

Sources of Variability and Systematic Error in Mouse Timing

Behavior

C. R. Gallistel*⁺, Adam King[†], Robert McDonald*

¹Abstract

With a view toward screening mice for genetically and pharmacologically induced distortions in temporal memory, we investigated the source of scalar error and scalar variability in the peak procedure. With head-poking as our response, we used target intervals of 5, 15 and 45 s in the standard procedure and in a variant in which there were multiple target temporal intervals in a single trial (the tripeak procedure). The observed pattern of changes in start and stop times and their variance and covariance imply that the decision criteria for starting to respond and ceasing to respond on a given trial are drawn from independent distributions that are not located equidistant from the time the subject expects responding to result in food delivery. The cross-target covariance implicates trial-to-trial variation in clock speed as the major source of scalar trial-to-trial variability in start and stop times. Thus, the systematic (often scalar) error in peak behavior is due largely if not entirely to the asymmetric location of start and stop decision criteria relative to the target time, and the scalar variability in starts and stops derives primarily from sources other than memory. The peak procedure cannot identify miscalibrations of memory unless they are such as to make the response interval miss the target time on most trials.

KEYWORDS

mice / peak procedure / scalar error / scalar variability / decision criteria / memory error / memory variability / clock variability

¹ *Department of Psychology
and Rutgers Center for Cognitive Science
Rutgers University
Piscataway, NJ 08854

[†]Fairfield University
Department of Mathematics and Computer Science

Identification of the physiological and molecular mechanisms underlying memory is one of the major challenges in neuroscience. One promising avenue of research involves the identification of genetic mutations that affect the mechanism(s) of memory. The development of a mouse strain that bred true for a mutation that altered a basic component of a molecular mechanism of memory would allow the identification of the location of the affected gene(s), and the location and action of their products.

Given that the content and processes of memory are currently only knowable through their behavioral consequences, it is desirable to develop behavioral screening techniques capable of measuring parameters of the memory mechanism itself. Broadly construed, memory probably comprises several distinct mechanisms. Processes on the input side (learning processes) and on the retrieval side (performance processes) may differ depending on the information to be acquired or used (Gallistel, 1980; 1990; 1999). However, by analogy to other information-storage systems (Gallistel, 2002), there may be a single, universal molecular mechanism that carries acquired information forward in time, just as there is a universal molecular mechanism for carrying hereditary information forward in time. Be that as it may, the focus of the screens we wish to develop is on the mechanism that carries simple quantitative information forward in time in mammals-- information about the duration of a temporal interval or about the distance and direction of a spatial location, the kind of information that is naturally represented by real numbers. We want to develop screens that distinguish between malfunction in that mechanism, on the one hand, and, on the other hand, alteration in the processes that extract from experience the information that goes into that memory (learning processes) or alteration in processes that translate the stored information into observable behavior (performance processes).

Genetic and/or pharmacological manipulations may affect measures of the strength, speed or frequency of a learned behavior through alteration of any of the three kinds of processes (learning, memory, and performance). Thus, measures such as the latency to reach the platform in the Morris water maze are not promising candidates for the kind of screen we seek. An alternative to measuring the strength of learned behavior is to screen for behaviorally manifest distortions or degradations in the content of memory (Church & Meck, 1988). If the mechanism that carries quantitative information forward in time delivers back values that differ systematically from those put in, then it introduces systematic error. If it delivers values that are more variable than the values put in, then it degrades the information. The question we pose is whether the systematic errors and trial-to-trial variability in timed behavior come in appreciable measure from the memory mechanism, as just delimited.

The probability of conditioned behavior often peaks at approximately the time at which reinforcement is expected, (Bevins, et al., 1997; Bolles & Moot, 1973; LaBarbera & Church, 1974; Roberts & Church, 1978; White, et al., 2000), which implies that information about the reinforcement latency has been extracted from past experience and carried forward in memory (Fantino, 1969; Gibbon, Church & Meck, 1984; Killeen, 1975; Savastano & Miller, 1998). The timing of the behavior commonly exhibits both systematic error in the mean and trial-to-trial variability (Gibbon, et al. 1984), and these parameters of the behavior differ from subject to subject (Gallistel & Gibbon, 2000), suggesting their utility in genetic screening.

The most frequently used, and most thoroughly analyzed, experimental paradigm for the study of temporal memory is the peak procedure (Catania, 1970; Cheng & Westwood, 1993; Cheng, Westwood, & Crystal, 1993; Church, Meck, & Gibbon, 1994; Roberts & Church, 1978). In this procedure, responding delivers reward at a fixed interval after the onset of a warning signal, commonly called the conditioned stimulus (CS). The duration of the interval between CS onset and the time after which responding produces food is called the target time. Typically, subjects begin to respond well before the target time, suggesting that there is some noise (hence, uncertainty) in their representation of the target time. On some trials, called probe trials, the reward is not delivered. On these trials, subjects cease responding at some point after the target time has past. Thus, responding on probe trials usually brackets the target time.

Importantly, the distribution of responding in time—when subjects start and when subjects stop relative to the target time—appears to be unaffected by the motivational state of the subject. By contrast, the strength (frequency) of responding in this same task depends strongly on motivation (Roberts, 1981). This is one example of a performance factor that has a large effect on the strength of the behavior but little effect on a measure of the contents of memory.

Figure 1 shows data from a mouse run in our version of the peak procedure (King, McDonald & Gallistel, 2001). The CS was the illumination of the feeding station in a standard mouse testing box. The presence of the mouse's head in the station triggered the release of a food pellet at or after the timing out of the programmed delay (the target time). The same subject was tested in blocks of multiple sessions at each of three different target times—10, 20 and 30 s.

On any given probe trial, the mouse poked its head into the feeding station when some proportion of the target time had elapsed, kept it there with few and short interruptions for a period that usually bracketed the target time, then withdrew its head, and poked only rarely and briefly thereafter. The interval during which its head was in the station most of the time corresponded to the "run" of high frequency discrete responses (lever presses or key pecks) identified in the pioneering work of Church, Meck and Gibbon (1994) on the analysis of individual trials in the peak procedure. In the mouse, as in the pigeon and the rat, the starts and stops of the run varied from trial to trial. Averaging over many trials, the probability that a run was in progress rose to nearly unity when some proportion of the target time had elapsed, then returned to near zero as the duration of the probe trial became considerably greater than the target time. The peak curves in Figure 1 are the cumulative distribution of run starts (rising phase) minus the cumulative distribution of stops (falling phase).

The peak curves from sessions with different target times very nearly superpose when plotted against the elapsed proportion of the target time (normalized trial time). This means that both the locations (central tendencies) and the spread (standard deviations) of the start and stop distributions were nearly fixed proportions of the target times. Scalar variability—variability proportionate to the target time—is a ubiquitous feature of timing data (Gibbon, 1977; Gibbon, 1991; Killeen & Weiss, 1987). Secondly, the superposed plots are not centered about the target time (which is at 1 on the normalized time axis). This means that the average center of a run, the point half way between the start and the stop, deviated significantly from the target time. This is a systematic error in the timing of the behavior. This systematic error was also proportional

to the target latency, a phenomenon known as scalar error. The distinction between systematic error and trial-to-trial variability in timing behavior is the same as the distinction between systematic error and random error in, for example, the targeting of a rifle. It is the distinction between the location of the mean relative to the target (the systematic error) and the dispersion in the shots about that mean (the variability). In our case, it is the distinction between the location of the average center of a run relative to the target time and the standard deviation of the starts and stops of the runs.

The scalar variability in run starts and stops and the scalar error in run centers have been attributed to the memory for target time (Gibbon, et al., 1984; Meck & Church, 1987). If this attribution can be strengthened, these easily measured aspects of timed behavior could become the basis for tests designed to detect genetically induced alterations in the properties of the memory mechanism.

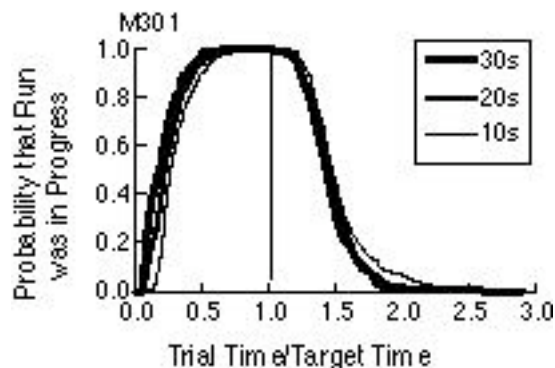


Figure 1. *Probability that a run was in progress versus normalized trial time in an individual mouse, performing in a version of the peak procedure. The target times in different blocks of sessions were 10, 20 and 30 s. These plots are called peak curves. The rising phase is the cumulative probability that the run had begun; the falling phase is the inverse of the cumulative probability that it had ended. The cumulative distributions of run starts and run stops from sessions with different target times superpose when plotted against normalized trial time. This means that the trial-to-trial variability in starts and stops is proportional to the target time (scalar variability). The stop distributions (trailing edge of the peaks) lie proportionately closer to the target time (vertical line at 1) than do the start distributions (leading edge), which means that the average center of a run occurred substantially earlier than the target time (systematic error). The systematic error was also scalar (proportionate to target time). (Reproduced from King, et al., 2001, by permission of the publisher.)*

An alternative, non-mnemonic^[PDI] source of the trial-to-trial scalar variability in start and stop proportions would be trial-to-trial variability in the relation between objective duration and subjective duration, such as would be produced by trial-to-trial variation in the speed of the timer. To the extent that this is true, the behaviorally measured variability is the result of variability in an information acquisition (learning) process (the timing of intervals) rather than variability in the memory mechanism, which preserves the results of past timings for later behavioral use.

