International Journal of Comparative Psychology, 2001, **14**, 232-257.

Screening for Mice that Remember Incorrectly

Adam Philip King,
Fairfield University, U.S.A.

Robert V. McDonald, and C. R. Gallistel
Rutgers University, U.S.A.

The memory mechanism carries information forward in time. Screening for genetic distortions in this information (screening for changes in the content of memory) is likely to do a better job of distinguishing between genetic effects on memory mechanisms and genetic effects on performance mechanisms than screening for changes in the strength of learned behavior. The peak procedure is the most commonly used screen for duration memory. Mice give peak data strikingly similar to data from the rat and the pigeon. The data, however, favor a model in which the decision criterion for starting the response (putting the head into the hole) and stopping the response are independently determined. For reasons we explain, this prevents the estimation of scalar memory error and memory variability from simple peak data.

Memory is a foundation of higher mental function and its destruction by degenerative diseases of the nervous system is a devastating and common health problem. Identifying the cellular and molecular mechanisms involved in memory is essential to our understanding of these diseases. Memory mutants offer hope for rapid progress in determining these mechanisms. If we had a strain of mice that bred true for a mutation in a gene that coded for an essential component of the molecular machinery of memory, we could use the increasingly powerful techniques of traditional and molecular genetics to locate and sequence the gene. It would then be possible to use molecular biological methods to find where and when that gene was expressed in the brain, and where the gene product was located within cells.

The identification and production of such memory mutants will depend fundamentally on effective behavioral screens, because memory is known only by its behavioral effects. In turn, the effectiveness of behavioral screens is determined by their diagnostic specificity. With regards to the molecular basis of memory, behavioral scientists must devise screens that distinguish between those genetic effects that bear directly on the mechanisms involved in memory and those that bear on processes that translate the remembered information into observable behavior. These latter performance factors are numerous; they include motivation, attention, and health, as well as many factors that do not fall readily under any of these headings (Rescorla, 1998; Wilkie et al., 1999). These processes affect behavioral indices of memory (i.e., the probability or vigor of a memory-based

This research was supported by NIMH Grant 1R21MH63866 to CRG. The assistance of Mario Walls and Yair Ghitza is gratefully acknowledged. Correspondence concerning this article may be addressed to Adam King, Fairfield University, North Benson Rd., Fairfield, CT 06430-5195, U.S.A. (king@cs.fairfield.edu).

response), but they are functionally independent of the memory itself. They play no role in carrying information forward in time; they affect only the expression of that information in observable behavior.

Most current behavioral screens for memory, such as the Morris water maze (e.g., Vicens et al., 1999), measure the *strength* of a memory-dependent behavior and therefore potentially confound performance and memory factors. In effect, they screen for animals that remember poorly, or, less often, for animals that remember unusually well (cf. Tang et al., 1999). Measures such as latency to reach the platform do not specifically implicate memory as the source of variation in behavior, in that there is no simple relationship between the strength of observed behavior and potential quantitative measures of the underlying engram. It is unclear what it means, physiologically speaking, to say that a mouse remembers the location of a platform more strongly under some circumstances than under others.

An alternative to screening for changes in the strength of memory-dependent behavior is to screen for distortions in the *contents* of memory, for example, "what does a subject remember a fixed interval to be" rather than "how likely is the subject to respond to a signal that predicts an event that will occur at that fixed interval" (cf. Church & Meck, 1988). In a wide variety of tasks, animals remember variables like distances, durations, and numerosity (Gallistel, 1990). This approach to studying memory has been most extensively developed in the field of timing behavior (Brunner, 1997; Fantino, 2000; Fetterman, 1995; Malapani, 1998; Rakitin, 1998; Roberts, 1998; Wilkie, 1995).

Among the several timing tasks that have been explored, the most thoroughly analyzed has been the peak procedure (Catania & Reynolds, 1968; Cheng & Westwood, 1993; Cheng et al., 1993; Church et al., 1994; Roberts & Church, 1978). In this procedure, subjects indicate with their responses (e.g., key pecks or lever presses) the target latency, at which they expect responding will have some consequence (e.g., food delivery) following a signal event (e.g., the illumination of a key or the extension of a lever). On some trials, the anticipated event fails to occur, in which case subjects cease responding when they no longer anticipate it. From these probe trials one gets data reflecting the temporal interval stored in memory. Subjects begin responding somewhat before the anticipated event and cease responding when they no longer expect it, thereby bracketing the time at which the event was expected. The advantage of using tasks like the peak procedure as a behavioral screen for changes in the quantitative properties of memory is that the temporal distribution of responses, which reflects the information about the remembered temporal interval, is demonstrably independent of the strength of the behavior (Roberts, 1981). For example, a pigeon will display the same temporal distribution of key pecks to obtain food in both food-deprived and relatively satiated conditions, although responding may be far more vigorous in the food-deprived condition. When normalized for maximum rate of pecking, these distributions will superpose.

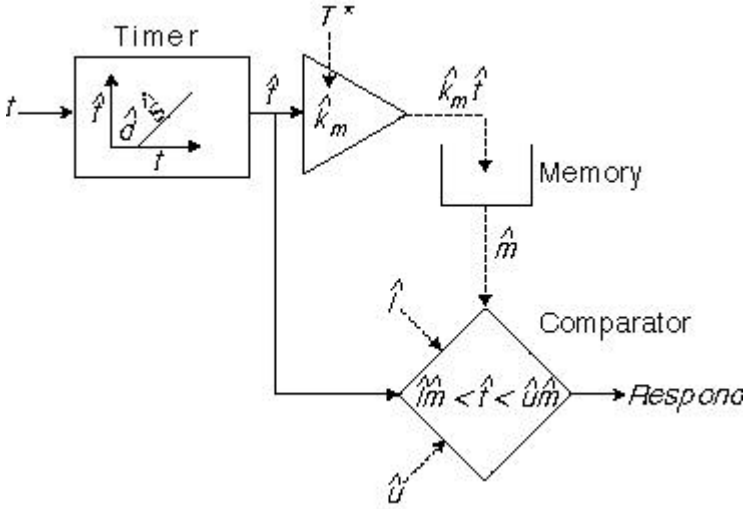

Figure 1. The SET framework. Solid arrows represent variables that vary continuously with time (e.g., subjective time); dashed arrows represent variables that take on enduring values at specific moments in time (e.g., memories for durations). Hatted variables are subjective quantities (not directly observable); variables without hats are objective quantities (directly measurable). t = objective trial time; $\hat{t}$ = subjective trial time; $\hat{d}$ = the delay before starting the timer; $\hat{s}$ = clock speed, the slope of the function relating subjective trial time to objective trial time; T^* = the target latency; $\hat{k}$ = the memory scalar; $\hat{m}$, a sample from memory used as the reference (or target) value on a probe trial; $\hat{l}$ = the proportion by which the start criterion anticipates the target value; $\hat{u}$ = the proportion by which the stop criterion exceeds the target value.

Scalar Expectancy Theory (SET; Gibbon, 1977, 1991; Gibbon et al., 1984) provides a framework that accounts for many facets of animal timing and it has been notably successful in accounting for data from the peak procedure (Cheng & Westwood, 1993; Cheng et al., 1993; Church et al., 1994; Gibbon & Church, 1992; Meck & Church, 1984; Meck & Church, 1987). SET makes explicit the basic operations presumably implemented in any system capable of solving the peak task (Figure 1). First, the system must have a timer (clock) that delivers an ongoing measure of elapsed duration. It must have a memory, a mechanism for carrying forward in time records of the elapsed intervals at which food has previously been delivered. Finally, it must have a comparator to decide whether the currently elapsed interval is close enough to the remembered interval to begin responding and whether it has grown sufficiently longer than the remembered latency to terminate responding. “Close enough” and “sufficiently longer” are specified by decision criteria.

Three quantitative features have been fairly well established when applying the SET framework to the data from the peak procedure (hereafter peak data). First, the subjective measure of trial duration is proportional to objective duration (Gallistel, 1999; Gibbon & Church, 1981). Second, the decision variable in SET is the ratio of the currently elapsing interval to the remembered interval (Aronson et al., 1993; Gibbon, 1991, 1992; Gibbon & Fairhurst, 1994). Third, the major sources of variability scale with the remembered temporal interval.

The left hand panels in Figure 2, from Gallistel and Gibbon (2000), show peak data from three different pigeons run with a signal-food interval of 30 s during one block of sessions and 45 s during another. The response measure is rate of pecking, averaged over many probe trials. The right hand panels in Figure 2 show the same data after the temporal axis has been normalized with respect to the signal-food interval and the response axis has been normalized with respect to the peak response. After normalization, the curves obtained at different feeding latencies superpose. This is a ubiquitous property of timing data, commonly referred to as time-scale invariance, or scalar property.

