

*CONTEXT EFFECTS IN A TEMPORAL DISCRIMINATION TASK: FURTHER TESTS OF THE SCALAR EXPECTANCY THEORY AND LEARNING-TO-TIME MODELS*JOANA ARANTES^{1,2} AND ARMANDO MACHADO¹¹UNIVERSITY OF MINHO, PORTUGAL²UNIVERSITY OF CANTERBURY, NEW ZEALAND

Pigeons were trained on two temporal bisection tasks, which alternated every two sessions. In the first task, they learned to choose a red key after a 1-s signal and a green key after a 4-s signal; in the second task, they learned to choose a blue key after a 4-s signal and a yellow key after a 16-s signal. Then the pigeons were exposed to a series of test trials in order to contrast two timing models, Learning-to-Time (LeT) and Scalar Expectancy Theory (SET). The models made substantially different predictions particularly for the test trials in which the sample duration ranged from 1 s to 16 s and the choice keys were Green and Blue, the keys associated with the same 4-s samples: LeT predicted that preference for Green should increase with sample duration, a context effect, but SET predicted that preference for Green should not vary with sample duration. The results were consistent with LeT. The present study adds to the literature the finding that the context effect occurs even when the two basic discriminations are never combined in the same session.

Key words: bisection procedure, context effect, temporal discrimination, timing models, key peck, pigeon

When temporal discrimination is studied using a matching-to-sample procedure, a pigeon is shown a light illuminated for 1 s or 4 s, for example, and is then given a choice between two keys, one red and the other green. Choices of red are reinforced following the 1-s sample and choices of green are reinforced following the 4-s sample. A well-trained pigeon may respond accurately on more than 90 percent of the trials (e.g., Stubbs, 1968; also Catania, 1970; Church & Deluty, 1977; Platt & Davis, 1983). In this task, also known as temporal bisection, how does the animal learn to discriminate the sample durations? What processes underlie its performance? Different theories of timing answer these questions differently.

Consider Scalar Expectancy Theory (SET), a model developed by Gibbon and his collaborators (Gibbon, 1977, 1981, 1991; for a

summary see Gallistel, 1990). According to SET, timing performance is regulated by an internal clock with the structure illustrated in the left panel of Figure 1. The clock comprises a pacemaker that generates pulses at a high but variable rate, an accumulator that adds the pulses during the to-be-timed interval, one or more memory stores that save the number of pulses in the accumulator at the end of the interval, and a comparator that compares the number in the accumulator with samples extracted from the memory stores. SET answers the question of learning by saying that at the end of training the animal has formed two memory stores, one containing the distribution of the number of pulses that were in the accumulator at the end of the short, 1-s samples—call it the “Red” store because it is associated with the Red choices—and the other the distribution of the number of pulses that were in the accumulator at the end of the long, 4-s samples, the “Green” store. Each memory store contains a distribution rather than a single number because SET postulates one or more sources of variability within the system (e.g., pacemaker rate may vary across trials; the number in the accumulator may be multiplied by a random variable before it is saved in reference memory). Moreover, the two distributions are scale transforms in the sense that their means and standard deviations are in the same 1-to-4 ratio as the two sample durations.

Research was supported by a grant from the Portuguese Foundation for Science and Technology (FCT) to Armando Machado. Joana Arantes was supported by a Ph.D. scholarship from the Portuguese Foundation for Science and Technology (FCT). We thank John Staddon, Randolph Grace, and Luís Oliveira for helpful comments on earlier versions of the paper, and Alexandra Lima for assistance with data collection.

Correspondence concerning this article should be addressed to Joana Arantes, University of Canterbury, Department of Psychology, Christchurch, New Zealand, or Armando Machado, Instituto de Educação e Psicologia, Universidade do Minho, 4710 Braga, Portugal (e-mail: joana.arantes@canterbury.ac.nz or armandom@iep.uminho.pt).

doi: 10.1901/jeab.2008-90-33

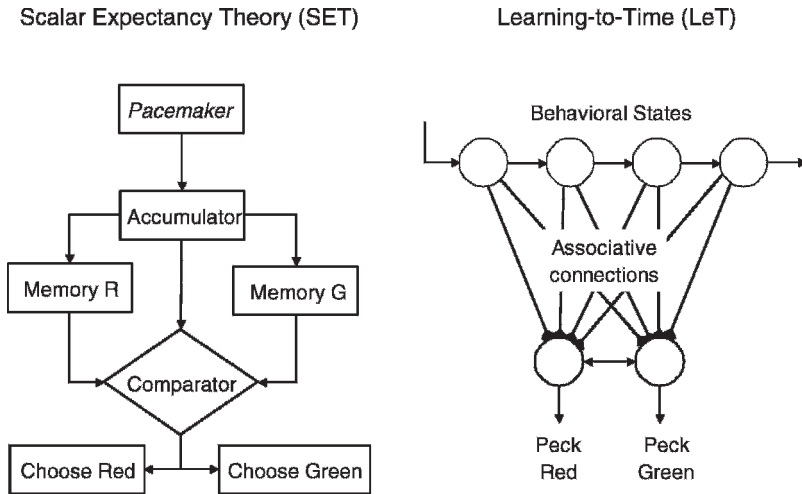


Fig. 1. Structure of the Scalar Expectancy Theory (SET) and the Learning-to-Time (LeT) models for the temporal bisection procedure.

For SET, performance in the temporal bisection task depends not only on the formation of the two memory stores but also on the animal's decision rule. The latter states that at the end of a sample with duration T , the number in the accumulator, X_T , is compared with two samples, one extracted from the Red store, X_R , and the other from the Green store, X_G . The animal will choose the Red or "short" key whenever $X_T/X_R < X_G/X_T$, or equivalently, $X_T < \sqrt{X_R \times X_G}$. In other words, the animal will peck the Red key if the sample duration is below the geometric mean of the two training stimuli (Gibbon, 1981), and it will peck the Green key otherwise.

According to SET the contents of the Red and Green memory stores depend only on the duration of the short and long samples, respectively. For example, the contents of the Green store depend on the duration of the long sample and are not affected by the duration of the short sample. This means that if the animal were trained with a short sample of 2 s, instead of 1 s, the contents of its Green store would remain the same because the 4-s sample did not change. We refer to the assumption that the contents of a memory store depend exclusively on the duration of its associated sample and not on the duration of the alternative sample as "context-independent memories"; this assumption will be the focus of the present study.

Another timing model is Learning-to-Time (LeT), a model developed by Machado (1997; Machado & Cevik, 1998) on the basis of Killeen and Fetterman's (1988) Behavioral theory of Timing (BeT). The right panel of Figure 1 illustrates its three-part structure (Appendix A identifies the relevant equations): a series of behavioral states (the equivalent of the pacemaker in SET), a matrix of associative connections linking the behavioral states to the operant responses (the equivalent of the memory stores in SET), and the operant responses themselves. The behavioral states embody our concepts of elicited, induced, adjunctive, interim, and terminal classes of behavior (see Killeen & Fetterman, 1988; Richelle & Lejeune, 1980; Staddon, 1977; Staddon & Simmelhag, 1971; Timberlake & Lucas, 1985) and according to LeT they underlie the sequential and temporal organization of behavior. In the bisection task, the onset of the sample activates the first state in the series but, as time elapses, the activation of each state flows to the next state. The speed of the activation flow varies with reinforcer rate, perhaps via changes in the animal's arousal level (e.g., Beam, Killeen, Bizo, & Fetterman, 1998). LeT answers the question of learning by saying that during training the strength of the associative link between a behavioral state and an operant response will increase if that response is reinforced and it will decrease if that response is extinguished. In addition, it

will increase also if the competing response is extinguished and decrease also if the competing response is reinforced. In other words, the links between the behavioral states and a response change with reinforcement and extinction of that response *and* with the reinforcement and extinction of the other response.

The model's decision rule states that after the sample the animal's choice will depend on which states are the most active at that moment—itsself a direct function of the sample duration—and on the strengths of the associative links between those states and the two responses. If the sample duration, T , is short, then the most active states at the moment of choice will be the first ones in the series and because during training the choice of Red was reinforced and the choice of Green was extinguished, these initial states will be coupled strongly with the Red response and weakly with the Green response; hence the animal is more likely to choose Red. Conversely, if T is long, then the most active states at the moment of choice will be subsequent states in the series and because during training the choice of Green was reinforced and the choice of Red was extinguished, those states will be coupled strongly with the Green response and weakly with the Red response; hence the animal is more likely to choose Green.

Note that in LeT the steady-state strengths of the links connecting the behavioral states to the operant responses correspond in SET to the distribution of counts in the memory stores. However, in contrast with SET's memory stores, the strengths of the links are context dependent. To understand this point, consider the links between the behavioral states and the Green response. Their final values will depend on the duration of the long *and* short samples. The model's context sensitivity stems from its learning rule—the links with the Green response change not only after 4-s samples (i.e., when Green is reinforced and Red extinguished) but also after 1-s samples (when Green is extinguished and Red is reinforced). Hence, if the duration of the short sample changes, the final values of the links connecting the states to the Green response will also change. To summarize, in the temporal bisection task, the difference between the two models amounts to this: Whereas LeT says that what the animal learns

regarding the 4-s sample and its associated choice, the Green key, will vary with the duration of the other sample, SET says that it will not.

The present study continues a series of studies designed to test the foregoing issue (Arantes, 2008; Machado & Arantes, 2006; Machado & Keen, 1999; Machado & Pata, 2005; and Oliveira & Machado, 2008). All of these studies used a double bisection task to separate the models' predictions. A pigeon learns two simple temporal discriminations. In the first it learns to choose a Red key after 1-s samples and a Green key after 4-s samples. In the second it learns to choose a Blue key after 4-s samples and a Yellow key after 16-s samples. To facilitate communication, we represent the two discriminations by two mappings, $\{S_1, S_4\} \rightarrow \{\text{Red, Green}\}$ and $\{S_4, S_{16}\} \rightarrow \{\text{Blue, Yellow}\}$. In each mapping the first set represents the sample durations and the second set represents the comparison stimuli or choice alternatives; the arrow means that the first and second elements in the choice set are reinforced following the first and second elements in the stimulus set, respectively. The two mappings occur during the same session, the (relatively) short set $\{S_1, S_4\} \rightarrow \{\text{Red, Green}\}$ on half of the trials and the (relatively) long set $\{S_4, S_{16}\} \rightarrow \{\text{Blue, Yellow}\}$ on the other half of the trials.