Gibbon and Church(1992) realized that an analysis of variance and covariance in the statistics of runs was a powerful tool for identifying the sources of the behavioral variability. Our work follows up on their insight and their analyses. If the variability in

starts and stops derives from memory noise, then samples from two different target memories should not covary. The trial-to-trial variations in the values read from one target memory (say, the memory for feedings that have occurred 5 s into a trial) should not predict the trial-to-trial variations in the values read from different target memories (the memories for feedings that have occurred 15 or 45 s into a trial). A late start on the run at the first target (5 s) should predict a late stop on that same run, because both the start and the stop of a given run are based on the same value read from memory, but it should not predict a late stop on the run at a later target in the same trial, because that target derives from the reading of a different memory (the memory for the target for that later run). Memory noise (variability) is equivalent in its effects on start and stop times to the effect of changing the locations of a bus stop on the arrival time of a bus. Moving the location earlier in the route causes the bus to arrive at that stop earlier; moving it farther along the route causes the bus to arrive later. But movements in the location of one stop should not affect arrival times at other stops.

On the other hand, if the trial-to-trial variability in starts and stops arises from trial-to-trial variability in how fast the timer runs, a late start on the run at the short latency target (due to a slower than usual timing) will predict that all subsequent starts and stops will also be late. To continue the bus analogy, if the bus is slow en route to the first stop, hence late getting to it, it will tend to be late at all the subsequent stops as well. It is this consideration that led us to adopt the tripeak procedure, which has three target times, all three of which control timed starts and stops on tripeak trials. By comparing within- and cross-target covariances in starts and stops, we estimate the extent to which trial-to-trial variability depends on variability in the timer (bus speed) rather than variability in the remembered targets (the locations of the bus's destinations).

An alternative, non-mnemonic source of the scalar error in the average center of a run would be systematic differences in how far from the target time the subject placed its start and stop criteria. Following Gibbon (1977), we assume that starts and stops are the results of a decision process in which the subject compares its constantly growing estimate of the time elapsed on this trial to the remembered target time. It starts to respond when the proportion between the two exceeds a critical value (the start criterion) and stops when it exceeds a higher one (the stop criterion). If a subject adopts a stop criterion proportionately closer to the target time than their start criterion, the average center of a run will be short of the target time. To the extent that systematic error is due to the asymmetric placement of start and stop criteria, it reflects performance factors rather than the miscalibration of memory.

In work leading to the suggestion that scalar error in memory reflected a miscalibration of memory, it was assumed that the placement of the start and stop criteria relative to the remembered target time was determined by a single quantity (Gibbon, et al., 1984; see also Cheng, 1992). We call this quantity the hedge factor, on the decision-theoretic assumption that the distance away from the remembered target time at which subjects place their start and stop criteria reflects the degree of their uncertainty about when the target is reached. The greater their uncertainty is, the greater is the amount by which they hedge their starts and stops by placing their decision criteria farther from the remembered target. If there is only one hedge factor, then any manipulation that moves the start criterion closer to the target time will likewise move the stop criterion closer to it, and by the same amount (cf Cheng, 1992). If on the other hand, subjects determine

their start and stop criteria independently, then a manipulation might cause them to move in different directions relative to the target time and/or move by different amounts. This consideration was our reason for running the same subjects with the same three target times first in the single-peak procedure and then in the tri-peak procedure. We thought that if start and stop criteria are independently placeable, then the shift from the single-peak procedure to the tri-peak procedure might cause asymmetric displacements of the start and stop criteria. This would be evidence that systematic errors in the single-peak data were likely to be the result of the asymmetric placement of decision criteria.

We first ran six mice in the peak procedure using an unusually broad spread of target intervals (5, 15, and 45 s—a 9-fold range). Then we ran the same subjects in a variant of the peak procedure similar to one first used by Fetterman and Killeen (1995), in which multiple target times are used (the tripeak procedure, see Matell, King and Meck (submitted) for an earlier use of this procedure with mice). In our version of the tripeak procedure, the subjects were presented with three illuminated feeding stations on each trial, only one of which paid off. When a feeding station paid off, it did so at the target time associated with that station throughout testing. Subjects, however, did not know which station would pay off on any given trial. Therefore, we expected them to sample them in the order of their target times, poking first in the 5 s station, then, if that failed, in the 15 s station, and finally, if that also failed, in the 45 s station.

Method

Subjects

The subjects were six female Swiss-Webster ND4 mice (Harlan, Indianapolis, IN), aged about 12 weeks and weighing between 19 and 22 g on the first day of the experiment. They were individually housed in plastic tubs, and maintained on a 12:12 h photoperiod, with lights on at 2200 h. Behavioral testing occurred during the dark phase of the photoperiod. Starting on the first day of the study, access to food was restricted. The amount of food obtained during the experimental sessions was supplemented at the end of each session to maintain body weight at approximately 90% of the free-feeding weight. Water was available ad lib. in both the home cage and the experimental chambers.

Apparatus

Experimental sessions took place in modular operant chambers (Med Associates, Georgia, VT, model # ENV307AW, 7 x 8 x 5 in), located in individual ventilated, sound-attenuating boxes. Each chamber was equipped with three pellet dispensers, each connected to one of three feedings station along one wall of the chamber (designated F1, F2, and F3 from left to right). A control station (identical to the feeding stations, but not connected to a pellet dispenser) was located on the opposite wall. The stations were cubic hoppers 2.4 cm on a side equipped with an infrared (IR) beam that detected nose pokes, and with a light that illuminated the interior of the hopper. The chambers were also equipped with a tone generator (2900 Hz) and a white noise generator (80 dB, flat 10-25,000 Hz). If the subject's head was already in the station when the program armed the feeder, the delivery of a 20 mg pellet (Research Diets, #PJA1 00020) was immediate; otherwise, it occurred at the onset of the next interruption. All aspects of the experimental

protocol were controlled by computer software (Med-PC, Med Associates), which also logged and time-stamped the events—the onsets and offsets of interruptions of the IR beams in the stations, the onsets and offsets of tones, noises, and station illuminations, and the delivery of food pellets. Event times were recorded with a resolution of 20 ms.

Procedure

Subjects were weighed at the start of each session, then placed in the operant chambers. During single peak procedure sessions (Phases 1-4) the opportunity to begin a trial was signaled by the illumination of the control station and the onset of white noise. These stimuli remained present until the subject poked its nose into the control station. When the poke was detected, the control station illumination and white noise terminated, the target feeding station was illuminated, the tone turned on, and the feeding clock started.

A random 4 out of 5 trials were feeding trials, on which the feeder was armed when the feeding clock reached the target latency (T^*). A single pellet was delivered at T^* if the subject's head was in the feeding station (interrupting the IR beam) when arming occurred; otherwise, the first poke thereafter triggered its delivery. The tone terminated immediately following pellet delivery. The illumination of the feeding station terminated 5 s later. When the mouse did not interrupt the IR beam in the feeding station at or after T^* , the trial ended after $3T^*$ plus an interval chosen from an exponential distribution with an expectation of T^* . Probe trials, when no food was delivered regardless of behavior at or after T^* , constituted a random 1 out of 5 trials. On these trials, the feeding station remained illuminated and the tone remained on for an interval equal to $2T^*$ plus an interval drawn from an exponential distribution with an expectation equal to T^* .

Following each trial, there was an ITI equal to $2T^*$ plus a random interval chosen from an exponential distribution with an expectation of T^* . At the end of the intertrial interval, the control station was again illuminated and the white noise turned back on, enabling the mouse to start the next trial.

Sessions lasted 10 hours, because these analyses require a great many trials, and we wanted to maximize the number of trials per session. Because the sessions were so long, there was a substantial chance that the mouse might go to sleep for parts of them. For that reason, we made the trials self-paced—the mouse initiated each trial. A second motivation for having the trials initiated by the mouse poking into a separate station (the control station) was that it guaranteed that a poke was not in progress at the start of a trial. Moreover, it fixed (controlled) the mouse's position at the start of each trial, which is important when the timing of the first response is a variable of interest. The number of trials in a session was largely determined by the subject, ranging from 75 to 250.

On both feeding and probe trials, we extracted from the recorded onsets and offsets of beam interruptions in the feeding station the sequence of head-out and head-in intervals. We refer to the head-in intervals as pokes.