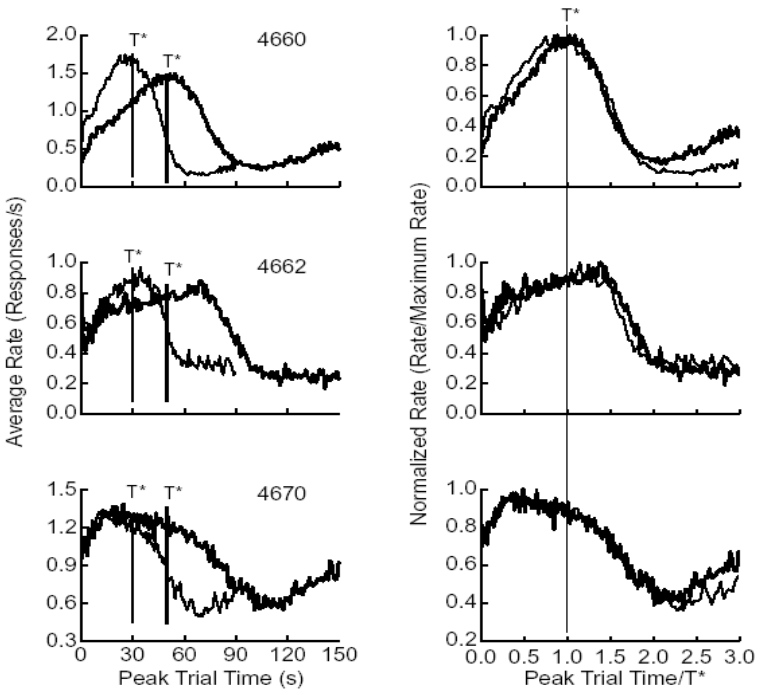

Figure 2. *Left*. Average pecking rates on probe (no-food) trials during sessions when the latency (*T**) between key illumination and food delivery was either 30 s or 45 s. Each panel gives data from a different bird (identified by the number on the panel). The feeding latencies are indicated by the vertical lines. *Right*. The same data after normalization with respect to target latency and peak response rate. (Reproduced with slight modifications from Gallistel & Gibbon, 2000).

For our purposes, the most interesting aspects of the peak data in Figure 2 are the individual differences in the location of the peak and the amount of spread about this location. The systematic errors in the location of the peak response are proportional to the pay-off latencies. Some birds are systematically late, some systematically early, and the magnitudes of these systematic errors are proportional to the target times (scalar error).

There are also systematic differences in the extent to which responding tightly or loosely brackets the target time. The curves from Bird 4660 are more sharply peaked than the curves for the other two birds. The variability from trial-

to-trial in the remembered target time would appear to be less in this bird than in the other two. These differences in trial-to-trial variability are likewise scalar; that is, the coefficient of variation (cv), which is the ratio between the standard deviation and the mean of a distribution, is constant within subjects across pay-off latencies and differs between subjects.

It is important to appreciate that between subject variations in the rates at which their timers run are unlikely to result in the between-subject differences in peak location seen in Figure 2. A subject with a slow timer would not be expected to have a peak located later than the target time, nor would a subject with a fast timer be expected to have a peak located before the target time. The location of the peak is determined by a comparison between two times, both of which presumably come from the same timer. One time is the currently elapsing interval, and one is the remembered pay-off latency. Because both times come from the same timer, between subject timer differences cancel out in determining the location of the peak relative to the target. The systematic error is also unlikely to reflect a strategy peculiar to the peak procedure because some subjects show a positive scalar error, some show a negative scalar error, and some show almost no scalar error (see Figure 2). While many performance factors could increase or decrease response latencies, they should have fixed, rather than proportional, effects (i.e., their effect should be independent of the latency remembered). The fact that the error is proportional to the remembered latency strongly constrains the plausible source(s) for that error; it (they) must multiply the contents of memory.

In SET, these systematic errors have often been accounted for by postulating a multiplicative calibration error between the payoff latencies as measured by the timer and the values that are retrieved from memory (Gibbon et al., 1984)[1]. Specifically, in Gibbon's (1984) model, the memory stage delivers to the comparator a remembered target latency whose expectation differs from the true latency by a multiplicative factor. This factor has usually been symbolized by K*. To be consistent with our notational conventions (Table 2 in the Appendix), we adopt the notation K_m.

If memory calibration can be shown to be the primary source of the systematic scalar errors in peak location, then between-subject variation in the memory calibration constant, K_m, can be viewed as analogous to the between-subject variation in *tau*, the period of the circadian clock. That is, the behaviorally observed memory errors may reflect genetically derived variability in the underlying molecular mechanism. This could give us a means of finding the relevant genes, just as behaviorally observed genetic variation in *tau* has done in the case of the circadian clock (Antoch et al., 1997; Dunlap, 1993)

The smooth rises and falls in Figure 2 are averaging artifacts. On any given peak trial, responding has the form of a square pulse, with an abrupt onset, an abrupt offset, and a steady high rate during the response interval. The onset and offset times are determined by the start and stop criteria, which are approximately proportional to the target time (the remembered payoff latency). That is, subjects start when the elapsed time has reached a certain percentage of the target time and stop when it has exceeded the target time by some percentage. Asymmetry of start and stop proportions is another potential source of systematic error in the location of the peak response probability. If, as has often been assumed, these proportions differ from the remembered payoff latency by a common hedge factor (minus and

plus the same percentage), then the center of the response interval is the remembered payoff latency. If, however, these proportions are independent of one another, then the center of the response interval need not be the remembered pay-off latency. For the peak procedure to be a successful memory screen, it is important to determine if this possibility can be ruled out.

We turn now to the second aspect of the data in Figure 2 that might be a good target for memory screening, the between subject differences in trial-to-trial variability. For the scalar variability (proportional spread) seen in the peak data, there are three potential sources. First, it is often assumed that that memory itself is a major source (Church, et al., 1994; Gibbon, 1994); the remembered payoff latency is assumed to vary from trial to trial. Other likely sources are trial-to-trial variability in the start and stop proportions, and in trial-to-trial variability in clock speed. Our ability to estimate the relative contributions of these sources determines the utility of the peak procedure as a screen for genetically controlled quantitative variation in the memory mechanism.

The successful use of the peak procedure (or similar tasks) as a behavioral screen for memory mutants depends on its application to individual subjects (group data is of little use here) and the use of subjects amenable to genetic analysis. A great deal of work has been done on analyzing and modeling the group performance of rats and pigeons in the peak procedure. However, little work has been done on analyzing and modeling the performance of individual subjects. In addition, there has been little peak procedure work using mice, which offer the greatest opportunity for genetic analysis.

In this paper, we report a variant of the peak procedure designed specifically for mice. Our main focus, however, is on the application of different analytic methods to the extraction from the data of estimates of the scalar memory error and memory variability, and on the closely related problems of distinguishing between scalar memory error and asymmetrical decision criteria, and between variability due to memory and variability from other sources.

Method

Subjects

The subjects were 2 female Swiss-Webster ND4 mice (Harlan, Indiana, U.S.A.) weighing between 25-26 g at the start of the experiment. Subjects were housed individually in clear shoebox cages and maintained on a 12:12 h light:dark schedule with lights on at 08:00 h. All training took place during the dark portion of the photoperiod. Subjects were left undisturbed in the colony room for 1 week after arrival. At the end of the first week, access to food was restricted and training was begun. During training, the subjects obtained 30-70% of their total food intake while in the operant chamber and this was supplemented with 1-3 g of additional food at the end of each training session. The amount of supplemental food delivered depended on the amount of food obtained during the training session and was adjusted on a daily basis to prevent excessive weight loss. Water was available ad lib. in both the home cage and operant chamber.

Apparatus

All training took place in operant chambers (MED Associates, Vermont, U.S.A., model # ENV307AW, 7 x 8 x 5 in). The chambers were located in individual ventilated, sound-attenuating boxes. Each chamber was equipped with three pellet dispensers each connected to a feeding station along one wall of the chamber. A control station (identical to the feeding stations, but not connected

to a pellet dispenser) was located on the opposite wall. Each station was equipped with an infrared beam that detected nose pokes into that station, and a light that illuminated the pellet delivery area. The chambers were also each equipped with a tone generator (2900 Hz) and a white noise generator (80 dB, flat 10-25,000 Hz). When activated, the pellet dispensers delivered a single 20-mg food pellet (Noyes, New Hampshire, U.S.A., #PJA/100020).

Procedure

Subjects were weighed at the start of each session and then placed in the operant chambers. 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 were terminated, the target feeding station was illuminated, the tone turned on, and the feeding clock started. Feeding trials were fixed-interval 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 beam was interrupted at that moment; otherwise, it was delivered at the first detected interruption following T^*. Illumination of the target feeding station was terminated 10 s following pellet delivery. On trials where no nose poke was detected 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, occurred on an average of one out of five trials. On these trials, the feeding station remained illuminated and the tone 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 this interval, the control station was again illuminated and the white noise turned back on, enabling the mouse to start the next trial. Sessions lasted 12 hours. The number of trials in a session was controlled by the subject's behavior, and ranged from 75 to 250. All aspects of the experimental protocol were controlled by computer software (MED-PC, MED Associates). The subject's interruption of the infrared beams was recorded with a temporal resolution of 20 ms.

Subjects were initially trained with a target latency of 20 s, and this protocol remained in effect for 16 sessions. Following this, the target latency was changed to 30 s for 15 sessions and then to 10 s for 10 sessions. For data analysis, we used all the sessions after the session at which visual inspection of the raster plots (see Figure 3) suggested that behavior had stabilized. For M302, this was the last 7 sessions in each condition. For M301, it was the last 9 sessions in the $T^* = 20$ s and 10 s conditions and the last 14 sessions in the $T^* = 30$ s condition.

Results

The subject's behavior on probe trials may be visualized by means of a raster plot in which the x-axis represents time since the beginning of a probe trial and the y-axis represents the sequence of probe trials within a session. A plotted point indicates that the subject had its head in the lit hopper at that moment (x axis) on that trial (y axis). Figure 3 shows raster plots for subject 301 and subject 302 for their final session with a T^* of 10 s and in the final session with a T^* of 30 s.