The critical feature of the double bisection procedure is that Green and Blue choices are reinforced following 4-s samples but their learning contexts differ because in one case (Green) the 4-s samples occur with 1-s samples, whereas in the other case (Blue) the 4-s samples occur with 16-s samples. To determine whether context matters, after the pigeon learns the double discrimination task, the experimenter introduces test trials during which sample duration ranges from 1 s to 16 s and the choice keys are Green and Blue. We represent the test trials with the notation $\{S_1..S_{16}\}:\{\text{Green, Blue}\}$ and note that the sample set includes durations that span the entire range of training durations, from 1 s to 16 s, and that a colon replaces the arrow to show that there are no correct choices.

The two models predict different results on the test trials. These predictions are illustrated in Figure 2. According to SET, the pigeon's preference for, say, Green should not vary with the sample duration. The reason is that the

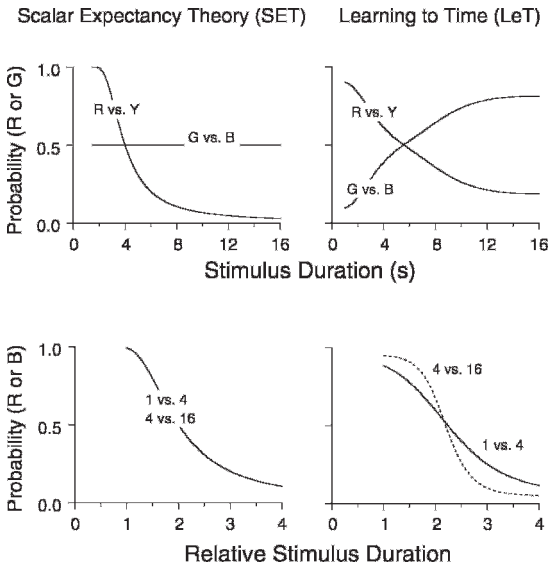


Fig. 2. Predictions of SET and LeT for the double bisection task. It is assumed that the animal has learned two mappings, $\{S_1, S_4\} \rightarrow \{\text{Red}, \text{Green}\}$ and $\{S_4, S_{16}\} \rightarrow \{\text{Blue}, \text{Yellow}\}$. The top panels show the predictions for the two types of stimulus-response generalization trials, $\{S_1..S_{16}\}:\{\text{Green}, \text{Blue}\}$ and $\{S_1..S_{16}\}:\{\text{Red}, \text{Yellow}\}$. The bottom panels show the predictions for the two types of stimulus-generalization trials, $\{S_1..S_4\}:\{\text{Red}, \text{Green}\}$ and $\{S_4..S_{16}\}:\{\text{Blue}, \text{Yellow}\}$.

“Green” and “Blue” memory stores formed during training contain identical distributions of counts. Therefore, during testing, the number in the accumulator at the end of a sample with duration T will be compared with two samples extracted from identical distributions. Hence, preference for Green should not vary with T . For SET, then, the psychometric function relating probability of choosing Green to sample duration should plot as a horizontal line (see the top left panel of Figure 2).

According to LeT, the pigeon’s preference for Green should increase with the sample

duration. To understand the prediction intuitively (for mathematical derivations see Appendix A and also Machado, 1997, and Machado & Pata, 2005), divide the behavioral states into three sets, “Early”, the set of states most active at 1 s, “Middle”, the set of states most active at 4 s, and “Late”, the set of states most active at 16 s (see Table 1). Let the strength of the link between a state and a choice response be represented by W (state, choice). Initially all states are associated equally with the four responses (i.e., $W = 0.5$ for all links), but these associations will change during training. Because of the mapping $\{S_1, S_4\} \rightarrow \{\text{Red}, \text{Green}\}$, the states in the “Early” set become strongly linked with Red and, more importantly, weakly linked with Green (i.e., $W(\text{“Early”}, \text{Green}) \rightarrow 0$). Similarly, because of the mapping $\{S_4, S_{16}\} \rightarrow \{\text{Blue}, \text{Yellow}\}$, the states in the “Late” set become strongly linked with Yellow and, more importantly, weakly linked with Blue (i.e., $W(\text{“Late”}, \text{Blue}) \rightarrow 0$). It follows that, during test trials, after 1-s samples, the “Early” states are the most active, and because their links with Green have decreased to 0 whereas their links with Blue have remained at 0.5, the animal prefers Blue. Conversely, after 16-s samples, the “Late” states are the most active, and because their links with Blue have decreased to 0 whereas their links with Green have remained at 0.5, the animal prefers Green. For LeT, then, the psychometric function relating probability of choosing Green to sample duration should plot as an increasing curve (see the top right panel of Figure 2).

Results from previous studies were consistent with LeT: When the pigeons chose between Green and Blue, the preference for Green increased with sample duration (Machado & Keen, 1999; Machado & Pata, 2005). Machado and Arantes (2006) tested the models’ predictions in a different way. If, at

Table 1
Strength of the links (W) between the behavioral states and the choice responses.

Sample	State	Choice Response			
		Red	Green	Blue	Yellow
1 s	“Early”	$W \rightarrow 1$	$W \rightarrow 0$	$W \approx 0.5$	$W \approx 0.5$
4 s	“Middle”	$W \rightarrow 0$	$W \rightarrow 1$	$W \rightarrow 1$	$W \rightarrow 0$
16 s	“Late”	$W \approx 0.5$	$W \approx 0.5$	$W \rightarrow 0$	$W \rightarrow 1$

Note. “Early”, “Middle” and “Late” represent the states most active after samples 1-s, 4-s, and 16-s long, respectively. Initially, all links equal 0.5. The arrows show the effects of training.

the end of training in the double bisection task, the pigeon is indifferent between Green and Blue, as SET predicts, then it should learn with the same speed the two new mappings $\{S_1, S_{16}\} \rightarrow \{\text{Green}, \text{Blue}\}$ and $\{S_1, S_{16}\} \rightarrow \{\text{Blue}, \text{Green}\}$. That is, there should be no difference between the speed of learning to associate a 1-s sample with Green and a 16-s sample with Blue, and the speed of learning to associate a 1-s sample with Blue and a 16-s sample with Green. However, if, at the end of training with the double bisection task, the pigeon's preference for Green increases with sample duration, as LeT predicts, then the mapping $\{S_1, S_{16}\} \rightarrow \{\text{Blue}, \text{Green}\}$ should be learned more rapidly because it is consistent with that preference; to learn the other mapping, $\{S_1, S_{16}\} \rightarrow \{\text{Green}, \text{Blue}\}$, the pigeon must reverse its initial preference. The results were consistent with LeT: For the consistent mapping, performance was at the steady state by the first session, whereas for the inconsistent mapping, performance was at the steady state only in the third session.

The main purpose of the present experiment was to explore the generality of the context effect observed in the double bisection experiments. In particular, we wanted to determine whether the context effect—preference for Green increases with sample duration—requires the integration of the two basic temporal discriminations in the same session. In the usual procedure, the two discriminations are trained separately, for example, $\{S_1, S_4\} \rightarrow \{\text{Red}, \text{Green}\}$ during the first sessions and $\{S_4, S_{16}\} \rightarrow \{\text{Blue}, \text{Yellow}\}$ during the next sessions, and then they are combined in the same session. Once choice proportions stabilize, the test trials begin. The issue then is whether the main effect persists if the two mappings are never combined in the same session. The present experiment addressed the issue. After the pigeons learned the two basic discriminations on separate sessions we assessed on probe trials how their preference for Green (versus Blue) varied with sample duration. Critically, the two discrimination tasks never occurred in the same session.

When the two discriminations occur during the same session, the subject cannot anticipate the sample durations or the choice keys that will be presented during a trial. But when the discriminations occur in different sessions, after the first trial of each session the subject

can anticipate which sample durations and choice keys will be present for the remainder of the session. It is conceivable that this difference in the predictability of sample durations and choice keys may affect the learning of the two basic discriminations and consequently alter the context effect reported above. If the context effect persists when the discriminations are not combined, then LeT's account is strengthened considerably because the account rests on the direct effects of the two basic discriminations and not on any direct or indirect effects of their integration in the same session. If the context effect disappears when the discriminations are not combined then LeT's account is invalidated.

The implications for SET are somewhat different. If the context effect persists, SET is further invalidated. If the context effect disappears, then SET can claim that when the discriminations are trained separately the memory stores remain context independent, as the model claims. SET could then be revised to try to derive the context effect from the integration of the basic discriminations.

In addition to the question of how preference for Green over Blue changes with sample duration when the discriminations are not combined, the experiment reported below asked two other questions. One was how preference for Red over Yellow changes with sample duration. To answer it, the experiment included test trials in which sample duration ranged from 1 s to 16 s and the choice was between Red and Yellow, that is, $\{S_1..S_{16}\}:\{\text{Red}, \text{Yellow}\}$. Test trials with the Red and Yellow comparisons are a sort of control condition because both SET and LeT predict that preference for Red should decrease monotonically with sample duration (see top panels of Figure 2). However, LeT makes a stronger prediction, also visible in Figure 2 and explained in Appendix A: The psychometric functions for Red and Green are symmetric around the horizontal line $y = .5$, which is equivalent to saying that if instead of plotting the preference for Red one plots the preference for Yellow, then the function will increase with sample duration exactly as the preference for Green increases with sample duration; that is, the two psychometric functions should overlap.