The design of the experiment is given in Table 1. In Phase 1, which ran for 20 days, the target time for half the subjects (Group A) was 15 s and the feeding station was F1; for the remaining subjects (Group B), the target time was 45 s and the feeding station

was F2. In Phase 2, which ran for 12 days, subjects who had the 15 s time and the F1 station in Phase 1, now had the 45 s time and F2, and vice versa for the other half of the subjects. In Phase 3, which ran for 8 days, the target time was 5 s and the feeding station was F3. In Phase 4, which ran for 14 days, the target time was again 45 s and the feeding station was again F2. Note that each of the three feeding stations was associated with a different feeding time.

In Phase 5, a poke in the illuminated control station led to the illumination of a randomly determined feeding station. The association between a feeding station and its target time was maintained: The illumination of F1 signaled the availability of food at that feeding station at a 15 s target latency, the illumination of F2 signaled food there after 45 s, and the illumination of F3 signaled food there after 5 s. There were no probe trials. As in earlier phases, when no poke occurred at or after T^* , the trial ended after $3T^*$ plus $E(T^*)$. The ITI was again equal to $2T^*$ plus $E(T^*)$.

After 6 days in Phase 5, all subjects showed clear evidence of being sensitive to the different target times at the different feeding stations, and so we proceeded to Phase 6, in which the trial-starting poke in the control station illuminated all three feeding stations. The station that would deliver food was randomly determined with equal probability. Whichever it was, it delivered a pellet only for a poke in progress at, or coming after, the target time associated with that station throughout the experiment. When the mouse failed to poke in the feeding station for a given trial at or after the target time for that station, the trial timed out at $3T^*$ plus ET^* , where T^* was equal to the longest of the three possible target latencies (45 s).

Table 1. Target Times (T) for the Two Groups of Subjects (A & B) in Successive Phases

<u>T</u>	<u>Phase 1</u>	<u>Phase 2</u>	<u>Phase 3</u>	<u>Phase 4</u>	<u>Phase 5</u> (1 F/trial)	<u>Phase 6</u> (3 F/trial)
5 s/F3			A & B		A & B	A & B
15 s/F1	A	B			A & B	A & B
45 s/F2	B	A		A & B	A & B	A & B

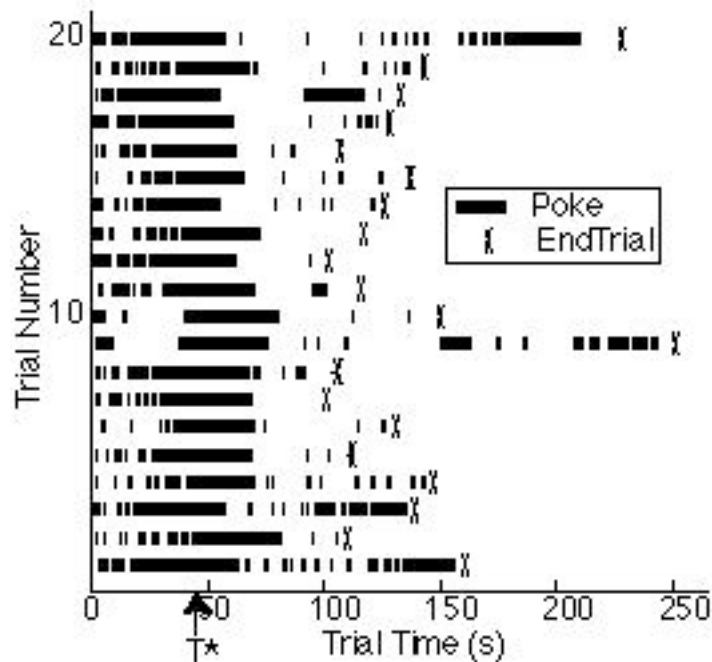
Note. In Phases 1 – 4, the same single station was illuminated on every trial. In Phase 5 (1 F/trial), a randomly chosen one station of the three stations was illuminated on any one trial. In Phase 6 (3 F/trial), all stations were illuminated on every trial, but only a randomly chosen one paid off. The target time at which a given illuminated station paid off was same throughout the experiment. F1, F2 and F3 specify the feedings stations associated with each target time.

Although we recorded the onsets and offsets of beam interruptions at all stations regardless of which station was the active station, the focus of our analysis of the Phase 6 data was on the trials on which the station with the 45 s target time was the active station. We expected that on those trials the mouse would first make a run at F3, the station with the 5 s target time, then at F1, the station with the 15 s target time, and finally at F2, the active station. More often than not, this is what happened. The many trials on which there was a clear succession of runs, first at the 5 s station, then at the 15 s station and finally at the 45 s station enabled us to look at the variance and covariance within and between runs at different temporal targets on the same trial.

Statistical Analysis

Single Peak Data. For each subject, at each of three intervals, the last 250 probe trials were selected for further analysis². The majority of these trials contained more than one poke at the target feeding station, but, on the great majority of trials, these multiple pokes presented a visually obvious "run" of more or less sustained poking around the anticipated feeding latency (see Figure 2).

Figure 2. *Sample performance on single peak probe trials in a session with a 45 s feeding latency. Pokes (intervals when the head interrupted the IR beam) are plotted against trial time.*



We developed a simple algorithm to determine the starts and stops of these runs. It divided each trial into three mutually exclusive and exhaustive intervals, an initial interval during which the head was out of the station more often than in it, a run interval during which it was in more than out, and a final interval during which it was again out more than in. These intervals were demarcated by reference to the running cumulative difference between the head-in and head-out times (Figure 3), that is, the cumulative time in the feeding station minus cumulative time out of the feeding station³. The cumulative difference function always began by going negative, because the head was always in the control station, hence out of the feeding station, when a trial started. At the start of the run, the function reversed its slope and trended upward, because, by definition the run is the period when the head is in more than out. At the end of the run, it resumed its downward trend, because, by definition, the run is over when the head is out more than it is in.

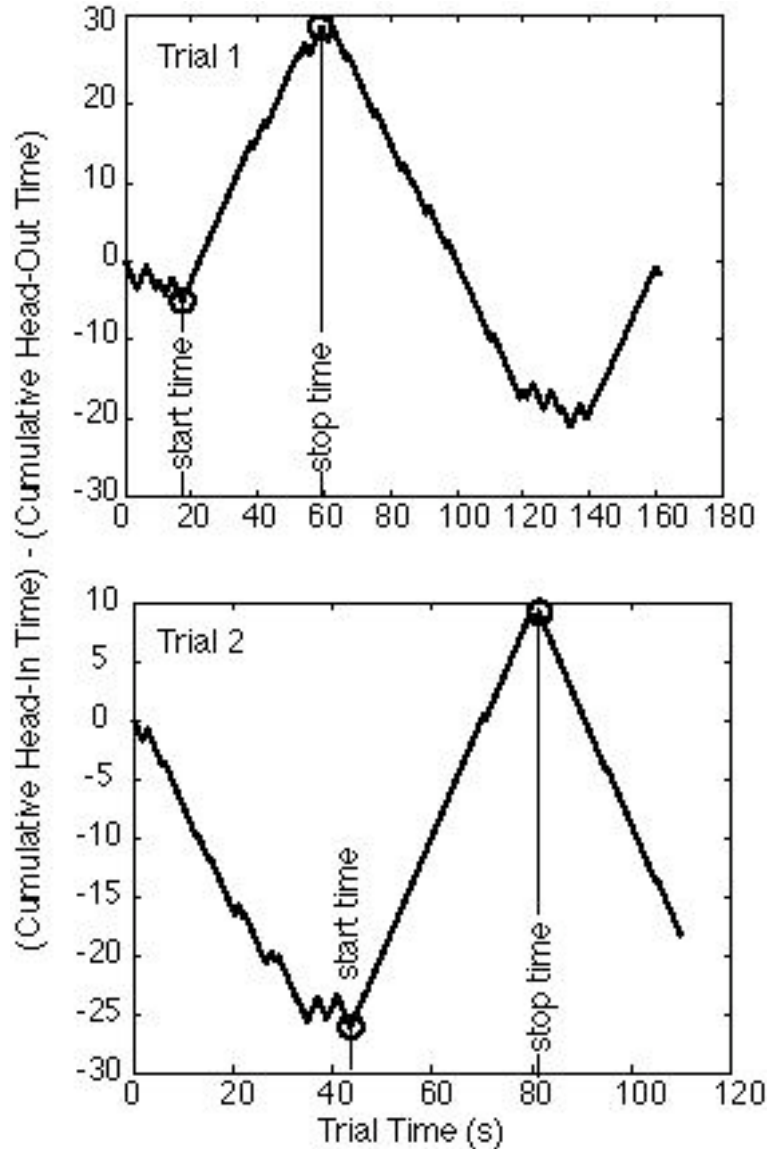
² Here we report only the analysis of the data from the second 45 s single peak sessions (Phase 4 data), but the analysis of the data at this latency in an earlier phase (1 or 2) gave very similar results.