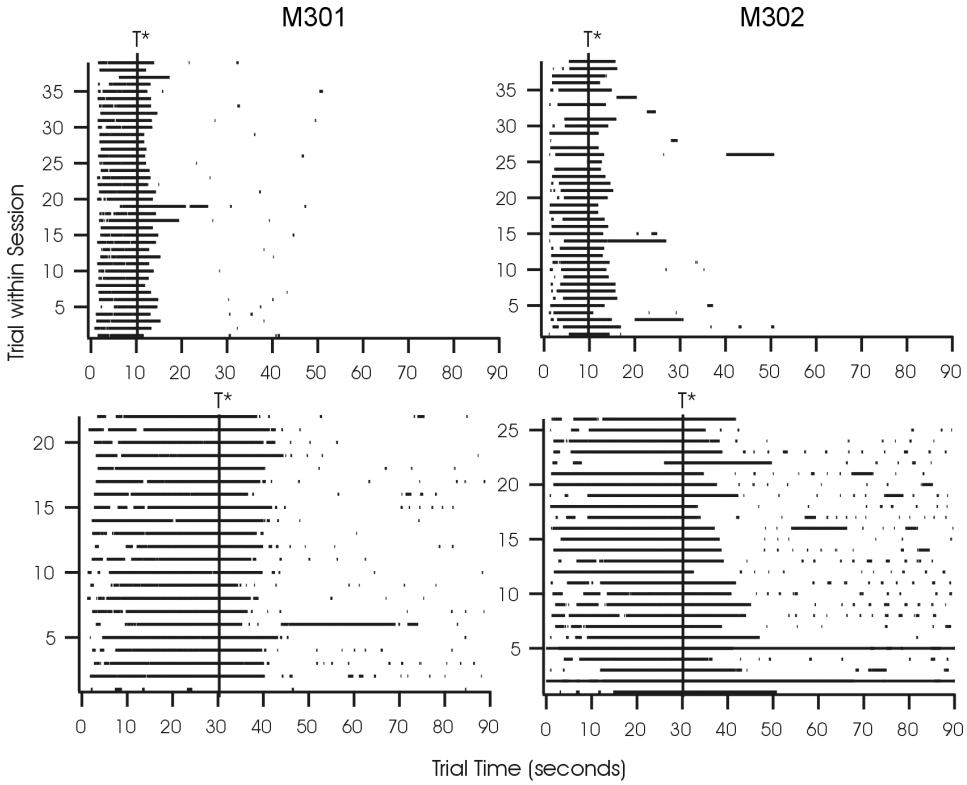

Figure 3. Raster plots of head-poke behavior on probe trials in sessions with different feeding latencies (*T** = 10 s in upper panels, 30 s in lower). Each horizontal line is one trial. The times during which the head was interrupting the beam in the feeding hopper are black. The vertical lines indicate T*.

It is evident in Figure 3 that the mouse's asymptotic behavior in our procedure is characterized by a break-run-break pattern, consistent with what has been found with the rat and pigeon (Cheng & Westwood, 1993; Church et al., 1994; Gibbon & Church, 1992). There is a substantial "run" interval around the time food is expected during which the mouse had its head in the hopper more or less continuously. Before and after this interval, the subject samples the hopper only intermittently. Figure 4 shows plots of the cumulative time that the head is in the hopper minus the cumulative time it is out of the hopper. Before and after the breaks, at the start and end of a run, the head is out of the hopper much more than it is in, so this plot has a strong downward slope before and after the run. During the run, the head is in much more than it is out, so the slope is strongly upward. The start and stop of the run are the locations of the pretarget minimum and the posttarget maximum of this plot. The reversal in slope at the start and stop was usually sharply defined. The uppermost plot in Figure 4 is representative. During the run, the head was in the hole 97% of the time, which is the median value for this measure over the roughly 1400 trials. The lower two plots show the worst cases in our data set, the runs with the lowest occupancies. These runs, with occupancies of 61% and 57%, are outliers. The mean and mean median occupancies were 95% and 97%, respectively, with an average standard deviation about the mean of 6 percentage points. In short, on the great majority of runs, the head was in the hole

more than 90% of the time during a run. Outside the bounds of a run, it was rarely in the hole, as may be seen from the examples in Figure 4.

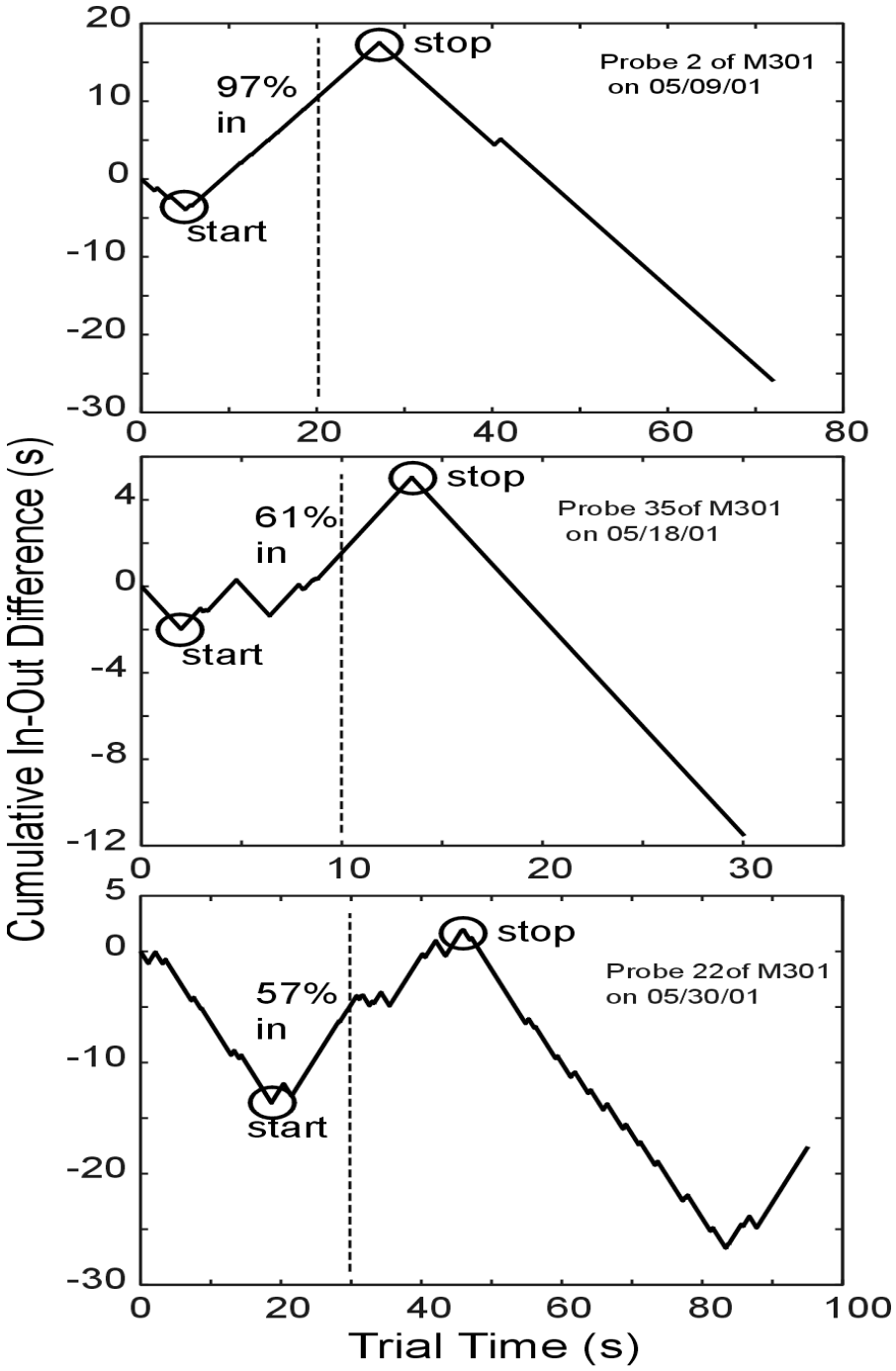

Figure 4. Cumulative time spent with the head in the feeding hopper minus cumulative time spent with the head out of the feeding hopper. The top panel shows data from a typical trial, and the middle and bottom panels show the two trials in our data set with the least amount of time spent with the head in the feeding hopper (the two 'worst' trials). Circles indicate the start and stop times determined for the response interval.

Following previous work, trials on which this interval started after, or ended before, *T** were not considered. In addition, data were discarded trials when the start time, stop time, or duration (stop-start time) was more than 3 standard deviations from the mean. These considerations combined led to the discarding of less than 2% of the probe trials.

A start time, stop time, duration (stop time-start time) and center time ([start time + stop time] / 2) were recorded for each valid probe trial. These values formed the basis for analysis of distributions, their variances and covariances. As pointed out by Church et al., (1994), there are a limited number of degrees of freedom in these measures. Specifying the means and variances of the start and stop times plus the correlation (covariance) between them, determines all other means, variances, and correlations. Thus, we focus our attention on these key measures.

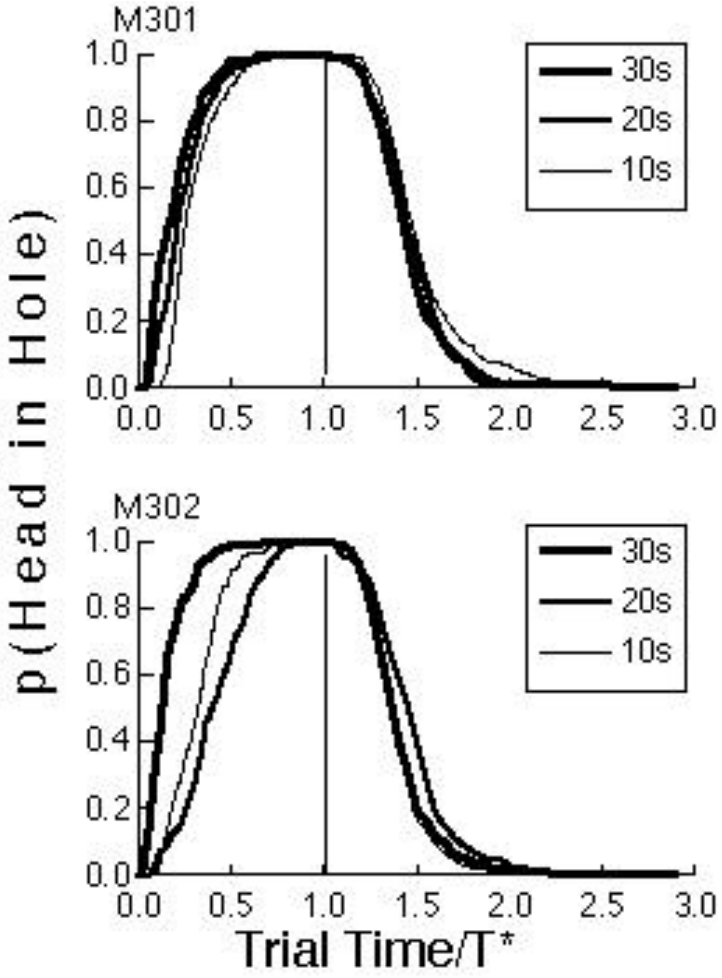

Figure 5. Normalized peak curves for two mice at three different feeding latencies (*T**). Only times falling between start times and stop times (as defined in text) were used in generating these plots. The probability that the mouse had its head in the feeding hopper at any given fraction of probe trial time (Trial Time/*T**) is the probability that it had started minus the probability that it had stopped, in other words, the cumulative distribution of starts minus the cumulative distribution of stops.