The other question was related to stimulus generalization and it asked whether the

psychometric function obtained by varying the sample from 4 s to 16 s and giving the animal a choice between Blue and Yellow (i.e., $\{S_4..S_{16}\}:\{\text{Blue, Yellow}\}$) is a scale transform of the psychometric function obtained by varying the sample from 1 s to 4 s and giving the animal a choice between Red and Green (i.e., $\{S_1..S_4\}:\{\text{Red, Green}\}$). In other words, do the two functions superimpose when plotted in relative time, the equivalent of Weber's law for timing? As we explain in the Discussion, the stimulus generalization tests do not differentiate the models as clearly as the stimulus-response generalization tests because, depending on the exact assumptions made by LeT, the models may or may not predict the same outcome. Specifically, whereas SET typically predicts superimposition of the psychometric functions (see bottom left panel of Figure 2), LeT may predict either superimposition or a steeper function for the "long" set. This latter case is illustrated in the bottom right panel of Figure 2.

To summarize, in the experiment reported below pigeons learned the two mappings $\{S_1, S_4\} \rightarrow \{\text{Red, Green}\}$ and $\{S_4, S_{16}\} \rightarrow \{\text{Blue, Yellow}\}$. These mappings were learned on separate sessions and throughout the experiment they never occurred during the same session. Afterwards the pigeons were exposed to two types of test trials, one type dealing with stimulus-response generalization and the other with stimulus generalization. The stimulus-response test trials ($\{S_1..S_{16}\}:\{\text{Blue, Green}\}$ and $\{S_1..S_{16}\}:\{\text{Red, Yellow}\}$) examined the context effect issue, and the stimulus generalization test trials (i.e., $\{S_1..S_4\}:\{\text{Red, Green}\}$ and $\{S_4..S_{16}\}:\{\text{Blue, Yellow}\}$) examined the superimposition issue. The results will have implications for our understanding of timing in general and the SET and LeT models in particular.

METHOD

Subjects

Seven experimentally naïve pigeons (*Columba livia*) maintained at 80% of their free-feeding body weights participated in the experiment. Water and grit were continuously available in their individual cages, and a 14:10-hr light/dark cycle (lights on at 7:00 a.m.) was in effect in the pigeon colony.

Apparatus

Two standard three-key operant chambers from Med Associates® were used. The keys were 20 mm in diameter, located 220 mm above the floor, and arranged in a row, 80 mm apart, center to center. The side keys could be illuminated with red, green, blue, or yellow lights, and the center key could be illuminated with white light. The 60 × 70-mm hopper opening was centered on the wall below the center key, providing access to mixed grain when the hopper was raised and illuminated with a 7.5-W white light. Located on the back wall of the chamber was the 7.5-W houselight. Chambers were enclosed by external boxes equipped with fans, which provided ventilation and masked extraneous sounds. Event scheduling and data recording were controlled by a personal computer programmed in the C++ language.

Procedure

The experiment consisted of three phases: baseline training, stimulus generalization testing, and stimulus-response generalization testing.

Baseline training. Sessions were conducted daily at approximately the same time. The pigeons learned to peck the keys using an autoshaping procedure (2–3 sessions). Then they were exposed to one of the two basic tasks, the discrimination between samples of 1 s and 4 s (3 birds) or the discrimination between samples of 4 s and 16 s (4 birds). The assignment of keylight colors to sample durations was counterbalanced with the following restrictions: (a) the color pairs red-green and blue-yellow always occurred together; and (b) green and blue were always associated with the two 4-s samples. However, for clarity, the procedure and the experimental results are described as though all birds had the following assignments $\{S_1, S_4\} \rightarrow \{\text{Red, Green}\}$ or $\{S_4, S_{16}\} \rightarrow \{\text{Blue, Yellow}\}$.

Sessions consisted of 60 trials, 30 with the short sample and 30 with the long sample. Each trial was preceded by a 30-s intertrial interval (ITI) during which the houselight was off. After the ITI, the houselight was turned on and the center key was illuminated with a white light. When the sample duration elapsed (e.g., 4 s), the center key was turned off and the side keys were illuminated with different

colors (e.g., red and green) to signal the choice phase. The position of the keylights was determined pseudorandomly with the constraint that the two colors were presented the same number of times on the left and right keys during the session. A peck at either key turned all keylights and the houselight off. If the choice was correct, the hopper was activated for a specific duration, which varied from 2 to 5 s across pigeons in order to maintain their body weights while minimizing postsession feeding. After the food delivery, the next ITI began. If the pigeon made an incorrect choice, the ITI began immediately and the trial was repeated (correction method). After three consecutive errors the trial was again repeated but only the correct key was illuminated during the choice phase.

Once the birds met the learning criterion on the first discrimination task (at least 80% correct choices for each sample, excluding repeated trials, for 5 consecutive sessions), training began on the second discrimination task. The sessions required to reach criterion ranged from 13 to 27 ($M = 21$) on the first discrimination, and from 13 to 44 sessions ($M = 22$) on the second. After the second discrimination was learned, the stimulus generalization testing phase began.

Stimulus generalization testing. There were two sets of generalization tests, one for each basic discrimination task. To illustrate how testing proceeded, suppose that a bird had learned the two discrimination tasks in the order $\{S_4, S_{16}\} \rightarrow \{\text{Blue, Yellow}\}$ and $\{S_1, S_4\} \rightarrow \{\text{Red, Green}\}$. In this case, the first generalization test session included samples ranging in duration from 1 s to 4 s. More specifically, each session consisted of 64 trials. Of these, 40 were regular training trials, 20 with the short sample (1 s in this case) and 20 with the long sample (4 s); correct responses were reinforced and incorrect responses repeated the trial. The remaining 24 trials were test trials. The sample durations were 1.41 s, 2.00 s, and 2.83 s, presented eight times each. Choice responses on test trials were never reinforced.

After five test sessions, the bird returned to the first discrimination task it had learned, $\{S_4, S_{16}\} \rightarrow \{\text{Blue, Yellow}\}$, and remained on that task until the accuracy criterion was again satisfied. Then it was exposed for five sessions to the second stimulus generalization test set. The

details were the same as during the first set except that the sample durations were four times longer (i.e., 4.0, 5.66, 8.0, 11.31, and 16.0 s). Note that on both generalization test sets, the five sample durations were logarithmically spaced and therefore the middle duration corresponded to the geometric mean of the anchor durations (i.e., 2 s is the geometric mean of 1 s and 4 s, and 8 s is the geometric mean of 4 s and 16 s).

If a bird learned the two basic discriminations in the opposite order, $\{S_1, S_4\} \rightarrow \{\text{Red, Green}\}$ followed by $\{S_4, S_{16}\} \rightarrow \{\text{Blue, Yellow}\}$, then during this phase it was exposed to the generalization set $\{S_4..S_{16}\}:\{\text{Blue, Yellow}\}$ for five sessions, returned to the basic discrimination task $\{S_1, S_4\} \rightarrow \{\text{Red, Green}\}$, and then it was exposed to the generalization set $\{S_1..S_4\}:\{\text{Red, Green}\}$ for five additional sessions. The number of training sessions separating the two generalization tests ranged from 6 to 13 ($M = 8.4$).

Stimulus-response generalization testing. Before the testing began, the pigeons received additional training on the two basic discrimination tasks. Each session consisted of 60 trials, 40 regular trials (correct responses were reinforced and incorrect responses repeated the trial) and 20 extinction trials (correct responses were not reinforced and incorrect responses did not repeat the trial). The two discrimination tasks alternated across sessions until the accuracy criterion was satisfied on both of them (range of 9 to 14 sessions, $M = 10.4$).

Then the birds were exposed to the stimulus-response generalization tests. Each test session included 40 regular training trials and 20 test trials. There were two sets of test sessions, one in which the regular trials were $\{S_1, S_4\} \rightarrow \{\text{Red, Green}\}$ —henceforth called “short” sessions—and another in which they were $\{S_4, S_{16}\} \rightarrow \{\text{Blue, Yellow}\}$, “long” sessions. The “short” and “long” test sessions differed not in terms of the test trials but in terms of the regular, training trials. For reasons presented below, the two types of session alternated every two sessions.

During the test trials, the sample duration was 1 s, 2 s, 4 s, 8 s, or 16 s, and the choice keys were Green and Blue or Red and Yellow. Because for each set of choice keys there were two spatial arrangements, there were 20 distinct test trials (2 choice sets \times 2 spatial arrangements \times 5 sample durations). These

20 trials were distributed over two consecutive sessions. One session included the trials with samples of 1 s, 4 s, and 8 s, and the other session included the trials with samples of 2 s, 4 s, and 16 s. Each sample was presented eight times per session (2 choice sets \times 2 spatial arrangements, repeated twice) except the 4-s sample which was presented only four times per session (2 choice sets \times 2 spatial arrangements). Because the test trials with 4-s samples were presented on both sessions, each of the 20 distinct test trials was presented the same number of times in each block of two test sessions.

To summarize, during the first two sessions tests occurred in the context of $\{S_1, S_4\} \rightarrow \{\text{Red, Green}\}$ training trials (“short” sessions); during the next two sessions the same tests occurred in the context of $\{S_4, S_{16}\} \rightarrow \{\text{Blue, Yellow}\}$ training trials (“long” sessions). The “short” and “long” sessions continued to alternate every two sessions for a total of 16 test sessions. At the end of this phase, each test duration had been presented 64 times, 32 with the $\{\text{Green, Blue}\}$ choice set and 32 with the $\{\text{Red, Yellow}\}$ choice set; within each set of 32, 16 were during “short” sessions and 16 during “long” sessions.

RESULTS

All pigeons learned the two basic discrimination tasks. During the last 5 sessions of the baseline that preceded the tests, the average proportion of correct responses was consistently high both across birds ($M = .92$) and sample durations (range = .86-.95). These results are consistent with previous studies (Machado & Arantes, 2006; Machado & Keen 1999; Machado & Pata, 2005; Oliveira & Machado, 2008).