³ Imagine two stopwatches that only run during a trial. One runs only when the head is in the station; the other only when it is out. The running cumulative difference at any point in the trial is the reading on the head-in stopwatch minus the reading on the head-out stopwatch. Both timers are zeroed at the start of a trial.

Figure 3. *Sample of running cumulative difference functions. These are for the first two trials in Figure 2. The starts and stops of the runs are circled. The start is the minimum prior to the longest uninterrupted poke. The stop is the maximum after the longest uninterrupted poke*

The algorithm for finding starts and stops first found the longest single poke, on the assumption that it would occur during the run. It looked backward from the middle of that poke to find the time at which the running cumulative difference reached its minimum, and took that to be the start time. It looked forward to find the time at which the running cumulative difference attained its maximum, and took that to be the stop time.

This algorithm differs in one potentially important respect from one we used for this purpose in a previous study (King, et al., 2001), where we followed the lead of Church, et al. (1994) in analyzing only runs that straddled the target time. Requiring the runs used for the statistical analysis to straddle the target time did not appear justified when we examined the tripeak data, because, the starts on the 15s feeding station moved much closer to the target time. As a result a non-trivial fraction of the runs did not straddle the target. Because many of the runs with late starts appeared to belong to the distribution of target-oriented runs, we did not want to exclude them from the analysis. However, we also wanted to compare the single-peak and tripeak results. This meant that we had to remove the target-straddling constraint from the single-peak data as well. To determine whether the removal of this constraint had a material effect on results of the single-peak analyses, we analyzed those data both with and without it. The straddling constraint is implemented by having the algorithm find the minimum of the running



difference function in the interval preceding the target time and maximum in the interval following the target time.

From the starts and stop times, we calculated the durations (stop time minus start time) and centers $[(\text{start time plus stop time})/2]$ of the runs. The removal of the target-straddling constraint brought into the data to be analyzed a varying number of clearly aberrant runs. These "bad trials" were excluded from previous analyses by the straddling constraint. To filter out these outliers, we eliminated runs on which the start occurred after the median of the distribution of stop times or the stop came before the median of the distribution of start times. We also removed runs whose duration was less than 1/3 the target time or more than 3 standard deviations longer than the mean duration.

Editing the data before computing descriptive statistics raises concerns about the effect of the editing on the results. Motivated by this concern, we did the single-peak analyses both with the data edited this way and with the data edited (indirectly) by imposing the independently motivated constraint that runs straddle the target time (as in previously published work of this kind)⁴. We show below that these two approaches yield very similar results. Because the imposition of our editing criteria on the single-peak data leaves trials whose summary statistics are very similar to the summary statistics from earlier work, we conclude that our criteria do not distort the resulting summary statistics. Also, we have made the raw data and the data on pokes available to interested readers by placing them on the APA website. Concerned readers may make their own determinations of the effects of this and other approaches to removing clearly aberrant runs prior to computing summary statistics. (For an indication of what is meant by "clearly aberrant see indicated trials in upper panel of Figure 4.)

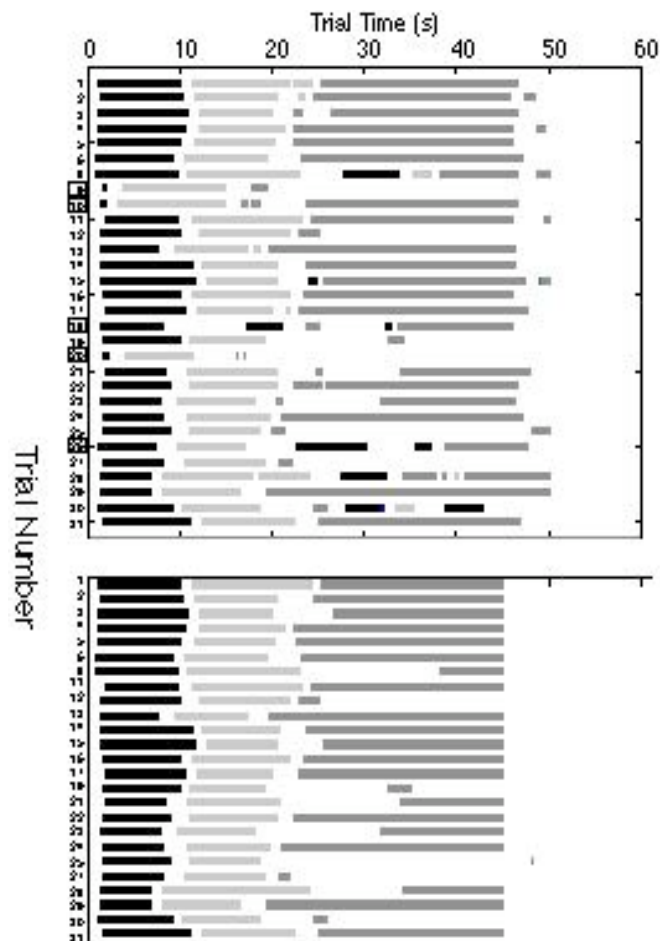
Tripeak Data. Using the data from the last eight days of Phase 6, we selected for analysis only the trials on which the station with the longest target time (45s) was the feeding station. The algorithm described above for finding runs was applied to these data to yield start and stop times for the runs at the 5 s and 15 s feeding stations and a start time for the run at the 45 s feeding station. We did not include the stop time for the 45 s target in our analyses, because most of these runs stopped at about 45 s when the feeder was activated. The duration and center of each 5 and 15 s run was then calculated, and the same filtering applied to the single peak data was applied to these data to eliminate outliers.

Out of concern for the possible impact of the editing on the results, we examined the effects of two different trial selection procedures. The first one took into account the starts of the runs at the 45 s feeding station, as well as the runs at the 5 and 15 s feeding stations. In this procedure, we eliminated trials on which there was no start at the 45s feeding station or trials in which the putative start on the 45s feeding station occurred less than 15 s into the trial. For the runs at the 5 and 15 s feeding stations, we applied the same filter that we applied to the single peak data. Only trials on which the runs at both feeding stations passed the filter were considered. The second selection procedure

⁴ The motivation for the target-straddling constraint (other than its power to purge the data of bad runs) is that it removes an analytic source of covariance between starts and stops. It does so by eliminating the overlap between start and stop distributions. When start and stop distributions overlap, there must be some positive covariance between starts and stops, because a stop cannot occur before a start. Thus, a start from the late tail of the start distribution cannot occur together with a stop from the early tail of the stop distribution. This analytic source of positive covariance is eliminated by the straddling constraint.

considered only these latter runs (at the 5 and 15 s feeding stations), ignoring the starts at the 45 s feeding station. The first procedure eliminated between 21 and 78% of the trials, depending on the subject, while the second procedure eliminated between 15 and 28% of the trials. Figure 4 gives a sample of tripeak behavior before and after filtering, marking the trials removed by the filtering. In studying it, one can see both that data editing of some kind is needed, and that (as we believe) our trial-rejection filter does not distort the big picture. The reader who wants to examine the effect of various editing strategies and see a larger sample of the kinds of trials eliminated by them is invited to pursue the question further using the data posted on the web.

Figure 4. *Sample of raw tripeak performance (top panel). Each numbered line is a trial. The intervals (pokes) during which the head interrupted the beam in each of the three feeding stations are plotted against trial time in different shades of gray. The trials whose numbers are boxed are "bad trials" to be eliminated by the filter. The bottom panel plots the good runs. Note that the filtered trials have been eliminated. There are no gaps in runs, because a run is by definition the continuous interval during which the head was in the station more than out of it.*



Results

Single Peak Sessions

All subjects showed peak behavior at all three target times (see Figure 5). On most trials, there was a well-defined run. The normalized peak curves computed from these runs did not, however, superpose (Figure 6). The peak curves plots the probability that the mouse was in a run, as a function of normalized trial time, that is, trial time divided by target time. If all of the factors determining whether the mouse has its head in the feeding station at a given point in a trial scaled with the target time, then these plots would superpose. If they did superpose, then the plots of normalized start and stop times in Figure 7 would be flat, but they are not. The departures from flatness—from constant proportionality—are significant in each subject (p generally $< .001$). Thus, the locations of

the start and stop times relative to the target time do not scale completely with target time.

When the target time is increased from 5 s to 45 s, the average start proportions and the average stop proportions both move toward 1, that is, starts and stops get proportionately closer to the target time. This makes the normalized peak curves narrower (Figure 6 = peak plots & Figure 7A = plot of start and stop proportions). Thus, at the longest target times, the duration of a run is proportionately shorter than at the shortest target times.