Figure 5 shows the normalized peak curves for the three target latencies for our two subjects. The normalized data from M301 superpose rather closely (the scalar property), whereas the data from M302 do not, the failure occurring mostly on the start side (rising side) of the peak curve. These individual subject peak plots (particularly that for M301) and the quantitative features of our data (Table 1) are similar to what has been reported for pigeons. In particular, these distributions exhibit scalar error in the location of the peak, that is, they are not centered on the target latency, and the amount by which they depart from the target latency is an approximately fixed proportion of that latency (at least in M301, see rightmost column of Table 1). The question is whether this error can be ascribed to the mechanism responsible for memory.

Table 1
Summary Statistics.

Subject,*T**	cvs				Correlations				
	Starts	Durations	Centers	Stops	start-stop	start-dur	start-center	dur-center	T_c/T^*
M301,10	0.49	0.21	0.19	0.18	0.29	-0.25	0.66	0.57	0.88
M301,20	0.57	0.20	0.17	0.15	0.18	-0.41	0.63	0.45	0.86
M301,30	0.67	0.17	0.16	0.13	0.22	-0.42	0.68	0.38	0.82
M302,10	0.52	0.19	0.16	0.11	0.29	-0.66	0.83	-0.12	0.87
M302,20	0.47	0.23	0.19	0.16	0.42	-0.45	0.82	0.15	0.95
M302,30	0.80	0.15	0.17	0.13	0.36	-0.34	0.75	0.36	0.78
M	0.59	0.19	0.17	0.14	0.29	-0.42	0.73	0.30	
SD	0.13	0.03	0.02	0.02	0.09	0.14	0.08	0.25	
Church, et al.									
M	0.50	0.28	0.22	0.21	0.31	-0.33	0.70	0.41	1.06
SD	0.07	0.04	0.02	0.03	0.08	0.10	0.07	0.14	0.08

Note. T* = feeder-arming latency; T_c = mean center (of the head-poke interval); cv = coefficient of variation; M = mean; SD = standard deviation.

Testing for the Scalar Property

The extent to which peak data obtained with different feeding latencies satisfy the scalar property may be assessed by comparing the normalized cumulative distributions (Figure 6). In these plots, the temporal dimension on the x-axis is normalized with respect to *T** (thus, 1 on the x-axis always corresponds to *T**). The cumulative normalized start distributions (top panels in Figure 5) attain asymptote before the value on the x-axis reaches 1, because all starts retained in the analysis occur before *T**. Similarly, the cumulative normalized stop distributions

(second row in Figure 5) begin to rise only when normalized trial time exceeds 1. The cumulative normalized distributions for the center times (third row of Figure 6) would have risen to half their asymptote at T^* if there were no scalar error. They are, in fact, well above this value at T^* in our two subjects because their head-pokes were centered too early (scalar error). The cumulative normalized distributions for the durations show most of their rise beyond 1, because most head-poke durations were longer than T^*.

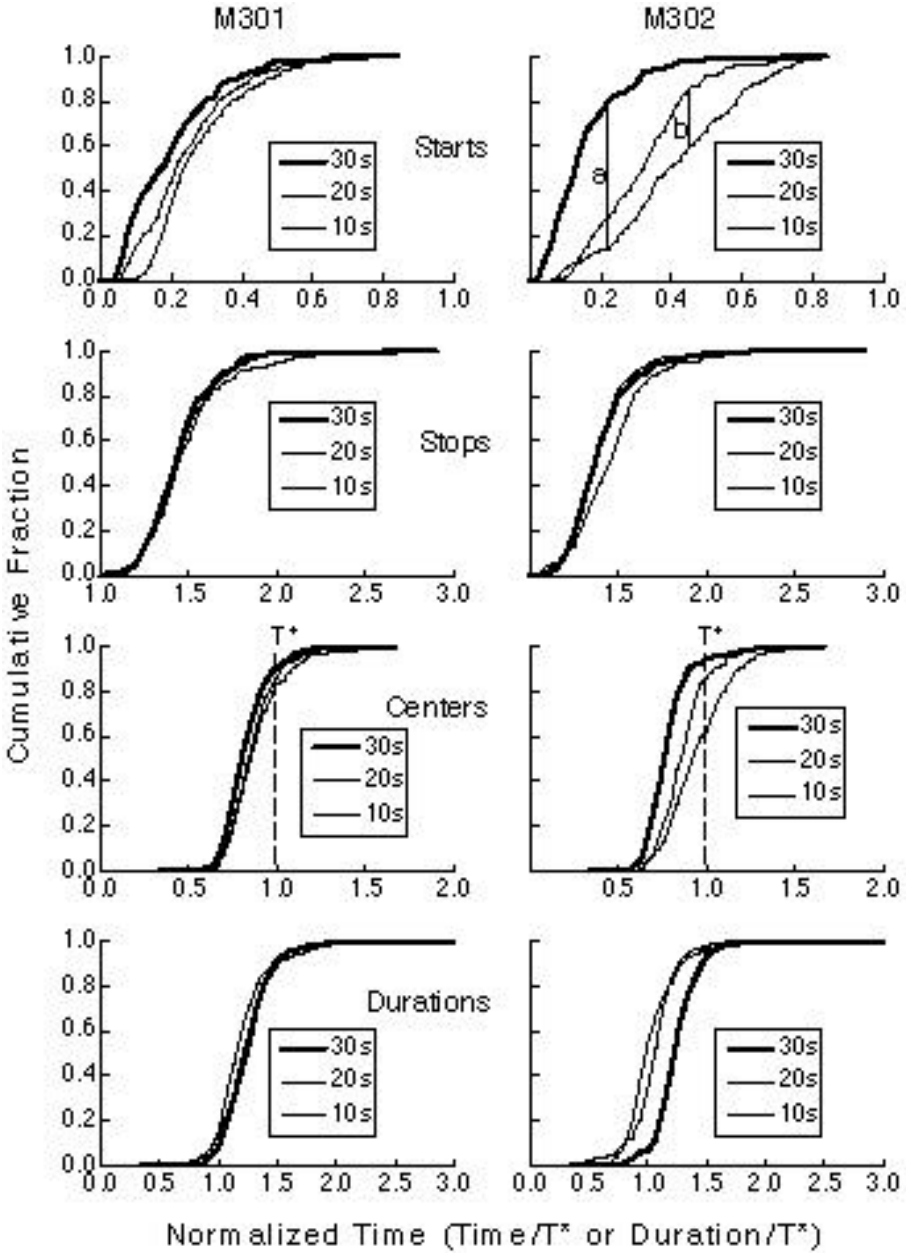

Figure 6. The normalized cumulative distributions. The Kolmogorov-Smirnoff K statistic is the maximum deviation between two normalized cumulative distributions (thin vertical lines labeled a & b in the uppermost right hand panel).

For Subject M301, the distributions (except for starts) are time-scale invariant; that is, they superpose when normalized. Moreover, the pattern of failure of time-scale invariance in the start distributions is perhaps explicable on the assumption that there is an additive factor (e.g., minimum latency to begin responding after starting a trial) truncating these distributions on their left. This would produce the observed pattern of failure, namely, that the 10 s distribution rises more steeply and lies to the right of the 20 s distribution, which in turn rises more steeply and lies to the right of the 30 s distribution.

For Subject M302, the violations of time-scale invariance are much greater, not confined to the start distributions, and less readily explicable. The thin vertical lines in the upper right panel of Figure 6 show this discrepancy for the comparison of the 20 s distribution to the 30 s distribution (line labeled a), and for the comparison of the 10 s distribution to the 20 s distribution (line labeled b). Calculation of the Kolmogorov-Smirnoff statistic suggests that both of the discrepancies seen in the upper right panel are unlikely to have arisen by chance ($p << 0.001$). In the case of M302, these discrepancies cannot have been simply a performance factor that established a minimum possible start latency, because the ordering of the distributions is wrong. It seems, rather, that the differing experimental conditions led to different average start criteria in this subject, but not to different average stop criteria. This suggests that the start and stop criteria are not determined by a common hedge factor.

The stop distributions for M302 are approximately time-scale invariant (right panel in row 2 of Figure 6). The departures from time-scale invariance in the distributions for the centers and durations in this subject are secondary consequences of the discrepancies in the start distributions. The distributions of the centers for the 10 and 20 s intervals are shifted to relatively later in the trials because the starts occur relatively later while the stops are relatively time-scale invariant. The distributions of the durations for the 10 and 20 s intervals are relatively shorter for the same reason. The results from M302 imply that, in this subject at least, the start distribution was independent of the stop distribution and influenced by different factors.