Stimulus Generalization Testing

The critical issue under examination is whether the two psychometric functions overlap when plotted on a common axis. Figure 3 shows the individual and average data. (Table B1 in Appendix B contains the individual data.) To plot the two sets of results on the same axis, the sample durations on the relatively long test trials were divided by 4. For all birds and both sets of tests, the probability of a “short” response [i.e., $p(\text{“Red”})$ or $p(\text{“Blue”})$], decreased monoton-

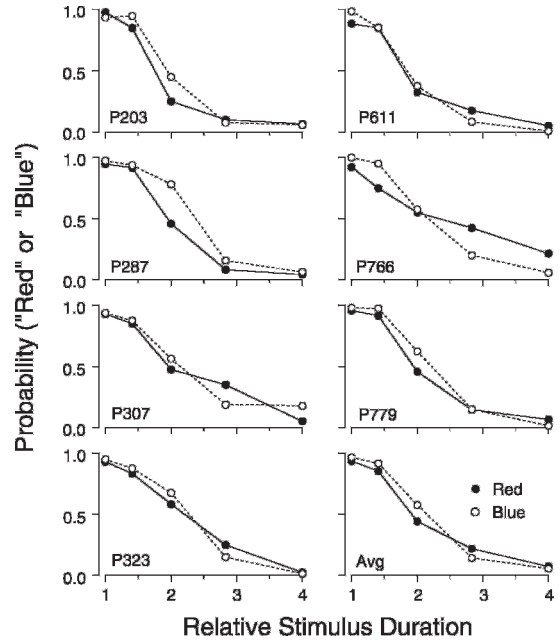


Fig. 3. Probability of “short” responses during the stimulus-generalization test trials. Filled and open circles correspond to the data from $\{S_1..S_4\}:\{\text{Red, Green}\}$ and $\{S_4..S_{16}\}:\{\text{Blue, Yellow}\}$ trials, respectively. The bottom right panel shows the average data.

ically with sample duration. Concerning superimposition, visual inspection shows that the curves for the long range (empty circles) tended to be slightly steeper than the curves for the short range, but with the exception of Pigeons P287 and P766, the differences were small. A repeated-measures analysis of variance (ANOVA) with the five durations as one factor and the two ranges as the other factor yielded a strong effect of stimulus duration, $F(4,24) = 266$, $p < .001$, no effect of stimulus range, $F(1,6) = 4.2$, $p = .09$, and, most important for the present purposes, a significant interaction between the two factors, $F(4,24) = 6.2$, $p < .01$. On average (see bottom right panel), the curve for the long range had higher values at the relative durations of 1.0 and 1.4, but lower values at the relative durations of 2.8 and 4.0.

To compare the present results with those from previous studies (see Discussion), we also fit a two-parameter logistic function to each pigeon’s data. The function has equation $p(\text{“short”}|T) = 1/[1+\exp((T-\mu)/(0.55\sigma))]$, where $p(\text{“short”}|T)$ is the probability of choosing the “short” comparison given a sample T -s long, μ is the mean or point of

subjective equality (PSE), and σ is the standard deviation or slope parameter (a smaller σ means a greater slope at the PSE). The logistic function fitted the data well, accounting for 96% to 100% of the variance in the data (mean $\omega^2 = 98\%$). The average values of the PSEs equaled 2.1 for the short and 2.2 for the long sets. The 95% confidence intervals (CI) included 2.0 s, the geometric mean of the training durations: 95% CI = [1.85, 2.32] for the short set and [1.99, 2.30] for the long set. The difference between the two means was not statistically significant, $t(6) = 0.75$.

As for the slope parameter, σ , visual inspection of the individual data shows that for 3 pigeons (P203, P287, and P779) the two psychometric functions were roughly parallel; for the other 4 pigeons, the function for the longer set was steeper. The average values of σ were 0.86 and 0.61 for the short and long sets, respectively, but the difference fell short of statistical significance, $t(6) = 1.69$, $p = 0.07$. The 95% CIs were [0.42, 1.29] for the short set and [0.48, 0.75] for the long set.

The results from the visual inspection of the data, ANOVA, and t -tests suggest that the psychometric functions for the short and long ranges sometimes differed in slope, with the function for the long range being on average steeper than the function for the short range, but the differences were small.

Stimulus-Response Generalization

The critical issue is whether the context effect (i.e., preference for Green over Blue increases with sample duration) remained even when the two basic discriminations were not integrated in the same session. Figure 4 shows the individual and average results (see Table B2 in Appendix B). For most birds and session types, as the test duration increased, the preference for Green also increased. This result is predicted by LeT but not by SET.

Any reliable differences between the two curves of each panel point to an effect of the type of session ("short" or "long") during which the test trials took place. To determine whether the apparent differences displayed in Figure 4 were reliable, we conducted a repeated-measures ANOVA with the five durations as one factor and the two types of session as the other factor. The results yielded significant effects of stimulus duration, $F(4,24) = 34.5$, $p < .001$, and type of session, $F(1,6) = 12.1$, $p < .001$, but not of their interaction, $F(4,24) = 0.8$.

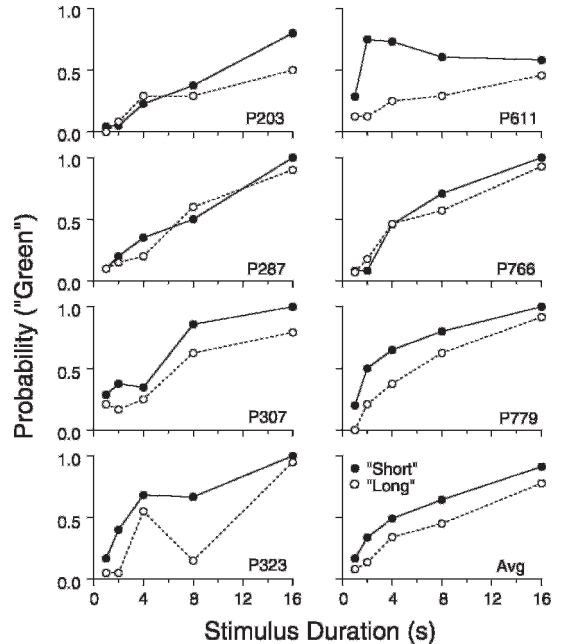


Fig. 4. Probability of choosing Green during the stimulus-response generalization test trials $\{S_1..S_{16}\}:\{Green, Blue\}$. The filled circles correspond to the test results obtained during the "short" sessions (regular trials= $\{S_1, S_4\} \rightarrow \{Red, Green\}$) and the open circles correspond to the test results obtained during the "long" sessions (regular trials= $\{S_4, S_{16}\} \rightarrow \{Blue, Yellow\}$).

We conclude that the main context effect (preference for Green increases with test duration) occurred during both types of sessions, but preference for Green was stronger overall during the "short" sessions.

Figure 5 shows that when the choice was between Red and Yellow, the keys paired with 1-s and 16-s stimuli, respectively, preference for Red decreased with stimulus duration (see Table B3 in Appendix B). This result is consistent with both the SET and LeT models. A repeated-measures ANOVA found significant effects of stimulus duration, $F(4,24) = 93.0$, $p < .001$, and type of session, $F(1,6) = 6.9$, $p = .04$; the interaction approached significance, $F(4,24) = 2.77$, $p = .05$. These results show that, during both types of session, as stimulus duration increased, preference for Red decreased. In addition, the probability of choosing Red was greater during "short" sessions but mainly for the shorter stimulus durations (see bottom right panel).

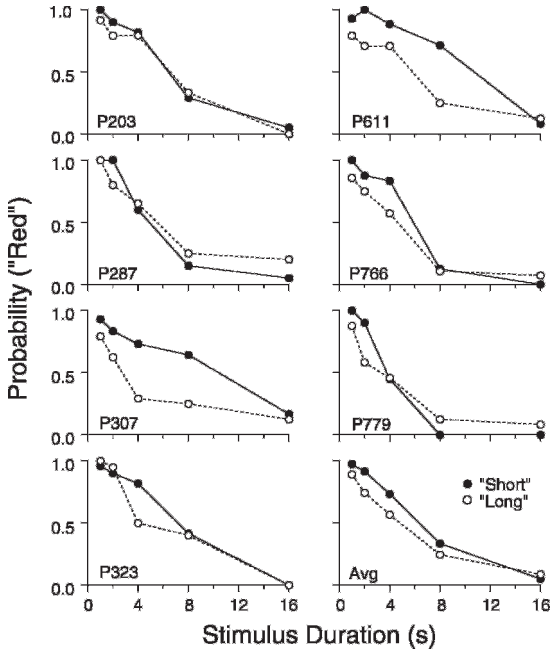


Fig. 5. Probability of choosing Red during the stimulus-response generalization test trials $\{S_1..S_{16}\}:\{\text{Red, Yellow}\}$. The filled circles correspond to the test results obtained during the “short” sessions (regular trials= $\{S_1, S_4\}\rightarrow\{\text{Red, Green}\}$) and the open circles correspond to the test results obtained during the “long” sessions (regular trials= $\{S_4, S_{16}\}\rightarrow\{\text{Blue, Yellow}\}$).

Fitting LeT to the Data

With one exception, the pattern of results described above is qualitatively consistent with LeT. We examine next whether it is also quantitatively consistent with the model. The exception is the effect of type of session, which showed that the preferences for Green and Red in the stimulus-response generalization tests were greater during the “short” than “long” sessions. For reasons described below, LeT has no principled way to account for this effect. Therefore, to fit the model we averaged the psychometric functions obtained in the two types of session. The pattern of results to be fitted comprised four functions, the two stimulus generalization functions $\{S_1..S_4\}:\{\text{Red, Green}\}$ and $\{S_4..S_{16}\}:\{\text{Blue, Yellow}\}$, and the two stimulus-response generalization functions, $\{S_1..S_{16}\}:\{\text{Green, Blue}\}$ and $\{S_1..S_{16}\}:\{\text{Red, Yellow}\}$.