Figure 5 (next page). *Runs on last 25 good trials from Mouse 1 under the three different single-peak conditions (top three panels) and the tripeak condition. Vertical lines indicate the target times.*

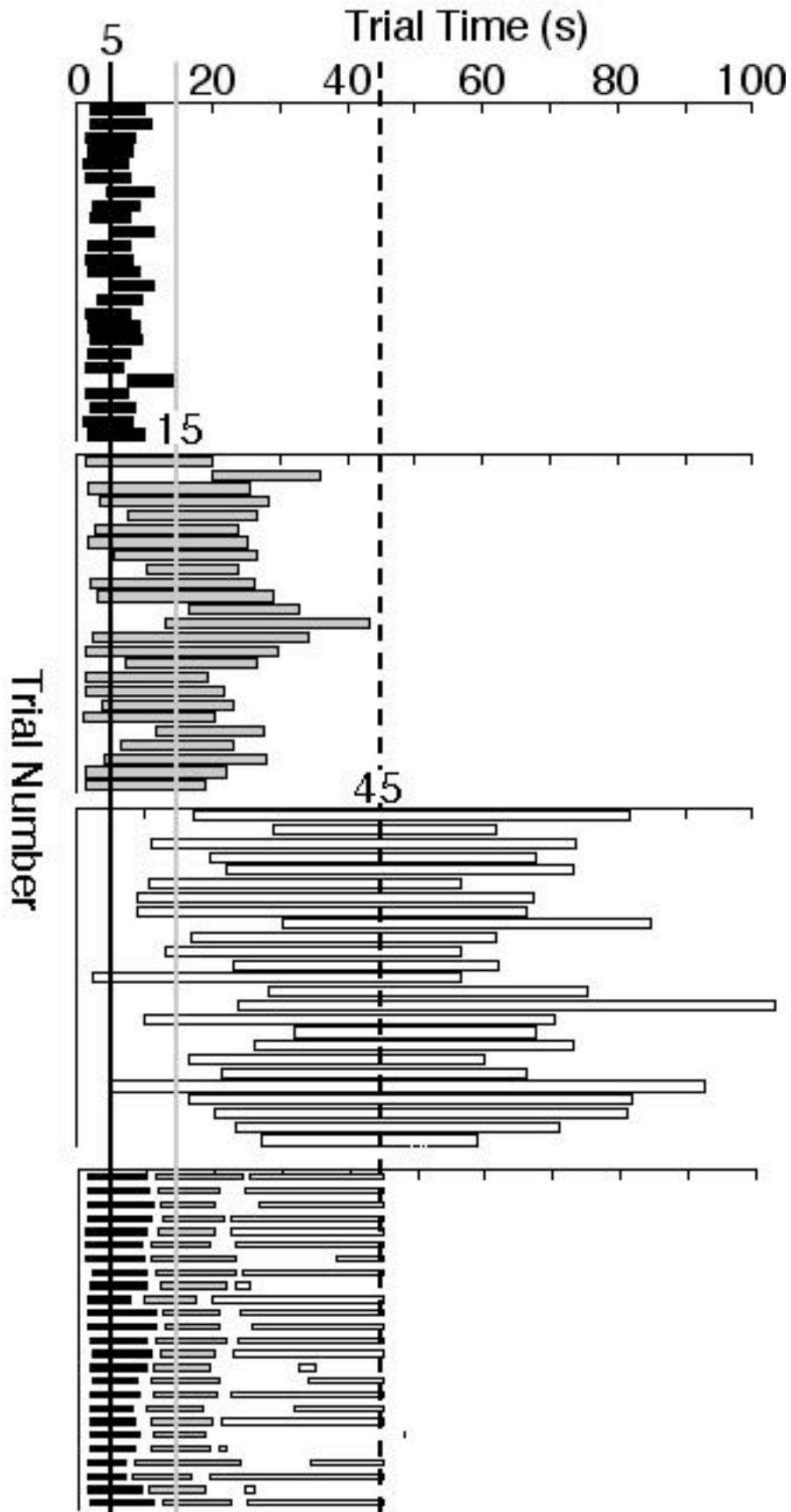


Figure 6. *Representative peak curve, which is the cumulative distribution of starts minus the cumulative distribution of stops. In this and all other subjects, the peak curve at 45 s was substantially narrower than the curves at the two shorter durations, which means that both starts and stops occurred proportionately closer to the target time.*

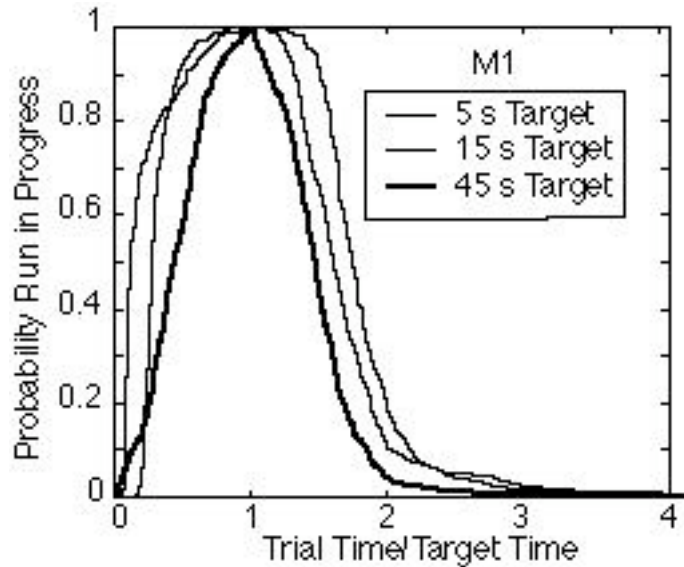
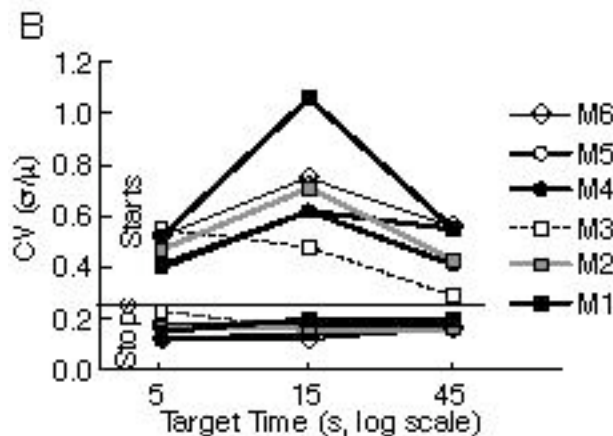
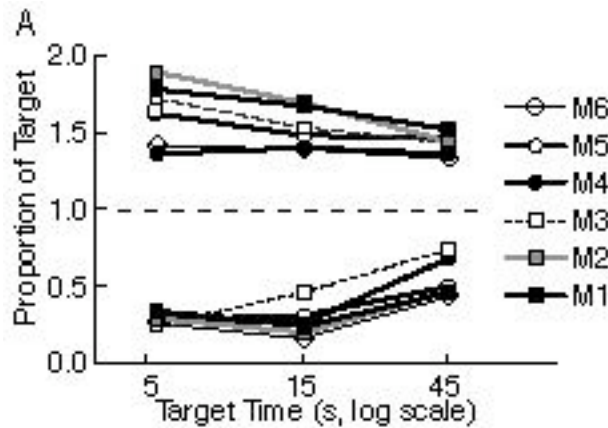


Figure 7[PD2]. A. *Mean starts and stops as a proportion of target time at three different target times. B. Coefficients of variation at three different target times. All start CVs lie above 0.25 (thin horizontal line) and all stop CVs lie below this value.*



The trend toward proportionately closer stops with longer target times is also seen at the intermediate target time; the stops were closer to 1 than at the shortest target time in 5 of the 6 subjects. However, in 5 of 6 subjects, the opposite was true for the proportion between start time and target time. The starts at 15 s occurred relatively farther from the target time (relatively sooner in the trial) than they did at both the shortest target time and the longest target time (Figure 7A). The difference

in the effect of the target time on the average start and stop proportions was marginally significant ($p = .06$).

The coefficients of variation for the starts were also not independent of target time; in 5 of 6 mice, they were greater at the intermediate target time than at either extreme (Figure 7B). Start CVs were invariably greater than stop CVs (Figure 7B). All of the latter fell below 0.25, usually close to their mean value of 0.16 (averaging across both subjects and target times). The start CVs were all greater than 0.3. Their mean value was 0.55, more than three times the mean stop CV. The values for start CVs were also notably more variable than for stop CVs, both within subjects (across target times) and between subjects. In short, under single-peak conditions, start CVs were much greater than stop CVs and much more labile.

By contrast, the CVs for stops did not vary with target time (Figure 7B). Evidently, stop proportion is the more tightly controlled parameter of the mouse's timing behavior, and its variability does not depend on target time. The standard deviation of the stop distribution remains at about 1/6 th ($0.16 \pm .03$) of the mean stop time across a 9-fold change in the target time.

The average CV for the centers of the runs was $0.20 \pm .03$. In all but one of the 18 cases (6 mice \times 3 target times), it was more than the coefficient of variation for the stops. This comparison is important in deciding between different models of the decision processes mediating the starting and stopping of runs. A single-hedge model predicts that the center CV should be less than both the start and the stop CVs (Church, et al., 1994; King, et al., 2001), which it consistently is not.