Modeling and Parameter Estimation

One Hedge vs. Two Hedges. Within the basic SET framework there are many variants. (For a consideration of several of these variants and the relevance of covariance patterns for deciding between them, see Gibbon & Church, 1990.) Two questions are of particular importance from our memory-screening perspective: What are the possible sources of the scalar error in the location of the peak and what are the principal sources of the scalar variability observed in the behavior?

If SET is to be used to estimate parameters of an underlying memory process from peak data, the first question to be addressed is how well the model describes the data from individual subjects. Given that few such analyses are available in the literature, we here examine SET analytically and compare analytically-derived parameter estimates with our empirical data. Due to length constraints, we report here only important results of these analyses. For the analytic estimation of model parameters, see Appendix A.

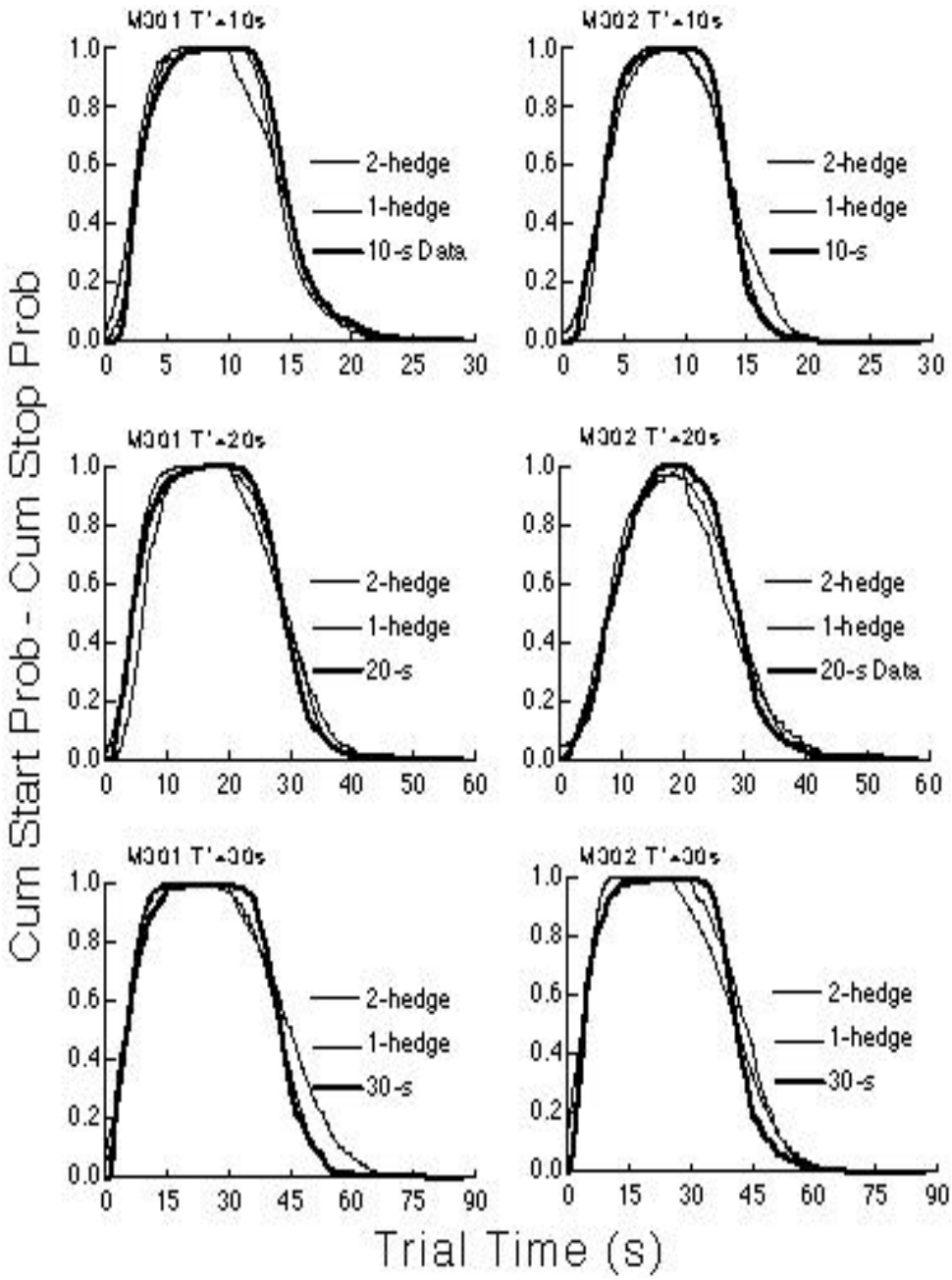

Figure 7. Four examples of peak data fitted by the single-hedge and double-hedge versions of SET, using analytically derived estimates of the model parameters.

Figure 7 shows the analytic fits of the 1-hedge and 2-hedge models to the data. These are the fits obtained when the model parameters are estimated analytically, rather than by an iterative search for the permissible parameters that minimize the residual variance. The parameter estimation formula can, and sometimes did, yield negative variance estimates for the distribution of the stop threshold in

the double-hedge model, because the variability in the stop distributions is about the same as the covariance between starts and stops, which means that, according to the double-hedge model, memory variability is the only significant source of variability in stop times. In these cases, we used that assumption to model the data; that is, we took the variability in the underlying distribution of stop proportions to be negligible.

By some standards, the fits in Figure 7 are good; they account for a high percentage of the variance. However, the discrepancies between the data and the model's predictions are clearly systematic. The observed distributions are more skewed than the distributions generated by the models. One reason for the skew in the start distributions may be that they are truncated on their left by the same factors that cause the failures to obtain scalar variability in start distributions. Given these truncating factors, it is perhaps surprising that the model's predictions fit the data as well as they do. The reasons for the skew in the stop data are less obvious. Church et al. (1991) investigated the sources of asymmetry in peak data at some length and identified several factors that skew the right side of the stop distribution.

The important point for our purpose is that the double-hedge model fits the data at least as well as the single-hedge model. On the one hand, this is not surprising, because the double-hedge model has one more parameter. However, as Church, et al. (1994) point out, the patterns of variance and covariance also permit one to determine which is the better model, and these favor the double-hedge model in our data, just as they did in the data of Church, et al. (1994). The single-hedge model predicts that the cv for distributions of the centers of the head-poke intervals should be less than the cv for both the start and the stop distributions, whereas, more often than not the cv of the stops is slightly less than the cv of the centers (Table 1). In other words, the variance in the centers is bigger than it ought to be relative to the variance in the stops. Also, the location of the center of a head-poke interval is not as strongly correlated with the start of the interval as the single-hedge model says it should be. Church, et al. (1994) showed that in the single-hedge model, this correlation should be equal to the ratio between the cv for centers [cv(d)] and the cv for durations [cv(d)], whereas in both our data and theirs, it is less than this ratio, often much less (Figure 8)

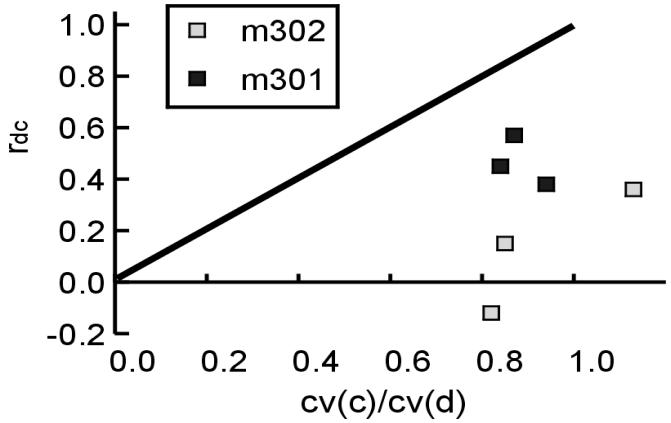

Figure 8. The correlations between durations and centers (ordinate) plotted against the ratio of the cvs. In the single hedge model, these points should cluster around the positive diagonal (solid line).

In summary, the data give no reason to reject a double-hedge model; on the contrary, they favor it. Therefore, if starts and stops are independently hedged, the deviation of the center of the peak curve from the target latency is not an estimate of scalar memory error; in a two hedge version of SET no such estimate is possible. Our modeling with two independent hedge factors emphasized this point by assuming that the scalar error in memory was zero. The expectation of the memory distribution was not a free parameter; it was set to the target time.

Memory Variability. Which model one adopts is also important for estimating the magnitude of memory variability. In the single-hedge model, noise in the hedge factor has equal and opposite effects on the start and stop thresholds. A smaller than usual hedge factor leads to a late start and an equally early stop. Thus, variation in the centers of the head-poke intervals is due entirely to memory variance. If, however, start and stop thresholds come from independent distributions, then variance in the decision criteria does contribute to variance in the centers. In the double-hedge model, however, the start-stop covariance is proportional to the memory variance, because an unusually long sample from the memory distribution leads to a late start and a late stop. Thus, which aspect of the peak data estimates the memory variance depends on which model one takes to be the right model.

In many published treatments of SET models, clock variance and memory variance are treated as interchangeable (e.g., Church, et al., 1994) on the implicit or explicit assumption that memory variability is clock variability, because target latencies sampled from memory come from the population of experienced target latencies. Under this assumption, the variability in the memory samples and the variability in the experienced target latencies are identical results of trial-to-trial variability in clock speed on feeding trials (Gibbon, 1977). Of course, it is possible that the variability in memory samples arises independently of variability in the input.

