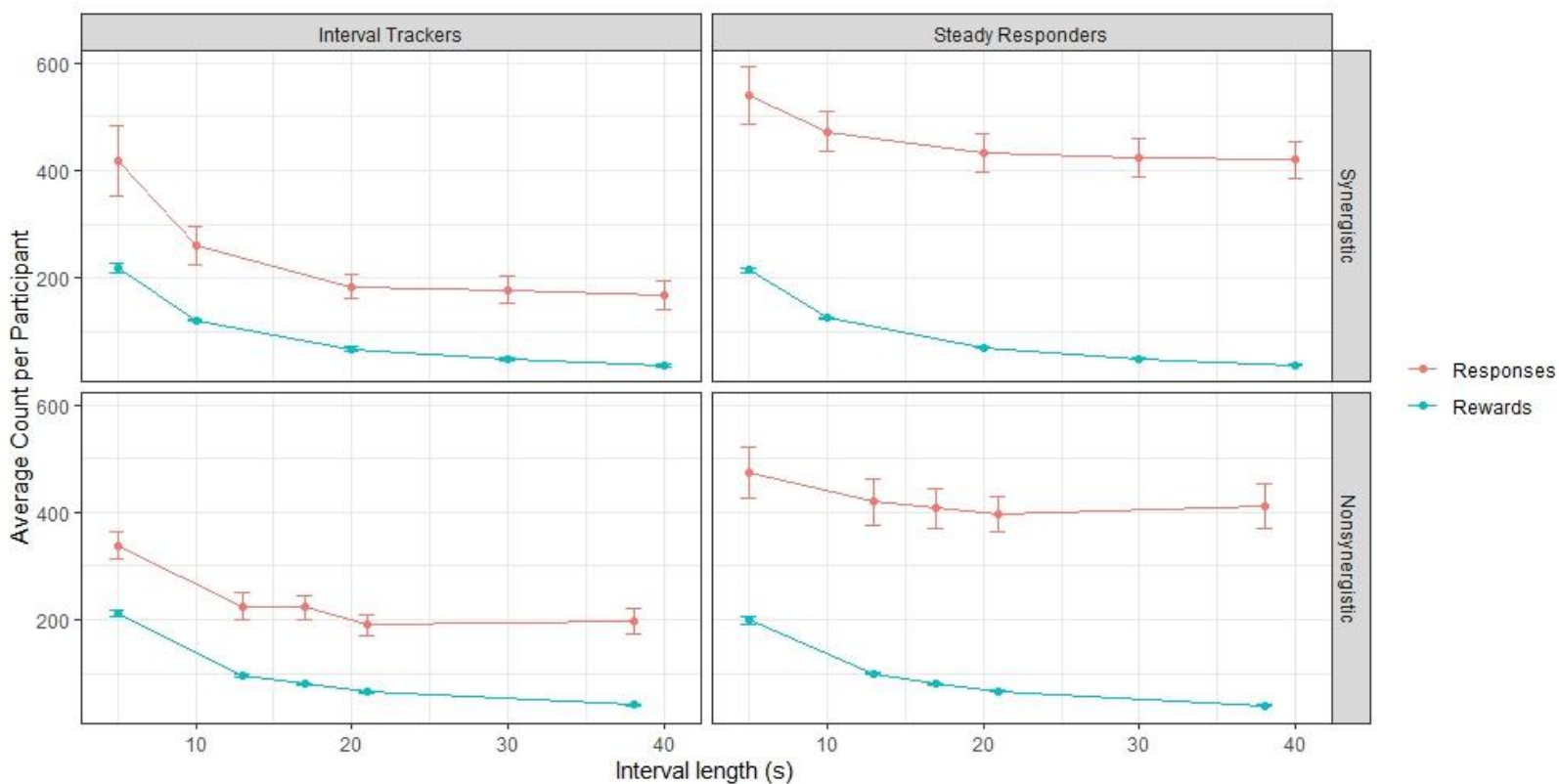
Steady Responders	5878	182	149	217	117	179	96	221	85	222	53
Steady Responders	6349	402	210	434	103	429	81	431	67	403	39
Steady Responders	6405	518	230	504	97	519	75	510	63	513	35
Steady Responders	6519	456	214	376	101	451	81	370	66	367	38
Steady Responders	6597	453	216	373	101	399	80	355	64	365	39
Steady Responders	6697	530	218	479	99	478	80	426	65	423	38
Steady Responders	7258	993	222	1,145	100	853	78	837	63	978	37
Steady Responders	7423	880	218	684	99	655	78	663	66	715	39
Steady Responders	8941	873	221	594	99	588	78	602	65	469	37
Steady Responders	9089	1,120	231	1,002	98	984	75	816	61	1,134	35
Steady Responders	9289	396	209	279	96	289	80	312	69	282	41
Steady Responders	9421	624	227	506	95	433	77	520	64	424	37
Steady Responders	9936	607	213	518	101	651	81	629	67	697	38
Interval Trackers	2475	502	206	421	100	309	78	268	67	222	39
Interval Trackers	2693	240	215	115	100	155	96	57	47	72	42
Interval Trackers	2731	368	197	428	102	365	83	314	71	345	47
Interval Trackers	3671	310	226	204	98	160	76	151	62	171	38
Interval Trackers	3719	245	212	167	97	161	80	159	69	165	42
Interval Trackers	4509	158	151	152	115	144	99	134	85	108	49
Interval Trackers	6518	425	252	143	81	160	71	138	59	163	37

Interval Trackers	7934	288	247	116	82	134	74	96	55	107	39
Interval Trackers	8040	407	239	219	91	265	74	209	61	189	35
Interval Trackers	9005	382	182	206	102	274	92	220	75	214	49
Interval Trackers	9038	572	206	420	102	458	81	347	64	436	42
Interval Trackers	9110	519	226	370	96	378	78	399	63	445	37
Interval Trackers	9182	359	216	187	86	215	78	197	68	219	43
Interval Trackers	9351	233	196	113	87	116	81	122	71	124	48
Interval Trackers	9505	296	222	230	100	177	76	165	63	197	39
Interval Trackers	9510	255	181	161	99	141	86	153	81	144	53
Interval Trackers	9573	233	207	161	100	179	86	164	66	106	36
Interval Trackers	9623	275	222	245	97	277	81	195	62	199	38
Interval Trackers	9692	366	237	229	95	174	75	136	57	131	36

Note. Resp = Responses. Rwd = Rewards.

Data from Appendix Tables G.1 and G.2 suggest sufficient contact with all reinforcement contingencies were made across participants.

Appendix Figure G.1 Mean Responses and Rewards by Interval Length, Cluster, and Condition



Note. This graph represents the mean values of all participants, based on data from Appendix Tables G.1 and G.2. Error bars represent ± 1 standard error.