The correlation between run centers and run durations was variable (mean = 0.29, $\square = .29$). In all 18 cases, however, it was less than the ratio of the center and duration CVs, often much less. Indeed, in 3 cases it was negative. Again, this comparison is important in deciding between different models. In the single hedge model, this correlation should be approximately equal to the ratio of the CVs, whereas in the double-hedge model, it should be less (Church, et al., 1994; King et al., 2001).^[PD3]

The start-stop correlations were positive in all 18 cases and did not differ as a function of target time (mean across subjects and targets = .45). The start-duration correlations were negative in 17 out of 18 cases. The average strength of this negative correlation increased significantly with increasing target time (from -.14 at 5 s to -.49 at 45 s). This reliably negative correlation between starts and durations is not consistent with timing models in which the start and the duration of the run are independently controlled (Gibbon & Church, 1992; Killeen & Fetterman, 1988).

No aspect of these results depended on the algorithm used to determine the starts and stops of runs. The results just reported were for the method in which we did not constrain the start and stop of a run to straddle the target time, but we did filter the runs to eliminate outliers. A very nearly identical pattern of results was obtained when we did constrain the starts and stops to straddle the target time, as has been the past practice (see Table 2, Constr and Unconstr rows).

Table 2. *Comparison of the Single Peak Results based on Constrained versus Unconstrained Runs and of Tripeak Data with and Without Consideration of the 45 s Starts*

		<u>Target Times (s)</u>		
		<u>5</u>	<u>15</u>	<u>45</u>
Normalized Starts	Constr	0.28	0.26	0.50
	Unconstr	0.29	0.27	0.55
	TrPk1	0.38	0.69	0.66
	TrPk2	0.39	0.69	
Normalized Stops	Constr	1.67	1.55	1.43
	Unconstr	1.63	1.53	1.42
	TrPk1	1.81	1.43	
	TrPk2	1.81	1.42	
StrtCVs	Constr	0.43	0.65	0.43
	Unconstr	0.48	0.71	0.46
	TrPk1	0.52	0.18	0.24
	TrPk2	0.53	0.18	
StpCVs	Constr	0.21	0.21	0.20
	Unconstr	0.16	0.16	0.17
	TrPk1	0.18	0.16	
	TrPk2	0.18	0.16	
CenterCV	Constr	0.21	0.22	0.20
	Unconstr	0.18	0.20	0.21
	TrPk1	0.21	0.15	
	TrPk2	0.21	0.15	
Start-Stop Correlation	Constr	0.29	0.26	0.26
	Unconstr	0.45	0.40	0.49
	TrPk1	0.42	0.61	
	TrPk2	0.43	0.60	
Center-Duration Correlation	Constr	0.71	0.39	0.04
	Unconstr	0.55	0.43	0.01
	TrPk1	0.43	0.64	
	TrPk2	0.38	0.64	
CtrCV/DrCV	Constr	0.91	0.91	0.63
	Unconstr	1.10	1.06	0.76
	TrPk1	0.97	0.62	
	TrPk2	0.98	0.61	

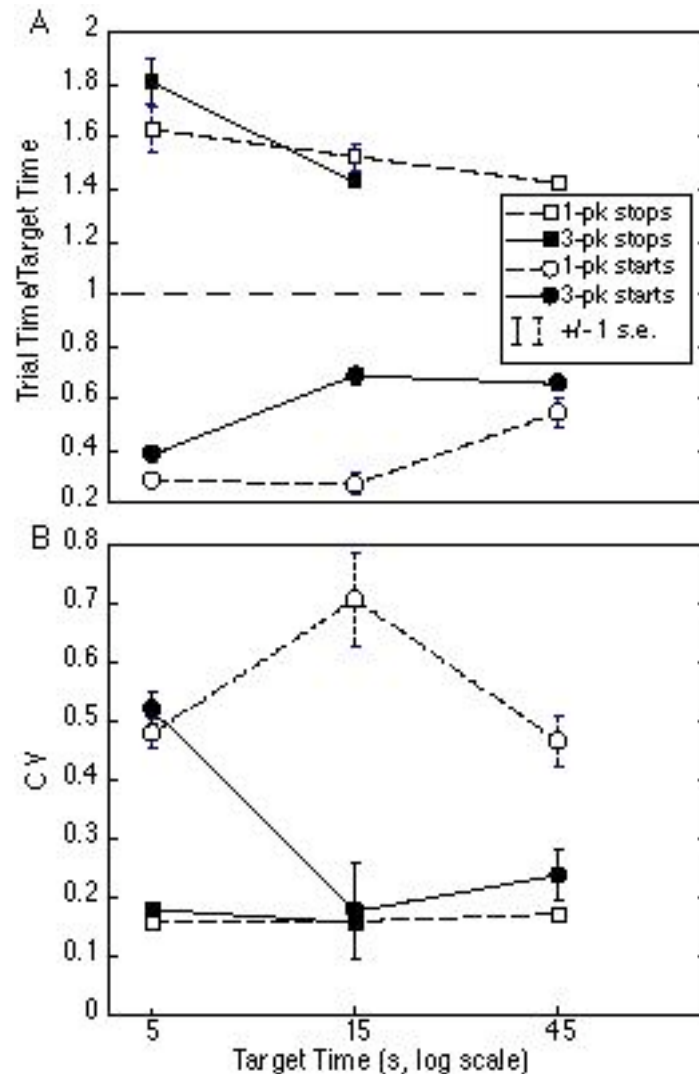
Note: "Constr" are single-peak statistics based on runs that were constrained to straddle the target time. "Unconstr" are single-peak statistics based on runs that were not so constrained but were filtered for outliers. "Tripeak 1" statistics are based on trials where the starts of 45 s runs were considered, in addition to filtered runs at the two shorter durations. Only trials on which the start at the 45 s feeding station was greater than 15 s and less than 45 s were considered. Tripeak 2 statistics are based on trials where only the (unconstrained but filtered) 5 s and 15 s runs were considered.

Tripeak Sessions

On trials when the long-latency feeding station paid off (tripeak trials), subjects behavior toward each feeding station was usually governed by the expected feeding time in that feeding station: they kept their head in the short latency feeding station for a run that generally straddled the expected feeding time for that feeding station, then shifted to the middle-latency feeding station, where they kept their head in the feeding station for a run that generally straddled the expected feeding time for that feeding station, and, finally, they shifted to the long-latency feeding station (Figure 4).

Some of the run statistics were, however, strikingly different. The tripeak procedure moved the starts closer to the target time, most strikingly for the 15 s target (Figure 8A). For the 15 s and 45 s starts, it greatly reduced the variability of starts, reducing start CVs to about the values seen with stops (Figure 8B). By contrast, the effect of the tripeak procedure on stops was minimal. It had no effect on the CVs and small and inconsistent effects on the locations of the stops at the 5 and 15 s feeding stations

Figure 8. *Comparison of single peak and tripeak start-stop statistics at different target times. A. Normalized starts and stops. B. Start and stop CVs.*



Of theoretical significance is the fact that the tripeak procedure had opposing effects on starts and stops at the 5 s feeding station. It moved the stops at this feeding station farther from the target, while moving the starts closer to it (interaction $p < .005$). For the 15 s feeding station, the tripeak procedure moved the starts much closer to the target value, while having little or no effect on the location of the stops (interaction $p < .001$). These differing effects on start and stop locations are direct evidence against the single-hedge assumption, which requires that any movement toward the target by the start be accompanied by an equal move toward the target by the stop.

In the tripeak procedure, the correlations between start times and stop times at a feeding station were more variable between subjects but broadly similar to those obtained in the single peak procedures (.43 at the 5 s feeding station and .60 at the 15 s feeding station). The start-duration correlation was consistently negative at the 5 s feeding station (average $-.29$), but variable, with an average not significantly different from zero, at the 15 s feeding station.

The average correlation between the stop time at the 5 s feeding station and the stop time at the 15 s feeding station was $0.55 (\pm .15)$. This was nearly as high as the average start-stop correlation at the 15 s feeding station ($.60 \pm .07$) and higher than the start-stop correlation at the 5 s feeding station ($.43 \pm .16$). Whatever it is that causes the start and stop times for a common remembered feeding latency to covary from trial to trial, it causes a similar degree of trial-to-trial covariation in stop times for different remembered temporal targets. Possible sources for this cross-target covariation are considered in the Discussion.

Discussion

The Sources of Systematic Error

Our results are not consistent with the single-hedge model, the model in which the start proportion and stop proportion deviate from 1 by the same value (the common hedge factor). Both the single-peak and the tripeak results give strong grounds for rejecting this model. The simplest, most direct evidence against it is our finding that the start and stop proportions are affected differently by different target durations in single peak sessions and by the tripeak procedure. In Figure 7A, one sees that the 15 s start proportion in 5 of 6 mice is farther from 1 than is the 5 s start proportion, while the stop proportion at 15 s is closer to 1 than at 5 s. This cannot happen in the single-hedge model. If one proportion gets closer to 1, so must the other, because their distance from the remembered target time is determined by a common hedge factor.