More importantly, if trial-to-trial variation in clock speed is the source of memory variability, then this clock variability should determine the observed variation in start and stop times in two ways, only one of which is reflected in the formulae in Table 3 (Appendix). On a trial when the clock runs slower than usual, the experienced feeding latency will be shorter than usual, resulting in a shorter than usual record in memory. Similarly, on a trial when the clock runs faster than usual, the experienced feeding latency will be longer than usual. As already explained, this variation in experienced (subjective) feeding latencies will show up as variation in the remembered feeding latencies on probe trials.[2] However, the variation in clock speed on the probe trials themselves will also result in variation in the observed start and stop times. Whatever the start criterion is, it will take more time to reach it on trials when the clock runs slow and less time on trials when the clock runs fast.[3]

To address this issue, we developed a computer program (Appendix B) capable of simulating a variety of SET models, including those where clock variance and memory variance are treated as separate sources of variability. Numerical simulation is necessary because when both memory variance and clock variance are treated as independent Gaussian variables, expressions involving the ratio between these variables are analytically intractable (they have undefined variances).

In running our first simulation, we set the parameters of the model so as to duplicate the assumptions made by Church et al. (1994) in deriving the formulae relating observed expectations and variances to the expectations and variances in SET variables: They implicitly assumed no variation in clock speed on probe trials, so we set this variation to zero in our simulation. We set the expectation of the memory distribution to T^* (the target time), and we set the expectations and standard deviations of the distributions for the start and stop threshold proportions (L and U) to the values we obtained analytically. We then varied the coefficient of variation for the distribution of $\hat{m}$ (that is, memory variability).

In the second simulation, we explored the effect of positing no memory variance, only clock-speed variance. We set memory variability to zero and varied the clock speed. In the third simulation, we allowed both memory variability and variability in clock speed. The first two simulations are the limiting cases of the third simulation, in which there is variation in both clock speed and memory samples, as there would have to be in any model that posited variation in clock speed as the source of variation in memory samples.

Figure 9 shows the results from the limiting cases (Simulations 1 and 2). A model in which clock speed varied but memory samples did not (dashed curves in Figure 9) tended to do better than a model in which memory samples varied but clock speed did not (solid curves in Figure 9). However, for most data sets, the goodness of the model's fit did not depend strongly on the values of these gammas. In particular, the fits obtained with gammas of zero were no worse and sometimes better than the fits obtained with the analytically estimated values of gamma.

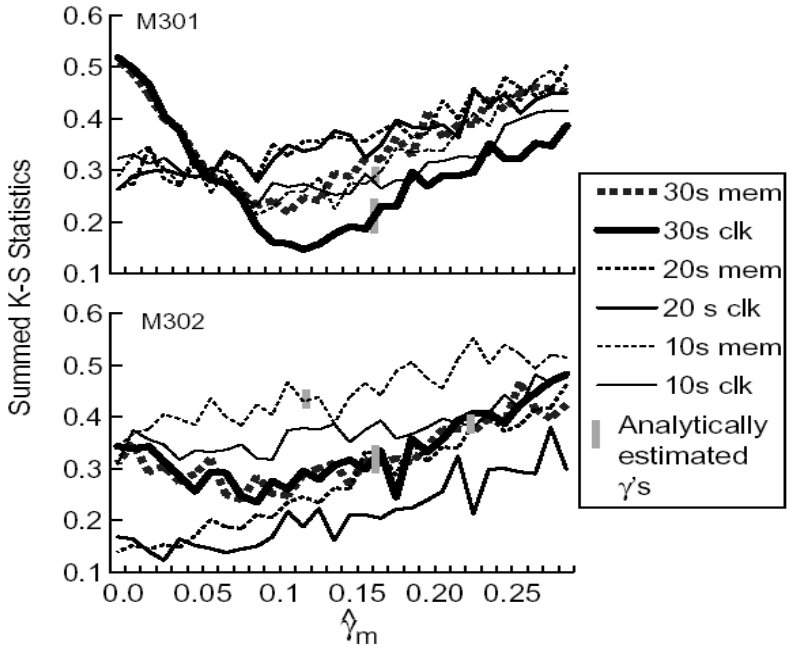

Figure 9. The effect of varying the amount of memory variance or the amount of clock variance on the goodness of the model's fit to the empirically obtained start and stop distributions. The measure of goodness of fit (summed K-S statistic) was the sum of the maximum deviations between the cumulative normalized start and stop distributions generated by the model and the empirically obtained distributions. $\hat{\gamma}_m$ is the coefficient of variation (in either the memory distribution or the clock speed distribution).

The simulation with both memory variation and clock variation did not change this picture: The best fit when both sources of covariation were present was only marginally better than the best fit in the limiting case where only clock speed was allowed to vary.

The most important conclusion that emerges from these simulations is that there are two potential sources of the modest covariation between starts and stops seen in our data and in the data of others (Cheng, et al., 1993; Cheng & Westwood, 1993; Church, et al., 1994; Gibbon & Church, 1990)–trial-to-trial variation in clock speed and trial to trial variation in the target value sampled from memory. Variation in clock speed may or may not be the source of the variation in memory samples. Whether it is or not, it would seem that in these data one cannot separately estimate the two contributions—memory variance and clock variance.

Discussion

Although our primary focus is theoretical rather than empirical, our peak-procedure data from two mice are strikingly similar to data from essentially the same procedure used with pigeons, rats (Church et al., 1994), and humans (Rakitin et al., 1998). Regardless of the subject species, the data are approximately time-scale invariant and the departures from time-scale invariance are greater for start distributions than for stop distributions. The patterns of variance and covariance are also the same across species: There is a positive start-stop correlation and a negative start-duration correlation, the cv for the centers is smaller than the cv for starts but the same or even slightly greater than the cv for stops, and the correlation between center and duration is greater than the ratio of the corresponding cvs, often much greater.

In short, the psychophysical characteristics of the system that mediates the attempt to reproduce a remembered interval appear to be invariant across a wide range of higher vertebrates. While it is possible that the particular failures of time-scale invariance in the present data are the result of behavioral tendencies particular to the mouse, we believe that a genetic dissection of this system in the mouse would have broadly applicable results.

The human data differ quantitatively from the rat, pigeon and our mouse data in two respects: The peaks are much sharper and narrower, and the start-stop correlations are higher. The ratio of the width of the peak at half-maximum (mean stop minus mean start) to its mode was 1.1-1.6 in our two mice and 1.2 +/-0.08 in Church, et al.'s (1994) rats, whereas it was 0.34 – 0.4 in the human data (see Table 2 on p. 22 of Rakitin et al., 1998). In rats and pigeons, and in our two mice, the start-stop correlations are roughly 0.3 ± 0.09, whereas in humans they are roughly 0.9 ± 0.1 (see Figure 6 on p. 24 of Rakitin et al., 1998).

We have used our data to investigate the question whether peak data alone permit one to estimate two quantitative properties of the memory mechanism: The trial-to-trial variability in a sample from memory, and the scalar error, that is the proportion by which the mean remembered duration differs from the mean of the originally experienced durations. We conclude that that peak data alone do not permit one to estimate these memory parameters.

Whether one can estimate the scalar error from peak data hinges on whether there are or are not separate hedge factors for the start and stop criteria. If

there is a single hedge factor, then the start and stop criteria on any trial are equidistant from the target sampled from memory, and systematic deviation from the target time implies systematic error in the memory samples. (Cheng, 1992) found that when he penalized pigeons for starting sooner than half way to the feeding (target) time, the pigeons started later and stopped sooner. That is, both the start criterion and the stop criterion moved closer to the target time. This counterintuitive result suggests that there is only one hedge factor.

However, data from our two mice, like comparable data from pigeons (Cheng & Westwood, 1993; Gibbon & Church, 1992), rats (Church et al., 1994), and humans (Rakitin et al., 1998), are more consistent with a two-hedge model than with a single-hedge model. In the single-hedge model, variance in the hedge factor contributes to variance in the starting and stopping times, but not to variance in the center because a small hedge produces a late start and an early stop, with mutually annulling effects on the center. Thus, only variance in the memory sample causes variance in the center. This source also contributes effects of equal magnitude to the start and stop times. Thus, the variance of the center should be relatively less than both of the latter variances. In fact, however, the cv for the center is consistently as large or larger than the cv for the stop–in our data and in others' (Cheng & Westwood, 1993; Church et al., 1994; Gibbon & Church, 1992).

Second, in the single-hedge model, variance in the memory sample is the sole source of the variance of the center but one of two sources for the variance of the duration (the other being variance in the hedge factor). If memory were the sole source for both center and duration variance, then the cvs for the center and the duration would be equal and the correlation between the two variables would be perfect; the duration of a head poke would vary from trial to trial in proportion as the centers were late or early. However, to the extent that variance in the hedge factor contributes to variance in the duration, the cv for the duration should be greater than the cv for the center, and the correlation between the two variables should be correspondingly less. Thus, the estimate for the correlation should on average equal the ratio of the two cv estimates. In fact, however, the estimated correlation is reliably less than this ratio (Figure 8), as predicted by a two-hedge model.

Finally, it is clear in our data and in that of others that various experimental factors can affect the start criterion either more or less than the stop criterion, which would not be possible if there were only one hedge factor common to the two criteria. In the data presented here for M302, changing the target latency had non-scalar effects on the start criterion but scalar effects on the stop criterion. Church et al. (1994, Figure 3, p. 138) found that the duration of peak trials affected the start criterion more than the stop criterion, whereas the stop criterion was affected more than the start criterion by whether the trial preceding the peak trial had been a feeding trial or another peak trial (a trial with no feeding—see their Figure 5, p. 139).

We conclude that peak data alone cannot distinguish variation in the degree of asymmetry between start and stop criteria from variation in systematic memory error when the start and stop criteria generally bracket the target time, as they do in our data (Figure 3). We note, however, that variation in the hedge factor or factors cannot shift the median of the start distribution beyond the target time nor the median of the stop distribution before the target time, whereas variation in

systematic memory error can shift either distribution arbitrarily far in either direction. In the human timing literature, taking Parkinson's patients off their L-DOPA can shift either the start or the stop distributions so as to place their medians close to or even slightly to the wrong side of the target value (Malapani et al., 1998), so effects of this magnitude could reasonably be looked for in genetic screening. A shift of the median to the wrong side of the target value would be strong evidence of systematic memory error. Our raster plots (Figure 3) display the raw data from the peak procedure and related procedures in such a way that an effect of this kind would be immediately evident, without statistical treatment or modeling.