The simplest version of LeT for the bisection tasks uses three free parameters (see Appendix A for mathematical details), λ , the

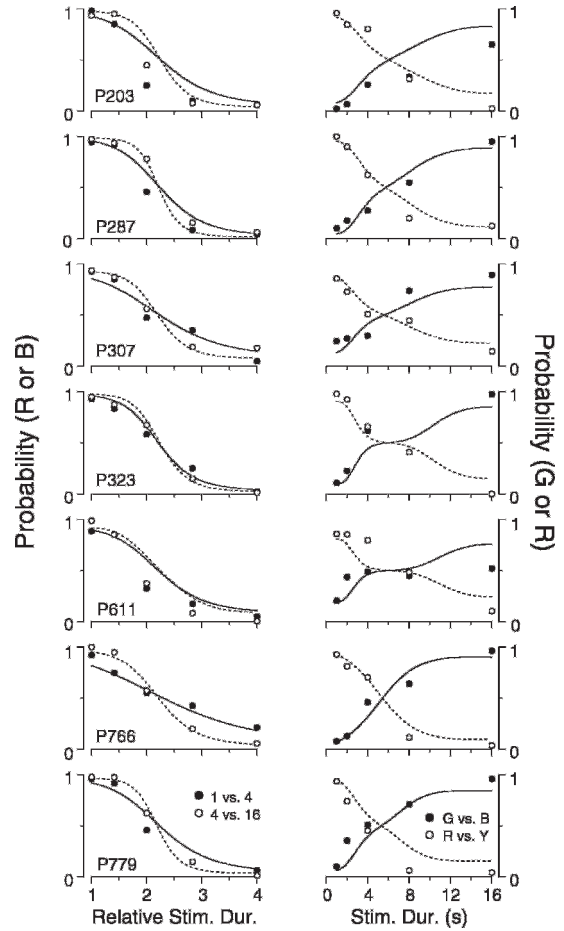


Fig. 6. The left panels show the stimulus-generalization data and the right panels show the stimulus-response generalization data averaged over the two session types. The curves show the prediction made by LeT using the parameters in Table 2.

speed of the activation flow across the behavioral states, γ , the effects of reinforcement and extinction on the strength of the links between the states and the operant responses, and δ , a parameter related to factors such as the discriminability of the comparison stimuli (Machado, 1997). We attempted to fit the four functions with the same set of three parameters.

Figure 6 shows the results for each pigeon. The left panels show the fits to the stimulus generalization data and the right panels show the fits to the stimulus-response generalization data. Table 2 gives the best-fitting parameter values and the overall variance accounted for (see the tables in Appendix B for the raw

Table 2

Best-fitting parameters used to fit the LeT model to the stimulus generalization and stimulus-response generalization data.

Pigeon	λ	γ	δ	ω^2
P203	1.32	0.03	3.94	0.83
P287	1.57	0.08	4.36	0.95
P307	1.63	0.15	2.53	0.91
P323	2.78	0.11	3.69	0.93
P611	2.78	0.08	2.51	0.78
P766	0.65	0.80	3.61	0.96
P779	1.76	1.00	3.35	0.90

data of each pigeon). In general, the model described the major trends in the data well, with ω^2 ranging from 79% to 96%. For the stimulus generalization tests (left panels), LeT predicted either superimposing functions (e.g., P323), or steeper functions for the long stimulus set (e.g., P766). However, LeT overestimated substantially the PSE of some pigeons (e.g., P611 and P203). This overestimation occurred because the model predicts a PSE slightly greater than the geometric mean (see Machado, 1997). If a pigeon displays a PSE at or below the geometric mean, then without additional assumptions LeT cannot account for its data.

For the stimulus-response generalization tests (right panels), LeT fit the data well for 5 pigeons. The exceptions were Pigeons P203 and P611. The poorest overall fit occurred for Pigeon P611 because in addition to a PSE below the geometric mean (left panel), this pigeon also displayed a constant preference for Green after stimulus durations greater than 1 s (see right panel, filled circles). As the solid lines show, LeT always predicts a monotonic increasing preference for Green (the context effect).

To summarize the model fit, Figure 7 shows the average of the data and the average of the individual fits. The top panel shows that, with respect to stimulus generalization, the only significant deviation between model and data seems to be the slight overestimation of the PSE, particularly for the short set (solid line and filled circles). The middle panel shows that, with respect to stimulus-response generalization, there were no large deviations between data and model. The theoretical curves on the two panels accounted for 96% of the variance in the data.

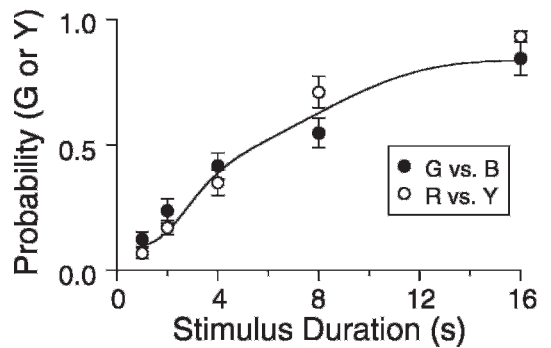
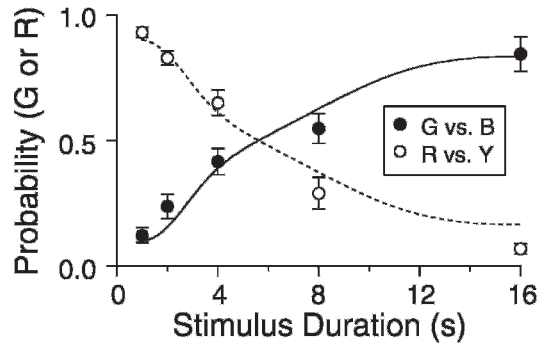
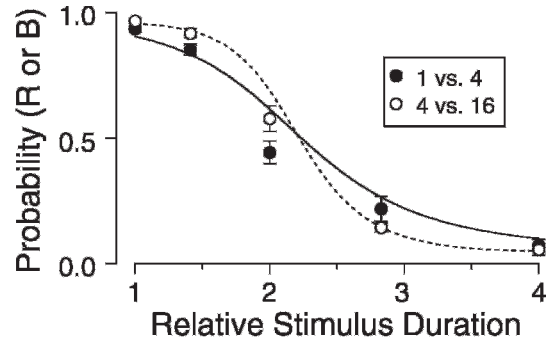


Fig. 7. The top and middle panels show the average of the data (± 1 SEM) and the average of curves predicted by LeT for the stimulus-generalization tests (top) and stimulus-response generalization tests (middle). The bottom panel shows the same stimulus-response data but instead of plotting the preference for Red it plots the preference for Yellow. LeT predicts that the two sets of symbols should overlap.

The bottom panel shows the stimulus-response generalization data in a different way. Instead of plotting the preference for Red over Yellow it plots the preference for Yellow over Red. LeT predicts that the preference for Yellow should increase with sample duration exactly as the preference for Green does (i.e.,

the two functions should overlap). Visually, the data seemed consistent with the prediction, but a two-way repeated measures ANOVA with duration and key color (Green or Yellow) as factors yielded a significant interaction, $F(4,32) = 3.3$, $p = .03$. As expected, there was also a strong effect of duration, $F(4,24) = 102$, $p < .001$, and no effect of keycolor, $F(1,6) = .23$. The interaction occurred because preference for Yellow was weaker at short durations but stronger at long durations than preference for Green.

We conclude that using only three free parameters LeT accounted well for the quantitative results of the stimulus–response generalization data, including the context effect, but its stronger prediction concerning symmetry of the two stimulus–response psychometric functions does not seem correct (see also Machado & Pata, 2005). Moreover, LeT did not account as well for the stimulus–generalization data because it tended to overestimate the PSE.

DISCUSSION

For a variety of tasks, the SET and LeT models make similar predictions. A case in point is the simple temporal bisection: Both models predict monotone psychometric functions with the point of subjective equality at (SET) or close to (LeT) the geometric mean of the training durations and superimposition of functions obtained with pairs of durations in the same ratio (e.g., 1 s vs. 4 s and 4 s vs. 16 s). To disentangle the models, we have developed the double bisection procedure.

In this procedure, pigeons learn two temporal discriminations, $\{S_1, S_4\} \rightarrow \{\text{Red}, \text{Green}\}$ and $\{S_4, S_{16}\} \rightarrow \{\text{Blue}, \text{Yellow}\}$ with the critical feature that the Green and Blue alternatives are associated with the same 4-s duration but in different contexts. The context for Green is the “1 s \rightarrow Red key” mapping, whereas the context for the Blue is the “16 s \rightarrow Yellow key” mapping. One could say also that the 4-s duration associated with Green is relatively long, whereas the 4-s duration associated with Blue is relatively short. Two questions may be asked, one pertaining to stimulus generalization and the other to stimulus–response generalization. First, will the psychometric functions $\{S_1..S_4\}:\{\text{Red}, \text{Green}\}$ and $\{S_4..S_{16}\}:\{\text{Blue}, \text{Yellow}\}$ superimpose when plotted

in relative time? And second, what is the shape of the psychometric functions $\{S_1..S_{16}\}:\{\text{Green}, \text{Blue}\}$ and $\{S_1..S_{16}\}:\{\text{Red}, \text{Yellow}\}$? We address each question in turn and then conclude with some implications of our findings for timing models.

Stimulus Generalization

When the two basic discriminations occur in the same session and no cue signals the forthcoming trials (as in the standard double bisection procedure), SET predicts superimposition of the two psychometric functions but LeT predicts that the function for the longer set, $\{S_4..S_{16}\}:\{\text{Blue}, \text{Yellow}\}$, should be steeper than the function for the shorter set, $\{S_1..S_4\}:\{\text{Red}, \text{Green}\}$. LeT’s prediction stems from its assumption that the activation flow across the behavioral states follows a Poisson process and, consequently, the ratio of the standard deviation to the mean (i.e., the coefficient of variation) decreases with the interval to be timed — the relative accuracy of a “Poisson clock” increases with sample duration (see Appendix A).