In the same vein, the tripeak procedure caused the start proportions at the 5 s feeding station to move closer to 1, while the stop proportions at this feeding station moved further from 1 (Figure 8A). It also caused the starts at the 15 s feeding station to move much closer to 1 (relative to their location in the single-peak procedure), while having no significant effect on the location of the stops.

Less simple but equally important is the failure of the correlation structure to conform to the predictions of the single-hedge model, as has been clearly recognized in previous work on the statistics of single trials (Brunner, et al., 1997; Church, et al., 1994). On this model, start and stop times should be negatively correlated, because a small hedge factor leads to a late start and an early stop. But starts and stops are invariably positively correlated, in both the single-peak and the tri-peak procedure; late starts tend to

be paired with late stops. Even when single peak runs were constrained to straddle the target time, eliminating the analytic source of covariance in starts and stops, the positive correlations ranged from .26 to .29 across target latencies.

Our results are also inconsistent with a model, such as the Killeen and Fetterman (1988) model in which the observed behavior is the result of a start decision and an (independent) duration decision. In such a model, the starts and durations of runs should be uncorrelated. Or, they could be positively correlated, if trial-to-trial variation in timer speed causes starts and durations to covary. But they should not be negatively correlated. In fact, however, durations are almost always negatively correlated with starts, as would be the case if duration were a secondary consequence of (partially) independent start and stop decisions. If the stops are completely uncorrelated with the starts, then the later a run starts, the shorter, on average, will be its duration.

Our results imply a model in which start and stop decisions are partially independent. Once it is recognized that start and stop proportions are at least partially independently determined, then it is clear that modest systematic error in the location of the peak curve is likely to be the result of asymmetric start and stop decision criteria. The start criterion on a given trial is the proportion of the remembered target time that the estimate of elapsed time on the current trial must exceed in order for the subject to start its run. Similarly, the stop criterion is the proportion by which the estimated elapsed interval must exceed the remembered target time in order for the subject to stop its run. If, on average, the subject puts its start criterion farther from 1 than its stop criterion, then the location of its peak curve will be centered left of 1. If the asymmetry goes the other way, it will be centered right of 1. Thus, memory miscalibration is unlikely to be the source of the modest systematic errors seen in normal subjects.

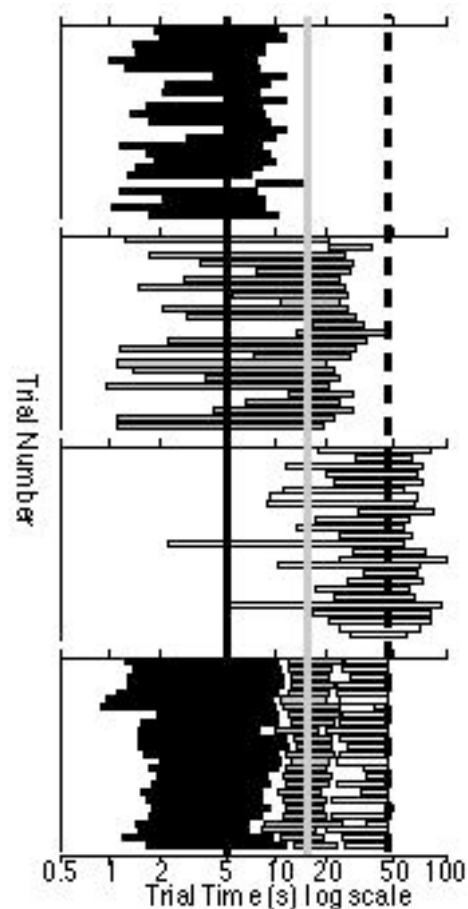
This conclusion is consistent with the clear lability in the start and stop proportions seen in our results. One consequence of this lability is that, although there are systematic errors in the locations of the peak curves, they are not scalar errors. Thus, for example, the fact that in the single-peak data, the start proportions lie farther from 1 when the target is 15 s than they do when it is 5 s, while the opposite is true for the stop proportions, means that in all five of those mice the systematic error is proportionately greater with a 15 s target than with a 5 s target (see, for example, Figure 6). When we use the tripeak procedure, the systematic error for a 15 s target reverses. In the tripeak procedure, the average center of a run at the 15 s feeding station, expressed as a proportion of the target, was 1.06 (s.e. = .02). In the single-peak procedure, it was .90 (s.e. = .03).

A miscalibrated memory mechanism, which delivers back values that deviate systematically from the values originally put in (Gibbon, et al., 1984), may be a factor in producing systematic error, but the effects of systematic memory error cannot be distinguished from the effects of asymmetrically placed decision criteria when the peak curve straddles the target value. If, however, one found a subject whose peak curve lay almost entirely to one side of the target time, then a gross systematic error of that kind could not be explained by asymmetric decision criteria. Presumably, a hedge factor cannot be less than zero. When a start hedge or a stop hedge factor is zero, the median of that distribution will be at the remembered target time. If the median of either distribution were to fall on the wrong side of the target time, then systematic misremembering of the target time (memory miscalibration) would be the more plausible explanation.

The Sources of Variability

Starts, stops and peaks. It is common in the analysis of peak data to treat the ratio of the width of the peak curve to its center as a measure of the variability in performance (e.g., Rakitin, et al., 1998). In the light of our results, this is not a valid measure of variability, because the width of the peak curve depends not on variability per se, but rather on how far apart the subject places its start and stop decision criteria. The more widely separated these criteria, the wider the resulting peak function. The peak curve for the 15 s feeding station is dramatically narrower in the tripeak condition than in the single peak condition (compare the second and fourth panels from the top in Figure 9), but that is not primarily because the behavior was less variable in the latter condition. It is primarily because the subjects' start criteria were much closer to the target value when the tripeak procedure was used.

Figure 9. *A representative sample of runs in the three single-peak conditions and the tripeak condition, plotted against the logarithm of trial time, in order to make equal changes in proportion into equal intervals on the plot. Note also the dramatic narrowing of the run time on the 15 s feeding station. (Compare second panel from top to bottom panel.)*



One component of the subjects' behavior was much less variable in the tripeak condition, namely, the start proportions. The variability in the other component, the stop proportions, was exactly the same in the single peak and tripeak conditions (see Figure 8B, 15 s target time). The rapprochement of the start proportion to the target and the decrease in its variability went hand in hand. And, the stop proportions, which are generally closer to the target than the start proportions, are also less variable. It seems likely that, as a rule, the smaller the variability in the proportions, the closer to the target they will be placed. This correlation is, however, by no means rigid. The start and stop proportions in the single peak condition are both considerably closer to the target at the longest target value than at the shortest (Figure 7A) but their variability is about the same at these two extremes (Figure 7B). Thus, in measuring variability in peak performance, it is essential to focus on its independent components, the start and stop distributions, rather than on their combined effect, which is the peak curve.

When we focus on the start and stop distributions rather than on the peak curves, we see striking time-scale invariance in one component, namely, the variability of the

stop distribution. The standard deviation of the stop distributions is close to $1/6^{\text{th}}$ of the mean stop time in all three single-peak conditions, and in the tripeak condition (Figure 8B).

Equally as striking is the lack of time-scale invariance in the other component. The character of the variability in the single-peak start distributions and the striking change brought about by the tripeak procedure are best appreciated by plotting runs against the logarithm of trial time (Figure 9), because equal intervals on such a plot correspond to equal proportions. If variability scales from one target value to another, then the raggedness of the starts and stops in these plots will be the same at different target durations.

The extreme raggedness of the starts in the single-peak conditions, relative to the stops, and relative to what is seen at the 15 s feeding station in the tripeak condition, suggests to us that the start distribution is a mixture distribution. It appears that on some trials, the subject does time its start relative to the remembered target, whereas on others, it starts, so to speak, as soon as it arrives at the feeding station. (Recall that the subject initiates a trial by poking into the control station on the other side of the test box.) These starts we might call "untimed" starts. On this hypothesis, the observed start distribution is a mixture of timed and untimed starts. As the target time gets longer, the proportion of untimed starts in this mixture gets smaller. When, in the tripeak procedure, the subject is given something else important to do prior to starting at the 15 s feeding station, untimed starts disappear altogether. When they do, starts are no more variable than stops. We conclude that there is an irreducible variability in the factors that govern timing behavior, that this variability is proportional to remembered duration (that is, it is scalar), and that it is measured by the variability in stop behavior in the peak procedure. Stops in the peak procedure are precisely timed, whereas starts are not.

Correlations. A positive correlation between starts and stops is reliably observed when single trials of peak data are analyzed for their correlational structure (Cheng & Westwood, 1993; Gibbon & Church, 1992; Church, et al., 1994; Rakitin, et al., 1998). This is clearly seen in our data as well. This could be explained by trial-to-trial variation in the remembered feeding latency. On trials where the sample from memory was longer than usual, both the start and the stop would tend to be late. Alternatively, it could be explained by trial-to-trial variability in clock speed. On trials where the clock ran slow, both the start and the stop would tend to be late. A major motivation for the use of the tripeak procedure was that it would distinguish between these different sources.