We also conclude that one cannot estimate the memory variance from peak data alone, again because of uncertainty about the correct version of SET. If, as appears likely, the start and stop criteria are separately determined, then the covariance of the start and stop thresholds is the measure of memory variance—but only if the effects of the variation in clock speed on the probe trials themselves (as opposed to on feeding trials) are assumed to be negligible. If, as has often been assumed, variation in clock speed is the source of variation in memory samples, then it is implausible to assume that this variation is not also a factor on the peak trials themselves. In any event, one needs a procedure that enables one to distinguish between covariation due to variation in the clock speed on peak trials and variation in the memory samples.

There are variants of the peak procedure that produce multiple possible targets on a single trial (Cheng et al., 1993; Fetterman & Killeen, 1995; Mattell et al., 2001). In general, they place responses controlled by different target latencies in competition, so that one response is appropriate during one part of the peak trial whereas another response becomes appropriate during another part of the same trial. These are likely to be useful in two respects: They should allow us to move start and stop criteria around independently (assuming that they can in fact shift independently), and they allow one to deconfound memory variance and clock variance (Cheng et al., 1993). Thus, we think that using variants of the peak procedure to screen for abnormalities in the quantitative properties of memory merits continued investigation.

References

Antoch, M. P., Song, E. J., Chang, A. M., Vitaterna, M. H., Zhao, Y., Wilsbacher, L. D., Sangoram, A. M., King, D. P., Pinto, L. H., & Takahashi, J. S. (1997). Functional identification of the mouse circadian Clock gene by transgenic BAC rescue. *Cell*, **89**, 655-667.

Aronson, L., Balsam, P. D., & Gibbon, J. (1993). Temporal comparator rules and responding in multiple schedules. *Animal Learning and Behavior,* **21**, 293-302.

Brunner, D., Fairhurst, S., Stolovitzky, G., & Gibbon, J. (1997). Mnemonics for variability: Remembering food delay. *Journal of Experimental Psychology: Animal Behavior Processes,* **23**, 68-83.

Catania, A. C., & Reynolds, G. S. (1968). A quantitative analysis of the responding maintained by interval schedules of reinforcement. *Journal of the Experirmantal Analysis of Behavior*, **11**, 327–383.

Cheng, K. (1992). The form of timing distributions in pigeons under penalties for responding early. *Animal Learning and Behavior*, **20**, 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 beahvior: 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**, 135-155.

Church, R. M., Miller, K. D., Meck, W. H., & Gibbon, J. (1991). Symmetrical and asymmetrical sources of variance in temporal generalization. *Animal Learning and Behavior*, **19**, 207-214.

Dunlap, J. C. (1993). Genetic analysis of circadian clocks. *Annual Review of Physiology*, **55**, 683-728.

Fantino, E., & Goldshmidt, J. N. (2000). Differences, not ratios, control choice in an experimental analogue to foraging. *Psychological Science*, **11**, 229-233.

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. (1990). *The organization of learning*. Cambridge, MA: Bradford Books/MIT Press.

Gallistel, C. R. (1999). Can a decay process explain the timing of conditioned responses? *Journal of the Experimental Analysis of Behavior*, **71**, 264-271.

Gallistel, C. R., & 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. (1992). Ubiquity of scalar timing with a Poisson clock. *Journal of Mathematical Psychology*, **36**, 283-293.

Gibbon, J., & Church, R. M. (1981). Time left: linear versus logarithmic subjective time. *Journal of Experimental Psychology: Animal Behavior Processes*, **7**, 87-107.

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.

Gibbon, J., & Fairhurst, S. (1994). Ratio versus difference comparators in choice. *Journal of the Experimental Analysis of Behavior*, **62**, 409-434.

Gibbon, J., Malapani, C., Dale, C. L., & Gallistel, C. R. (1997). Toward a neurobiology of temporal cognition: Advances and challenges. *Current Opinion in Neurobiology*, **7**, 170-184.

Leak, T. M., & Gibbon, J. (1995). Simultaneous timing of multiple intervals: Implications for the scalar property. *Journal of Experimental Psychology: Animal Behavior Processes*, **21**, 3-19.

Malapani, C., Rakitin, B. C., Levy, R., Meck, W. H., Deweer, B., Dubois, B., & Gibbon, J. (1998). Coupled temporal memories in Parkinson's disease: A dopamine-related dysfunction. *Journal of Cognitive Neuroscience*, **10**, 316-331.

Mattell, M. S., King, G. R., & Meck, W. H. (2001). Differential modulation of peak times in the tri-peak procedure by the chronic administration of intermittent or continuous cocaine. MS (pp. 29).

Meck, W. H., & Church, R. M. (1984). Simultaneous temporal processing. *Journal of Experimental Psychology: Animal Behavior Processes*, **10**, 1-29.

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**, 15-33.

Rescorla, R. A. (1998). Instrumental learning: Nature and persistence. In M. Sabourin & F. I. M. Craik & M. Roberts (Eds.), *Proceedings of the XXVI International Congress of Psychology: Vol. 2. Advances in psychological science: Biological and cognitive aspects, Vol. 2* (pp. 239-258). London: Psychology Press.

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.

Roberts, W. A., & Boisvert, M. J. (1998). Using the peak procedure to measure timing and counting processes in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, **24**, 416-430.

Tang, Y.-P., Shimizu, E., Dube, G. R., Rampon, C., Kerchner, G. A., Zhuo, M., Liu, G., & Tsien, J. Z. (1999). Genetic enhancement of learning and memory in mice. *Nature*, **401**, 63-69.

Vicens, P., Bernal, M. C., Carrasco, M. C., & Redolat, R. (1999). Previous training in the water maze: Differential effects in NMRI and C57BL mice. *Physiology and Behavior*, **67**. 197-203.

Wilkie, D. M., Willson, R. J., & Carr, J. A. R. (1999). Errors made by animals in memory paradigms are not always due to failure of memory. *Neuroscience and Biobehavioral Reviews*, **23**, 451-455.

Footnotes

[1] The error may arise either in the write-to-memory operation or in the read-from-memory operation or as a combined result of quantitative errors in these two elementary operations.

[2] The resulting distribution of experienced feeding latencies in memory will be skewed toward longer latencies, because the experienced feeding latency tends to zero as clock speed tends to zero and it tends to infinity as clock speed tends to infinity. The difference between a finite expectation and zero is equal to the expectation (hence, finite) while the difference between a finite expectation and infinity is itself infinite. However, the skew is small when the coefficient of variation in clock speed is of the order seen in our data (0.2).

[3] This effect also yields a skew towards longer times for the same reasons: observed start times tend toward infinity as clock speed tends toward zero.

Appendix A: Formulae and Derivations

The formulae used to estimate model parameters from the data are given in Table 1A. They derive from formulae in Church et al., (1994), which go back in turn to derivations in the Appendix to Gibbon and Church (1992). We briefly recapitulate essential steps, because it is difficult to pull them together.

Table 2
Notation.

Variable	Symbol
Objective time	t
Subjective time	$\hat{t}$
Target reward. time (arming latency)	T^*
Observed time of reward	t_r
Subjective time of reward	$\hat{t}_r$
Observed start time	t_l
Observed stop time	t_u
Observed center $(t_u - t_l)/2$	t_c
Memory scalar (whose expectation [$\hat{K}_m$] is commonly symbolized by *K*)*	$\hat{k}_m$
A reward latency sampled from memory	$\hat{m}$
Clock speed on a given trial	$\hat{s}$
Delay in starting the clock on a given trial	$\hat{d}$
Start criterion (as a proportion of $\hat{m}$)	$\hat{l}$
Stop criterion (as a proportion of $\hat{m}$)	$\hat{u}$

Start hedge factor (1-$\hat{l}$)	$\hat{b}_{\mathrm{l}}$
Stop hedge factor (1-$\hat{u}$)	$\hat{b}_{\mathrm{u}}$

Conventions: An italicized small letter is the name of a variable; for example, t = elapsed time in a trial. A hat (e.g. $\hat{t}$) indicates that it is a subjective variable (not directly measurable); thus $\hat{t}$ = subjective elapsed time in a trial. Absence of a hat indicates that it is an objective (directly measurable) quantity. Many variables are assumed Gaussian random, in which case, the upper case version of the letter is the expectation of the distribution, a lower case sigma subscripted by the letter (in text not italic form) is the standard deviation of the distribution, and a lower case gamma, similarly subscripted, is the coefficient of variation. An italicized *i* as a subscript indicate that a particular value, the *i*th value, of the random variable is being referred to. Thus, $\hat{m}$ refers to reinforcement times sampled from memory. $\hat{M}$ and $\hat{\sigma}_{\mathrm{m}}$ are the expectation and standard deviation of the sampled distribution, and $\hat{m}_i$ is the sample on trial *i*.

Table 3
Formulae for Estimating Model Parameters from Data.