The first study with the double bisection procedure (Machado & Keen, 1999) used 8 pigeons and obtained the average data shown in the top panel of Figure 8. The repeated-measures ANOVA conducted with the individual data yielded a strong effect of duration ($p < .001$) and no effect of range ($p = .17$); the effect of the interaction approached significance ($p = .07$). Machado and Keen also fit the logistic function to each individual data set. As the top panel shows, the average of the fitted curves was steeper for the long than the short sets. The t tests revealed that the differences in slope were statistically significant, but the differences in PSE were not. Taken together, these results suggest that the two psychometric functions may indeed differ in slope, as LeT predicts, but these differences were not shown by all birds and even when they were shown they tended to be relatively small.

In the present study the standard double bisection procedure was changed such that the two discriminations occurred always in separate sessions. This procedural difference induces another one, a difference in local reinforcement rates between the two sessions. When the sample durations are 1 s and 4 s the local reinforcer rate is four times greater than

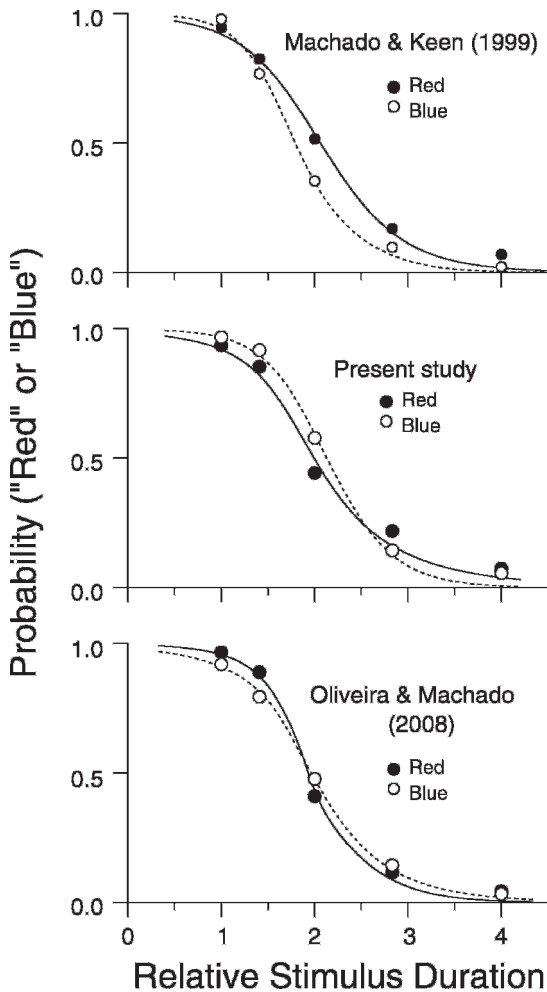


Fig. 8. Average (scaled) data and logistic curves from the stimulus-generalization test trials of three studies. Filled and open circles correspond to the "short" and "long" discrimination sets, respectively. *Top*. In Machado and Keen's (1999) study, the two discrimination sets were combined in the same session. *Middle*. In the present study, the two discrimination sets were not combined in the same session. *Bottom*. In Oliveira and Machado's (2008) study, the two discrimination sets were combined in the same session but each was signaled by a distinctive cue. In the top and middle panels the discrimination sets were $\{S_1, S_4\} \rightarrow \{\text{Red}, \text{Green}\}$ and $\{S_4, S_{16}\} \rightarrow \{\text{Blue}, \text{Yellow}\}$; in the bottom panel they were $\{S_{1.5}, S_6\} \rightarrow \{\text{Red}, \text{Green}\}$ and $\{S_6, S_{24}\} \rightarrow \{\text{Blue}, \text{Yellow}\}$.

when they are 4 s and 16 s. For SET, this difference has no effect and superimposition is still predicted. Although the speed of the pacemaker can increase during "short" sessions, as long as the coefficient of variation of pacemaker speed remains constant, superim-

position holds (see Oliveira & Machado, 2008). The case is different for LeT and slightly more complex. If we assume that parameters λ and γ are strictly proportional to local reinforcer rate, then LeT predicts superimposition (Machado, 1997; also Killeen & Fetterman, 1988); if the assumption is not made and, in particular, if λ increases with the local reinforcer rate but not proportionately, as subsequent empirical research has suggested (see Bizo & White, 1994, 1995a, 1995b; Fetterman & Killeen, 1991; Killeen & Fetterman, 1988; Morgan, Killeen, & Fetterman, 1993), then LeT predicts a steeper function for the long range. In summary, whereas SET predicts superimposition, LeT predicts either superimposition or steeper functions for the longer discrimination set.

The results of the generalization tests were similar to those of the previous study. The repeated-measures ANOVA yielded a significant interaction, which corroborated the visual impression that the functions differed in slope on the average, but the *t*-test on the slope parameters of the logistic fits (see the curves in the middle panel of Figure 8) only approached significance.

The bottom panel shows the average data and functions obtained in a third study (Oliveira & Machado, 2008) with 9 pigeons. In this case the two discriminations were integrated in the same session, but each was signaled by a distinctive cue. The sample from the short set consisted of the center key illuminated with a horizontal bar and the samples from the long set consisted of the center key illuminated with a vertical bar. Thus, similar to Machado and Keen (1999), the two discriminations were presented in the same session but, similar to the present study, on each trial the relative duration of the sample and the choice keys were predictable. The repeated-measures ANOVA results yielded a significant effect of duration ($p < .001$), but not of range ($p = .15$) or their interaction ($p = .22$). The *t* tests on the parameters provided by the logistic fits (see the curves in the bottom panel of Figure 8) revealed no significant differences in either slope or PSE.

We conclude that when the two discrimination tasks, and therefore the two local reinforcer rates that these tasks induce, are not signaled as in the original study, or when they are signaled by global cues as in the present

study, the psychometric function for the long set tends to be slightly steeper on the average than the psychometric function for the short set. When they are signaled by local cues (vertical or horizontal bars) as in Oliveira and Machado's (2008) study, the functions tend to superimpose. These tentative conclusions should not hide the fact that there were substantial individual differences in all these studies, differences whose sources remain to be investigated. Because of these differences, the results from the stimulus generalization tests are not as informative concerning model comparisons as the results from the stimulus-response generalization tests to which we now turn.

Stimulus-Response Generalization

The critical finding reported in previous studies with the standard double bisection procedure was that preference for Green increases with stimulus duration. This context effect is consistent with LeT but not with SET. However, in previous studies, the two discriminations were integrated in the same session before the Green-Blue tests were carried out and therefore it was not known whether such integration was necessary to obtain the effect. When two discriminations are integrated in the same session, the animal cannot anticipate on each trial whether the sample duration will be from the short or long sets, or the comparison keys. In contrast, when the discriminations occur in different sessions, after the first trial of each session the animal can anticipate both the sample duration range and the comparison keys. To determine whether this procedural difference changes the context effect, in the present study the two discriminations were never mixed in the same session either before or during the Green-Blue test trials.

The results showed that the context effect was maintained: Preference for Green increased with sample duration. The result was robust, for all pigeons showed it. Moreover, it was similar in magnitude to the results obtained without integration. Figure 9 compares the average data obtained in the present study with the average data obtained in two studies without integration. As the top panel shows, with or without integration, preference for Green increased from about 0.1 to about 0.7 as the sample duration increased from 1 s

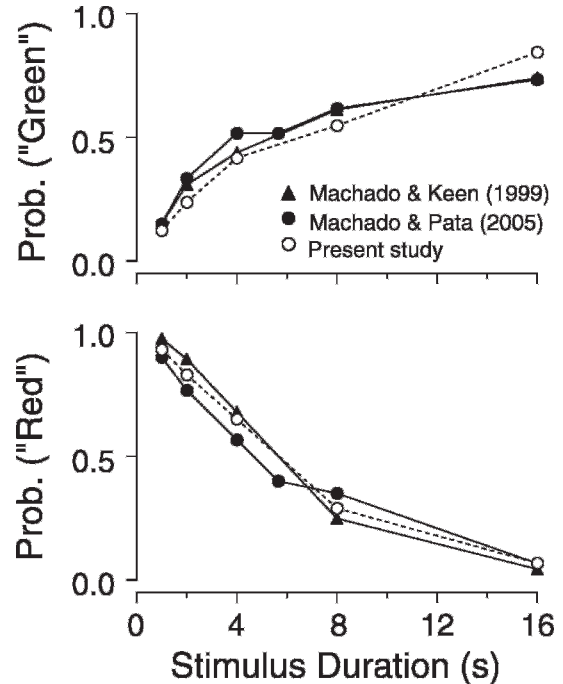


Fig. 9. Average data from the stimulus-response generalization test trials of three studies: Machado and Keen (1999), Machado and Pata (2005), and the present study. The top and bottom panels show the data from the $\{S_1..S_{16}\};\{\text{Green, Blue}\}$ and $\{S_1..S_{16}\};\{\text{Red, Yellow}\}$ trials, respectively. In the former two studies, the discrimination sets were combined in the same session.

to 16 s. The generality of the context effect is consistent with LeT but casts serious doubt on SET's account of temporal discrimination in bisection tasks.

In the other test set, the comparisons were the stimuli associated with the most extreme durations, Red and Yellow. As both models predicted, preference for Red decreased with sample duration. The result also was robust across pigeons and, as the bottom panel of Figure 9 shows, its magnitude was similar to that observed in studies without integration.