The first explanation predicts no cross-target correlation between unconstrained components of run behavior in the tripeak task, whereas the second predicts that these components should be as strongly correlated as the single-station starts and stops. Stops on the 5 s feeding station and starts on the 15 s feeding station are likely to be correlated for purely analytic reasons, the subject cannot start on the latter feeding station until it has stopped on the former. However, after the filtering out of bad trials, the distributions of stops on the 5 s feeding station does not overlap the distribution of stops on the 15 s feeding station, so there is no analytic reason for the stop times on the two feeding stations to be correlated. They are, however, strongly correlated ($\rho = .55$, $\sigma = .15$), about as strongly correlated as the start and stop times at the 15 s feeding station ($\rho = .60$, $\sigma = .07$). Therefore, whatever the source of the start-stop covariation, it is common to runs controlled by two different remembered temporal targets. Clock variance, that is, trial-to-

trial variability in timer speed, is the most obvious such source. Even if the clock rate varies during the course of the trial, when the clock is slow to approach and pass the first remembered target, then it will on average be later in approaching and passing subsequent targets as well. (Consideration of the bus analogy used in the introduction may help to see why this is so.)

The cross-target start-start and start-stop correlations also imply a source of variability that is common to runs regulated by different samples from memory. However, these correlations are not as suitable for estimating the importance of this common factor, given the likelihood that the start distribution at the 5 s feeding station is a mixture of timed and untimed starts.

Our results with mice are very similar to results obtained with pigeons, rats, and humans (Cheng & Westwood, 1993; Gibbon & Church, 1992; Church, et al., 1994; Rakitin, et al., 1998) in all essential respects, although this previous work has not led to the conclusions we now draw from these results. In non-human subjects, start CVs are always considerably greater than stop CVs, suggesting that the start distributions are usually mixture distributions. Even in humans, peak curves often fail to superpose in ways similar to those seen in our data (Rakitin, et al., 1998). The correlational structure we observe in mice is also seen in the other three species of subjects. That correlational structure is strong evidence against the single-hedge model (Brunner, et al., 1997; Church, et al., 1994; King, et al., 2001). On the other hand, our tripeak correlational structure, which implicates clock variability as the source of the covariation that has heretofore been ascribed to memory variability (Church, et al., 1994), is new.

We conclude, first, that the most likely source of modest systematic error in peak curves is the asymmetric placement of the start and stop decision criteria relative to the remembered feeding latency. In other words, the error arises from performance factors (decision criteria), not memory. We conclude, second, that trial-to-trial variability in clock speed explains most of what has heretofore been ascribed to memory variability. In other words, the variability arises from an aspect of the learning process (interval timing) rather than from memory. Finally, we note that an attraction of this kind of quantitative content-oriented approach to the study of memory is that it permits the functional localization of the sources of error and variability. In assessing the behavioral effects of genetic manipulations, establishing the functional locus of a genetic effect is as important as establishing the chromosomal location of the mutant gene.

References

- Bevins, R. A., McPhee, J. E., Hauhut, A. S., and Ayres, J. J. B. (1997). Converging evidence for one-trial context fear conditioning with an immediate shock: Importance of shock potency. Journal of Experimental Psychology: Animal Behavior Processes 23, 312-324.
- Bolles, R. C., and Moot, S. A. (1973). The rat's anticipation of two meals a day. Journal of Comparative and Physiological Psychology 83, 510-514.
- Brunner, D., Fairhurst, S., Stolovitzky, G., & Gibbon, J. (1997). Mnemonics for variability: Remembering food delay. Journal of Experimental Psychology: Animal Behavior Processes, 23(1), 68-83.

- Catania, A. C. (1970). Reinforcement schedules and psychophysical judgments. In Theory of reinforcement schedules, W. N. Schoenfeld, ed. (New York, Appleton-Century-Croft), pp. 1-42.
- Cheng, K. (1992). The form of timing distributions in pigeons under penalties for responding early. Animal Learning and Behavior, 20(2), 112-120.
- Cheng, K., & Westwood, R. (1993). Analysis of single trials in pigeon's timing performance. Journal of Experimental Psychology: Animal Behavior Processes, 19, 56-67.
- Cheng, K., Westwood, R., & Crystal, J. D. (1993). Memory variance in the peak procedure of timing in pigeons. Journal of Experimental Psychology: Animal Behavior Processes, 19, 68-76.
- Church, R. M., & Meck, W. H. (1988). Biological basis of the remembered time of reinforcement. In M. L. Commons & R. M. Church & J. R. Stellar & A. R. Wagner (Eds.), Quantitative analyses of behavior: Biological determinants of reinforcement (Vol. 7, pp. 103-119). Hillsdale, NJ: Erlbaum.
- Church, R. M., Meck, W. H., & Gibbon, J. (1994). Application of scalar timing theory to individual trials. Journal of Experimental Psychology: Animal Behavior Processes, 20(2), 135-155.
- Fantino, E. (1969). Conditioned reinforcement, choice, and the psychological distance to reward. In Conditioned reinforcement, D. P. Hendry, ed. (Homewood, IL, Dorsey Press), pp. 163-191.
- Fetterman, J. G., & Killeen, P. R. (1995). Categorical scaling of time: Implications for clock-counter models. Journal of Experimental Psychology: Animal Behavior Processes, 21, 43-63.
- Gallistel, C. R. (1980). The organization of action: a new synthesis (Hilldale, NJ, Lawrence Erlbaum).
- Gallistel, C. R. (1990). The organization of learning. Cambridge, MA: Bradford Books/MIT Press.
- Gallistel, C. R. (1999). The replacement of general-purpose learning models with adaptively specialized learning modules. In The cognitive neurosciences. 2nd ed., M. S. Gazzaniga, ed. (Cambridge, MA, MIT Press), pp. 1179-1191.
- Gallistel, C. R. (2002). The principle of adaptive specialization as it applies to learning and memory. In Principles of human learning and memory, R. H. Kluwe, G. Lüer, and F. Rösler, eds. (Basel, Birkenaeuser), pp. 250-280.
- Gallistel, C. R., and Gibbon, J. (2000). Time, rate and conditioning. Psychological Review 107, 289-344.
- Gibbon, J. (1977). Scalar expectancy theory and Weber's Law in animal timing. Psychological Review, 84, 279-335.
- Gibbon, J. (1991). Origins of scalar timing. Animal Learning and Behavior, 22, 3-38.
- Gibbon, J., & Church, R. M. (1992). Comparison of variance and covariance patterns in parallel and serial theories of timing. Journal of the Experimental Analysis of Behavior, 57, 393-406.
- Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. In J. Gibbon & L. Allan (Eds.), Timing and time perception (Vol. 423, pp. 52-77). New York: New York Academy of Sciences.

- Killeen, P. (1975). On the temporal control of behavior. Psychological Review 82, 89–115.
- Killeen, P. R., & Fetterman, J. G. (1988). A behavioral theory of timing. Psychological Review, 94, 455-468.
- Killeen, P. R., & Weiss, N. A. (1987). Optimal timing and the Weber function. Psychological Review, 94, 455-468.
- King, A.P., McDonald, R.V., & Gallistel, C.R. (in press). Screening for mice that remember incorrectly. International Journal of Comparative Cognition.
- LaBarbera, J. D., and Church, R. M. (1974). Magnitude of fear as a function of the expected time to an aversive event. Animal Learning and Behavior 2, 199-202.
- Matell, M.S., King, G.R., & Meck, W.H. (submitted). Differential Modulation of Clock Speed in the Tri-Peak Procedure by the Chronic Administration of Intermittent versus Continuous Cocaine.
- Meck, W. H., & Church, R. M. (1987). Cholinergic modulation of the content of temporal memory. Behavioral Neuroscience, 101, 457-464.
- Rakitin, B. C., Gibbon, J., Penney, T. B., Malapani, C., Hinton, S. C., & Meck, W. H. (1998). Scalar expectancy theory and peak-interval timing in humans. Journal of Experimental Psychology: Animal Behavior Processes, 24(1), 15-33.
- Roberts, S. (1981). Isolation of an internal clock. Journal of Experimental Psychology: Animal Behavior Processes, 7, 242-268.
- Roberts, S., & Church, R. M. (1978). Control of an internal clock. Journal of Experimental Psychology: Animal Behavior Processes, 4, 318-337.
- Savastano, H. I., and Miller, R. R. (1998). Time as content in Pavlovian conditioning. Behavioural Processes 44, 147-162.
- White, N. E., Kehoe, E. J., Choi, J.-S., and Moore, J. W. (2000). Coefficients of variation in timing of the classically conditioned eyeblink in rabbits. Psychobiology 28, 520-524.

Acknowledgements

This work was supported by NIMH Grant R21 MH63866 to CRG