Model & Model Parameter	*Estimation Formula*
Single Hedge Factor	
$\hat{M}$ = expectation of target latency distribution	$\mathrm{Med}(t_{\mathrm{c}})$
$\hat{\sigma}^2_{\mathrm{m}}$ = variance of remembered target latency distribution	σ^2_{c}
$\hat{B}$ = expectation of the hedge factor distribution	$\dfrac{\mathrm{Med}(t_{\mathrm{u}}) - \mathrm{Med}(t_{\mathrm{l}})}{2\mathrm{Med}(t_{\mathrm{c}})}$
$\hat{\sigma}^2_{\mathrm{b}}$ = variance of the hedge factor distribution = the variance of both the start & stop proportions	$\dfrac{\sigma^2_{t_{\mathrm{l}}} + \sigma^2_{t_{\mathrm{u}}} - 2\hat{\sigma}^2_{\mathrm{m}}(1+\hat{B}^2)}{2\left(\hat{\sigma}^2_{\mathrm{m}} + \hat{M}^2\right)}$, the average of the estimates obtained from the distributions of start and stop times
Two Hedge Factors	
$\hat{L}$ = expectation of start proportions	$\mathrm{Med}(t_{\mathrm{l}}) / T^*$
$\hat{U}$ = expectation of stop proportions	$\mathrm{Med}(t_{\mathrm{u}}) / T^*$
$\hat{\sigma}^2_{\mathrm{m}}$ = variance of remembered feeding latencies	$\dfrac{\mathrm{Cov}(t_{\mathrm{l}}, t_{\mathrm{u}})}{\hat{U}\hat{L}}$
$\hat{\sigma}^2_{\mathrm{l}}$ = variance of start proportions	$\dfrac{\sigma^2_{t_{\mathrm{l}}} - \hat{L}^2\hat{\sigma}^2_{\mathrm{m}}}{\hat{\sigma}^2_{\mathrm{m}} + T^{*2}}$
$\hat{\sigma}^2_{\mathrm{u}}$ = variance of stop proportions	$\dfrac{\sigma^2_{t_{\mathrm{u}}} - \hat{U}^2\hat{\sigma}^2_{\mathrm{m}}}{\hat{\sigma}^2_{\mathrm{m}} + T^{*2}}$

Notation for observed or known quantities

T^* = target latency (programmed feeding time)

t_{l} = start time

t_{u} = stop time

t_{c} = center point (middle) of a head-poke interval = $\dfrac{t_{\mathrm{l}} + t_{\mathrm{u}}}{2}$

$l = t_{\mathrm{l}}/t_{\mathrm{c}}$

$u = t_{\mathrm{u}}/t_{\mathrm{c}}$

σ^2_{c} = variance of center points

The memory for the target latency is sampled once on each trial, yielding a quantity denoted $\hat{m}$. Either one or two variable hedge factors are also sampled. The observed variability in starts and stops is determined by trial-to-trial variability in $\hat{m}$ and in the hedge factors. Although Church et al. (1994) refer to $\hat{\sigma}_{\mathrm{m}}^2$, the variance of $\hat{m}$, sometimes as memory variance and sometimes as clock variance, their formulae implicitly assume that trial-to-trial variation in clock speed on probe trials makes a negligible contribution to the observed variation in starts and stops. The observed variance in start times, $\sigma_{\mathrm{t_l}}^2$, is due to the variance, $\hat{\sigma}_{\mathrm{lm}}^2$, of the start criteria, $\hat{l}\,\hat{m}$ which is equal to the variance in the products of $\hat{l}$ and $\hat{m}$. From the rule for the variance of products, we have:

$$\sigma_{\mathrm{t_l}}^2 = \hat{\sigma}_{\mathrm{l}}^2\hat{\sigma}_{\mathrm{m}}^2 + \hat{L}^2\hat{\sigma}_{\mathrm{m}}^2 + \hat{M}^2\hat{\sigma}_{\mathrm{l}}^2 \qquad (1)$$

(Capital letters denote expectations.) The right hand side of Equation (1) may be rewritten in the form, $\hat{\sigma}_{\mathrm{l}}^{\,2}(\hat{\sigma}_{\mathrm{m}}^{\,2} + \hat{M}^2) + \hat{L}^2\hat{\sigma}_{\mathrm{m}}^{\,2}$. The same reasoning leads to the formula relating model parameters to the observed variance in stops:

$$\sigma_{\mathrm{t_u}}^2 = \hat{\sigma}_{\mathrm{u}}^2\hat{\sigma}_{\mathrm{m}}^2 + \hat{U}^2\hat{\sigma}_{\mathrm{m}}^2 + \hat{M}^2\hat{\sigma}_{\mathrm{u}}^2 = \hat{\sigma}_{\mathrm{m}}^2(\hat{\sigma}_{\mathrm{u}}^2 + \hat{U}^2) + \hat{M}^2\hat{\sigma}_{\mathrm{u}}^2 \qquad (2)$$

In the single-hedge model, it is assumed that there is only one hedge factor, $\hat{b}$, with $\hat{l} = 1 - \hat{b}$ and $\hat{u} = 1 + \hat{b}$. Scalar error in the observed mean center, T_{c}, is then an estimate of the scalar memory error, K_{m}, and the observed variance of the center points, σ_{c}^2, is an estimate of the variance of the memory distribution (for derivations, see Gibbon & Church, 1992).

In the double-hedge model, the proportions $\hat{l}$ and $\hat{u}$ are drawn from different distributions. Then, the only source of covariance between start and stop times is the memory sample. This covariance, when divided by the product of the expectations of the start and stop proportions, gives an estimate of the memory variance. The derivation follows from the general formula for covariance:

$$2\mathrm{Cov}(t_1, t_{\mathrm{u}}) = 2\mathrm{Cov}(\hat{l}\hat{m}, \hat{u}\hat{m}) = Var(\hat{l}\hat{m} + \hat{u}\hat{m}) - (\hat{\sigma}_{\mathrm{lm}}^2 + \hat{\sigma}_{\mathrm{um}}^2) \qquad (3)$$

Because $\hat{l}$ and $\hat{u}$ are assumed to vary independently, their variation does not contribute to the covariation in $\hat{l}\hat{m}$ and $\hat{u}\hat{m}$, so we can replace $\hat{l}$ and $\hat{u}$ with their expectations, $\hat{L}$ and $\hat{U}$, which are constants. Examining one by one the two terms on the right hand side of Equation (3), we find that:

$$\mathrm{Var}(\hat{L}\hat{m} + \hat{U}\hat{m}) = \mathrm{Var}[\hat{m}(\hat{L} + \hat{U})] = \hat{\sigma}_{\mathrm{m}}^2(\hat{L} + \hat{U})^2 = \hat{\sigma}_{\mathrm{m}}^2\left(\hat{L}^2 + 2\hat{L}\hat{U} + \hat{U}^2\right)$$

and

$$\left(\hat{\sigma}^2_{\text{Lm}} + \hat{\sigma}^2_{\text{Um}}\right) = \hat{L}^2\hat{\sigma}^2_{\text{m}} + \hat{U}^2\hat{\sigma}^2_{\text{m}} = \hat{\sigma}^2_{\text{m}}\left(\hat{L}^2 + \hat{U}^2\right).$$

The squares of the expectations and the factor 2 cancel out, leaving

$$\text{Cov}(t_{\text{l}}, t_{\text{u}}) = \hat{U}\hat{L}\hat{\sigma}^2_{\text{m}}.$$

The peak curve (whether predicted or observed) is the cumulative start distribution minus the cumulative stop distribution.

Appendix B: Simulation

The SET framework implemented in the simulation is the one portrayed in Figure 1. Specifying different parameter values and different memory-sampling options creates variants of the basic SET model. The program runs a user-specified number of trials, containing a user-specified proportion of probe trials, to generate distributions of starts, stops, centers and spreads. The program compares the cumulative normalized start and stop distributions it generates to the cumulative normalized distributions obtained from individual subjects. The degree of mismatch between the model output and the obtained distributions is measured by the sum of the K-S statistics for the start and stop distributions. The program can vary user-specified model parameters over a range of user-specified steps in a search for the parameter combination that minimizes this measure of the discrepancy between model output and data. The user can then examine the variance and covariance patterns generated by the model using the best-fitting parameter values to determine the extent to which those patterns correspond to the patterns in the data. The program for simulating SET models is available upon request from APK.

The simulation allows one to specify the following parameters and memory-sampling options:

- The expectation ($\hat{D}$) and standard deviation (σ_{d}) of the delay ($\hat{d}$) in starting the timer.
- The standard deviation ($\hat{\sigma}_{\text{s}}$ of the clock speed ($\hat{s}$), which is the slope of the function relating objective duration to subjective duration. The expectation of this slope distribution is 1, because, in this framework, variation in its expectation has no observable effect.
- T^*, the feeding latency
- The expectation ($\hat{K}_{\text{m}}$) of the memory scalar ($\hat{k}_{\text{m}}$), which is the multiplicative factor relating an experienced feeding latency ($\hat{t}_{\text{r}}$) to the record of that same latency when retrieved from memory. In the figure, this scalar distortion is shown on the input side of memory, but conceptually it could be either on the input or the output. It allows for a calibration error in memory, a systematic discrepancy between originally experienced durations and remembered durations. A calibration error exists if the expectation is some value other than 1.

- The standard deviation of the memory scalar ($\hat{k}_m$). The noise in this scalar can create a population of different remembered feeding latencies in the absence of variation in T^* and/or clock speed.
- The memory sampling process on probe trials: The options are:
  i) Pick a random value from the population in memory resulting from trials where food was obtained (feeding trials)
  ii) Take the mean or median of the population in memory
  iii) Take the minimum value in memory
  iv) Take a user-specified fixed value (e.g., T^*) that is independent of the values experienced on feeding trials
- $\hat{\gamma}_m$ = the coefficient of variation of the distribution of memory values (target values) on probe trials. If the memory sampling process yields a fixed or little varying value (as happens if options ii - iv above are specified), then this allows for the introduction of variation in the samples actually used as the targets, variation that is independent of variation in the inputs to memory.
- $\hat{L}$ = the expectation of the start proportion
- $\hat{\sigma}_l$ = the standard deviation of the start proportion
- $\hat{U}$ = the expectation of the stop proportion
- $\hat{\sigma}_u$ = the standard deviation of the stop proportion

Received September 29, 2001.
Revision received March 19, 2002.
Accepted May 2, 2002.