The effects of stimulus duration on preference for Green over Blue and for Red over Yellow were consistent with LeT. However, the effects of session type, "short" or "long", were not predicted by either LeT or SET. When presented with a choice between Green and Blue following a sample T -s long, preference for Green was more pronounced during "short" than "long" sessions. A similar result was obtained in Oliveira and Machado's

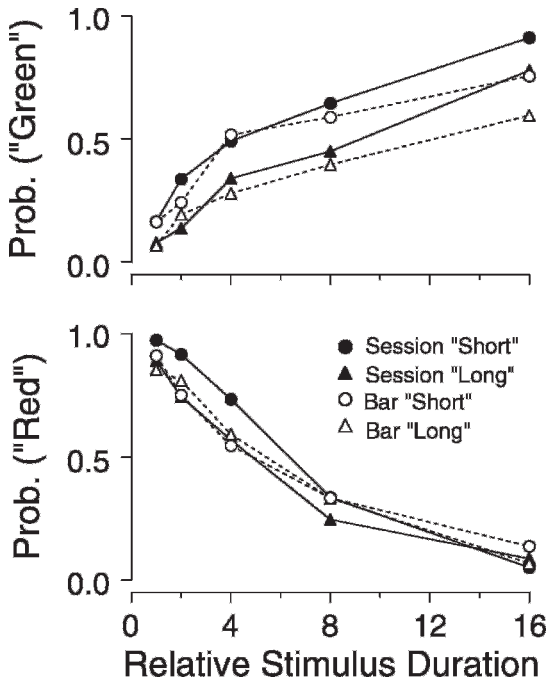


Fig. 10. Average data from the stimulus-response generalization test trials of two studies. In the present study (filled symbols) the “short” and “long” discrimination sets were presented in different sessions. The sets were $\{S_1, S_{16}\} \rightarrow \{\text{Red}, \text{Green}\}$ and $\{S_1, S_{16}\} \rightarrow \{\text{Blue}, \text{Yellow}\}$. In Oliveira and Machado’s (2008) study (open symbols), the two sets were combined in the same session but each was signaled by a distinctive cue (horizontal or vertical bar). The sets were $\{S_{1.5..S_{24}}\} \rightarrow \{\text{Red}, \text{Yellow}\}$ and $\{S_{1.5..S_{24}}\} \rightarrow \{\text{Green}, \text{Blue}\}$. All stimulus durations were divided by 1.5 to fit in the 1s-16s scale.

(2008) study in which horizontal and vertical bars cued the forthcoming trial. The top panel of Figure 10 displays the average results from both studies.

When interpreting the sample cue effect displayed in Figure 10, Oliveira and Machado (2008) considered three hypotheses, one based on changes in pacemaker speed (SET) or activation flow (LeT), another based on changes in mediating behaviors, and yet another based on associative, biasing factors. They rejected the first two hypotheses and tentatively entertained the third, which stated the following: When a sample cue signals, say, a “short” trial, the pigeon is slightly biased towards Green because, during training, Green (but not Blue) was reinforced in the presence of that cue. Conversely, when the sample cue signals a “long” trial, the pigeon is

slightly biased towards Blue. Thus, according to the hypothesis, the pigeon’s choice during test trials depends on two variables, the sample duration (causing an increasing preference for Green) and the sample cue (adding to that preference an approximately constant amount whenever the sample signals a “short” trial).

The session effects obtained in the present study are consistent with this associative account provided we assume that “short” and “long” sessions are functionally equivalent to the horizontal and vertical bars. In this view, it does not matter whether a cue is local (bar) or global (session type), but rather whether in its presence choices of one or the other comparison stimuli were reinforced. Because during “short” sessions choices of Green (but not Blue) were reinforced, whereas the opposite happened during “long” sessions, preference for Green is higher in “short” than in “long” sessions. Similarly, because during “short” sessions choices of Red (but not Yellow) were reinforced, whereas the opposite happened during “long” sessions, preference for Red is higher in “short” than in “long” sessions. Both session effects were observed.

Further studies should examine this hypothesis more directly and, in addition, clarify the conditions under which the biasing effect is obtained. To illustrate, presumably the discrimination between Red and Yellow, the colors associated with the most extreme durations, is easier than the discrimination between Green and Blue, the colors associated with the same duration. If the biasing effect is stronger when the discrimination is more difficult, then the hypothesis could account for the different amounts of vertical shifts displayed in Figure 10.

Implications for Timing Models

The studies with the double bisection procedure have different implications for SET and LeT. For SET, the major challenge is to make the memory stores context dependent, for otherwise the model will not be able to account for the context effect. The memory store that represents one particular sample must be sensitive to the duration of the other sample. But Machado (1997) also showed that SET has logical problems related to how the memory structures postulated by the model work in some temporal tasks (see Gallistel,

2007, and Machado & Silva, 2007a, 2007b, for a discussion of this issue). These logical problems with SET suggest that the numbers in the accumulator at the end of the same sample should not be saved in a single store. The logical and empirical problems mentioned above (see also Guilhardi, McInnis, Church, & Machado, 2007; Machado & Guilhardi, 2000; Staddon, 1999) seriously question the overall memory structure postulated by SET.

For LeT, two major challenges are apparent, how to increase its flexibility to deal with stimulus generalization functions with PSEs at or below the geometric mean, and how to account for the type of session effect displayed in Figures 4, 5, and 10. The former problem questions the Poisson dynamics of the behavioral states, and the latter problem invites us to address the issue of how timing models can include associative effects (see Church, 1997; Kirkpatrick & Church, 1998). Perhaps a hybrid model that preserves LeT's assumptions concerning the associative links but replaces the Poisson dynamics of the states with SET's scalar dynamics will be able to circumvent the shortcomings of both models.

In conclusion, the present study showed that the context effect obtained in the standard double bisection task does not depend on the integration of the two discrimination tasks in the same session. Together with the other studies it suggests that the effect has considerable generality. The effect is inconsistent with SET, the dominant model in the field, because SET assumes that the memory for a particular duration is independent of the alternative durations with which it paired during training. In contrast, the context effect is consistent with LeT. However, neither LeT nor SET can account for the full pattern of results (stimulus generalization and stimulus-response generalization) obtained with the double bisection procedure. A more inclusive model is necessary.

REFERENCES

- Arantes, J. (2008). Comparison of Scalar Expectancy Theory (SET) and the Learning-to-Time (LeT) model in a successive temporal bisection task. *Behavioural Processes*, *78*, 269–278.
- Beam, J. J., Killeen, P. R., Bizo, L. A., & Fetterman, J. G. (1998). How reinforcement context affects temporal production and categorization. *Animal Learning & Behavior*, *26*, 388–396.
- Bizo, L. A., & White, K. G. (1994). The behavioral theory of timing: Reinforcer rate determines pacemaker rate. *Journal of the Experimental Analysis of Behavior*, *61*, 19–33.
- Bizo, L. A., & White, K. G. (1995a). Biasing the pacemaker in the behavioral theory of timing. *Journal of the Experimental Analysis of Behavior*, *64*, 225–235.
- Bizo, L. A., & White, K. G. (1995b). Reinforcement context and pacemaker rate in the behavioral theory of timing. *Animal Learning & Behavior*, *23*, 376–382.
- Catania, A. C. (1970). Reinforcement schedules and the psychophysical judgments: a study of some temporal properties of behavior. In W. N. Schoenfeld (Ed.), *The theory of reinforcement schedules* (pp. 1–42). New York: Appleton-Century-Crofts.
- Church, R. (1997). Quantitative models of animal learning and cognition. *Journal of Experimental Psychology: Animal Behavior Processes*, *23*, 379–389.
- Church, R. M., & Deluty, M. Z. (1977). Bisection of temporal intervals. *Journal of Experimental Psychology: Animal Behavior Processes*, *3*, 216–228.
- Fetterman, J. G., & Killeen, P. R. (1991). Adjusting the pacemaker. *Learning and Motivation*, *22*, 226–252.
- Gallistel, C. R. (1990). *The organization of learning*. Cambridge, MA: Bradford Books/MIT Press.
- Gallistel, C. R. (2007). Flawed foundations of associationism: Comment on Machado and Silva (2007). *American Psychologist*, *62*, 682–685.
- Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, *84*, 279–325.
- Gibbon, J. (1981). On the form and location of the psychometric bisection function for time. *Journal of Mathematical Psychology*, *24*, 58–87.
- Gibbon, J. (1991). Origins of scalar timing theory. *Learning and Motivation*, *22*, 3–38.
- Guilhardi, P., MacInnis, M. L. M., Church, R. M., & Machado, A. (2007). Shifts in the psychophysical function in rats. *Behavioural Processes*, *75*, 167–175.
- Killeen, P., & Fetterman, J. G. (1988). A behavioral theory of timing. *Psychological Review*, *95*, 274–285.
- Kirkpatrick, K., & Church, R. M. (1998). Are separate theories of conditioning and timing necessary? *Behavioural Processes*, *44*, 163–182.
- Machado, A. (1997). Learning the temporal dynamics of behavior. *Psychological Review*, *104*, 241–265.
- Machado, A., & Arantes, J. (2006). Further tests of the Scalar Expectancy Theory (SET) and the Learning-to-Time (LeT) model in a temporal bisection task. *Behavioural Processes*, *72*, 195–206.
- Machado, A., & Cevik, M. (1998). Acquisition and extinction under periodic reinforcement. *Behavioural Processes*, *44*, 237–262.
- Machado, A., & Guilhardi, P. (2000). Shifts in the psychometric function and their implications for models of timing. *Journal of the Experimental Analysis of Behavior*, *74*, 25–54.
- Machado, A., & Keen, R. (1999). Learning to Time (LET) or Scalar Expectancy Theory (SET)? A critical test of two models of timing. *Psychological Science*, *10*, 285–290.
- Machado, A., & Pata, P. (2005). Testing the Scalar Expectancy Theory (SET) and the Learning to Time model (LeT) in a double bisection task. *Learning and Behavior*, *33*, 111–122.

- Machado, A., & Silva, F. (2007a). Toward a richer view of the scientific method: the role of conceptual analysis. *American Psychologist*, *62*, 671–681.
- Machado, A., & Silva, F. (2007b). On the clarification of concepts: A reply to Gallistel and Lau. *American Psychologist*, *62*, 689–691.
- Morgan, L., Killeen, P. R., & Fetterman, J. G. (1993). Changing rates of reinforcement perturbs the flow of time. *Behavioural Processes*, *30*, 259–272.
- Oliveira, L., & Machado, A. (2008). The effect of sample duration and cue on a double temporal discrimination. *Learning and Motivation*, *39*, 71–94.
- Platt, J. R., & Davis, E. R. (1983). Bisection of temporal intervals by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *9*, 160–170.
- Richelle, M., & Lejeune, H. (1980). *Time in animal behavior*. Oxford, England: Pergamon Press.
- Staddon, J. E. R. (1977). Schedule-induced behavior. In W. K. Honig, & J. E. R. Staddon (Eds.), *Handbook of operant behavior* (pp. 125–152). Englewood Cliffs, NJ: Prentice-Hall.
- Staddon, J. E. R. (1999). Time and memory: Towards a pacemaker-free theory of interval timing. *Journal of the Experimental Analysis of Behavior*, *71*, 215–251.
- Staddon, J. E. R., & Simmelhag, V. (1971). Superstition experiment – a reexamination of its implications for the principles of adaptive behavior. *Psychological Review*, *78*, 3–43.
- Stubbs, D. A. (1968). The discrimination of stimulus duration by pigeons. *Journal of the Experimental Analysis of Behavior*, *11*, 223–238.
- Timberlake, W., & Lucas, G. A. (1985). The basis of superstitious behaviour — change contingency, stimulus substitution, or appetitive behaviour. *Journal of the Experimental Analysis of Behavior*, *44*, 279–299.

Received: July 12, 2007

Final Acceptance: March 20, 2008

APPENDIX A

DERIVATIONS FROM LeT

The derivation of LeT’s equations for the temporal bisection tasks are detailed in Machado (1997) and Machado and Pata (2005). Here we summarize the major results.

Simple bisection. Assume the short (S) and long (L) samples occur with probability .5, the correct choices following them are the Red (R) and Green (G) keys, respectively, and training continues for a total of m trials with each sample.

LeT has three components, the behavioral states, their associative links with the operant responses, and the operant responses themselves (see Figure 1). To obtain its predictions, we need to know the activation of state n after a sample of T seconds, $X(T,n)$, and the strength on trial m of the associative links between state n and the R and G responses, $WR(n,m)$ and $WG(n,m)$, respectively.

In LeT, $X(T,n)$ follows the Poisson distribution, a minimal, parsimonious assumption with equation

$$X(T,n) = \frac{\exp(-\lambda T)(\lambda T)^n}{n!} \quad (1)$$

where $\lambda > 0$ is the activation speed across the states. The Poisson distribution has mean and standard deviation equal to λT and $\sqrt{\lambda T}$, respectively. Hence, the coefficient of variation equals $1/\sqrt{\lambda T}$, which decreases with T .

During training, the changes in $WR(n,m)$ and $WG(n,m)$ are always symmetric—when $WR(n,m)$ increases because response R was reinforced or decreases because it was extinguished, $WG(n,m)$ changes by the same amount but in the opposite direction. The same is true when response G occurs and $WG(n,m)$ changes— $WR(n,m)$ also changes by the same amount but in the opposite direction. To simplify, we assume there is no initial bias and set $WR(n,0) = WG(n,0) = 0.5$. From the symmetry of change and the initial values, it follows that $WG(n,m) = 1 - WR(n,m)$.

The expected values of $WR(n,m)$ and $WG(n,m)$ at the end of trial m , designated by $E[WR(n,m)]$ and $E[WG(n,m)]$, equal

$$E[WR(n,m)] = \frac{X(S,n)}{X(S,n) + X(L,n)} - \left(1 - \gamma \frac{X(S,n) + X(L,n)}{2}\right)^m \times \left(\frac{X(S,n)}{X(S,n) + X(L,n)} - 0.5\right) \quad (2)$$

and

$$E[WG(n,m)] = 1 - E[WR(n,m)] \quad (3)$$

From the expressions for $X(T,n)$, $E[WR(n,m)]$, and $E[WG(n,m)]$ one gets the strengths of the R and G responses after a signal of duration T , $RR(T,m)$ and $RG(T,m)$, respectively,

$$\begin{aligned}
 RR(T, m) &= \sum_n X(T, n) E[WR(n, m)] \\
 RG(T, m) &= \sum_n X(T, n) E[WG(n, m)]
 \end{aligned}
 \tag{4}$$

Because of Equation (3) and the fact that $\sum_n X(T, n) = 1$, $RG(T, m) = 1 - RR(T, m)$, which leads to the symmetry prediction mentioned in the text when discussing the bottom panel of Figure 7.

Finally, the probability of choosing the R key on trial m at the end of a sample T -s long equals

$$\begin{aligned}
 P(R|m, T) &= \frac{\exp[\delta \times RR(T, m)]}{\exp[\delta \times RR(T, m)] + \exp[\delta \times RG(T, m)]} \\
 &= \frac{1}{1 + \exp\{\delta[RG(T, m) - RR(T, m)]\}}
 \end{aligned}
 \tag{5}$$

where $\delta > 0$ is a sensitivity parameter.

In summary, Equation (1) describes the activation of the behavioral states and uses parameter λ , Equations (2) and (3) describe the effects of learning and use parameter γ , and Equation (5) yields choice probability on the basis of the two response strengths given by Equation (4) and uses parameter δ .

Double bisection. Assume that R and G are associated with 1 s and 4 s samples, as before, and Blue (B) and Yellow (Y) are associated with 4 s and 16 s signals, respectively. At the end of m trials with each sample, the state of the animal is characterized by the two pairs of vectors of associative links, $WR(n, m)$ and $WG(n, m)$ on the one hand, and $WB(n, m)$ and $WY(n, m)$ on the other hand. As illustrated above, we obtain $WR(n, m)$ using Equation (2) with $S = 1$ s and $L = 4$ s, and then obtain $WG(n, m)$ using Equation (3). Similarly, we obtain $WB(n, m)$ using Equation (2) with $S = 4$ s and $L = 16$ s and then obtain $WY(n, m)$ using Equation (3). During testing, when two novel keys are presented, say G and B, at the end of a sample T -s long, we obtain the probability of choosing G by a) inserting the values of $E[WG(n, m)]$ and $E[WB(n, m)]$ in Equation (4) to obtain $RG(T, m)$ and $RB(T, m)$, and then inserting these values into Equation (5). The curves in Figures 6 and 7 were obtained using the parameter values shown in Table 2 and an m value of 2000 trials.

APPENDIX B
DATA FROM INDIVIDUAL SUBJECTS

Table B1

Proportion of choices of the “short” key (Red or Blue) for each sample duration during the two sets of stimulus generalization tests.

Set	Sample (s)	Pigeon						
		P203	P287	P307	P323	P611	P766	P779
Short	1.00	0.98	0.95	0.93	0.93	0.89	0.92	0.96
	1.41	0.85	0.92	0.85	0.83	0.85	0.75	0.92
	2.00	0.25	0.46	0.48	0.58	0.32	0.55	0.46
	2.83	0.10	0.08	0.35	0.25	0.18	0.42	0.14
Long	4.00	0.06	0.04	0.05	0.02	0.05	0.21	0.06
	4.00	0.94	0.97	0.94	0.95	0.99	1.00	0.98
	5.66	0.95	0.94	0.88	0.88	0.85	0.95	0.98
	8.00	0.45	0.78	0.56	0.68	0.38	0.58	0.62
	11.31	0.08	0.16	0.19	0.15	0.08	0.20	0.15
	16.00	0.06	0.06	0.18	0.01	0.00	0.06	0.01

Note. During the Short set, the choice was between the Red and Green keys; during the Long set the choice was between the Blue and Yellow keys.

Table B2

Proportion of choices of the Green key for each sample duration during the two sets of stimulus–response generalization tests (“Short” and “Long” sessions).

Sessions	Sample (s)	Pigeon						
		P203	P287	P307	P323	P611	P766	P779
Short	1.00	0.04	0.10	0.28	0.17	0.28	0.08	0.20
	2.00	0.05	0.20	0.38	0.40	0.75	0.08	0.50
	4.00	0.23	0.35	0.35	0.68	0.73	0.46	0.65
	8.00	0.38	0.50	0.86	0.67	0.61	0.71	0.80
Long	16.00	0.80	1.00	1.00	1.00	0.58	1.00	1.00
	1.00	0.00	0.10	0.21	0.05	0.12	0.07	0.00
	2.00	0.08	0.15	0.17	0.05	0.12	0.18	0.21
	4.00	0.29	0.20	0.25	0.55	0.25	0.46	0.38
	8.00	0.29	0.60	0.62	0.15	0.29	0.57	0.62
	16.00	0.50	0.90	0.79	0.95	0.46	0.93	0.92

Note. The choice was between the Green and Blue keys.

Table B3

Proportion of choices of the Red key for each sample duration during the two sets of stimulus–response generalization tests (“Short” and “Long” sessions).

Sessions	Sample (s)	Pigeon						
		P203	P287	P307	P323	P611	P766	P779
Short	1	1.00	1.00	0.93	0.96	0.93	1.00	1.00
	2	0.90	1.00	0.83	0.90	1.00	0.88	0.90
	4	0.82	0.60	0.73	0.82	0.88	0.83	0.45
	8	0.29	0.15	0.64	0.42	0.71	0.12	0.00
	16	0.05	0.05	0.17	0.00	0.08	0.00	0.00
Long	1	0.92	1.00	0.79	1.00	0.79	0.86	0.88
	2	0.79	0.80	0.62	0.95	0.71	0.75	0.58
	4	0.79	0.65	0.29	0.50	0.71	0.57	0.46
	8	0.33	0.25	0.25	0.40	0.25	0.11	0.12
	16	0.00	0.20	0.12	0.00	0.12	0.07	0.08

Note. The choice was between the Red and Yellow keys.