

*THE INTERACTION OF TEMPORAL GENERALIZATION GRADIENTS PREDICTS THE CONTEXT EFFECT*

ANA CATARINA VIEIRA DE CASTRO AND ARMANDO MACHADO

UNIVERSITY OF MINHO, PORTUGAL

In a temporal double bisection task, animals learn two discriminations. In the presence of Red and Green keys, responses to Red are reinforced after 1-s samples and responses to Green are reinforced after 4-s samples; in the presence of Blue and Yellow keys, responses to Blue are reinforced after 4-s samples and responses to Yellow are reinforced after 16-s samples. Subsequently, given a choice between Green and Blue, the probability of choosing Green increases with the sample duration—the context effect. In the present study we asked whether this effect could be predicted from the stimulus generalization gradients induced by the two basic discriminations. Six pigeons learned to peck Green following 4-s samples ( $S^+$ ) but not following 1-s samples ( $S^-$ ) and to peck Red following 4-s samples ( $S^+$ ) but not following 16-s samples ( $S^-$ ). Temporal generalization gradients for Green and Red were then obtained. Finally, the pigeons were given a choice between Green and Red following sample durations ranging from 1 to 16 s. Results showed that a) the two generalization gradients had the minimum at the  $S^-$  duration, an intermediate value between the  $S^-$  and the  $S^+$  durations, and the maximum at the  $S^+$  as well as more extreme durations; b) on choice trials, preference for Green over Red increased with sample duration, the context effect; and c) the two generalization gradients predicted the average context effect well. The Learning-to-Time model accounts for the major trends in the data.

*Key words:* context effect, temporal generalization, temporal discrimination, quantitative model, Learning-to-Time model, key peck, pigeon

With the aim of contrasting some models of timing, Machado and Keen (1999) developed the double bisection task, a modified version of one of the most well-known procedures used to study timing in animals and humans—the temporal bisection task. In a typical temporal bisection task, the subjects learn to discriminate between two samples of different durations. For instance, a pigeon is presented with a sample light illuminated for either 1 or 4 s and then with two comparison stimuli, a Red key and a Green key. The pigeon receives food if it pecks Red following 1-s samples and Green following 4-s samples. After the discrimination is acquired the animal is presented with samples of intermediate durations and the preference for one of the keys, say, Green, is assessed. It is commonly found that the proportion of responses to Green increases

with sample duration from about 0 to about 1, with the indifference point close to the geometric mean of the two training durations (Catania, 1970; Church & Deluty, 1977; Fetterman & Killeen, 1991; Platt & Davis, 1983; Stubbs, 1968; for summaries, see Gallistel, 1990; Richelle & Lejeune, 1980; Shettleworth, 1998).

In the double bisection procedure (see Figure 1) the animals learn not one but two temporal discriminations, which we call Type 1 and Type 2. In the Type 1 discrimination they learn to choose Red over Green after 1-s samples and Green over Red after 4-s samples. In the Type 2 discrimination they learn to choose Blue over Yellow after 4-s samples and Yellow over Blue after 16-s samples. Next, generalization tests are conducted in which new pairs of comparisons are introduced following different sample durations. Of critical importance to contrast timing models are the test trials on which Green and Blue are presented together, for these two comparisons were reinforced following the same sample duration of 4 s.

Figure 2 shows the typical result of these test trials: Preference for Green over Blue increases with sample duration (Machado & Keen, 1999). This finding is called the context effect because, even though the choices of Green

---

The authors thank the students from the Animal Learning and Behavior laboratory of the University of Minho for their helpful comments on the paper. Ana Catarina Vieira de Castro was supported by a PhD fellowship and Armando Machado by a grant from the Portuguese Foundation for Science and Technology (FCT).

Address correspondence to Ana Catarina Vieira de Castro or Armando Machado at Escola de Psicologia, Universidade do Minho, 4710 Braga, Portugal (e-mail: castro.anacatarina@gmail.com or armandom@psi.uminho.pt).

doi: 10.1901/jeab.2012.97-263

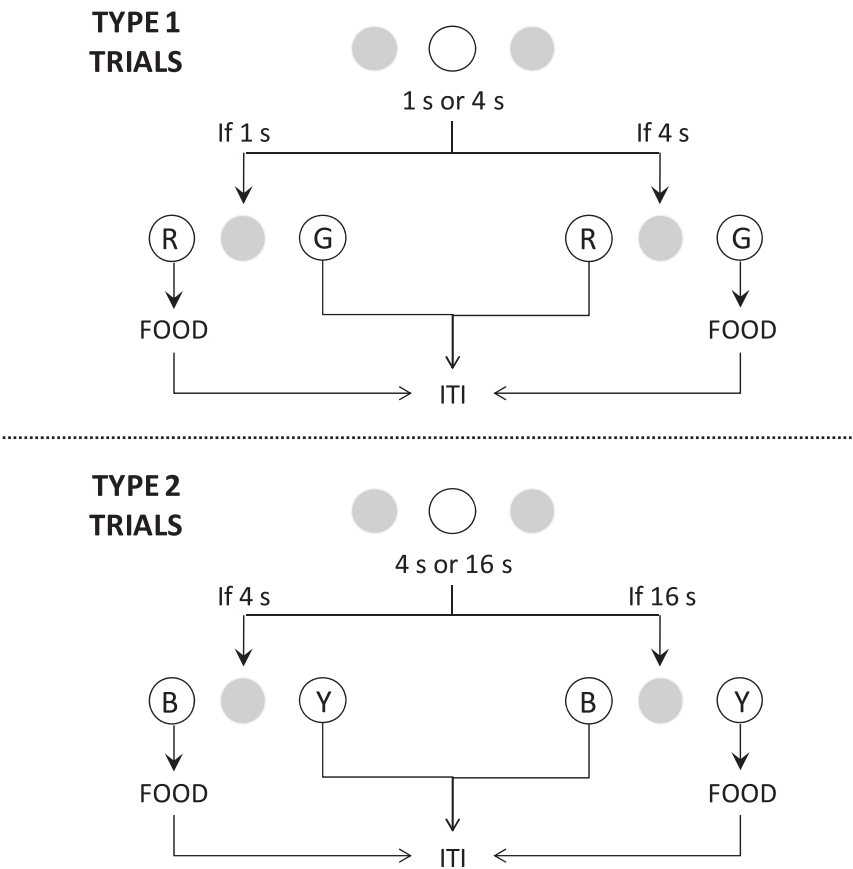


Fig. 1. Structure of the two types of training trials in a double bisection task. On *Type 1 trials*, pecking red or green is reinforced after 1-s or 4-s samples, respectively. On *Type 2 trials*, pecking blue or yellow is reinforced following 4-s or 16-s samples, respectively.

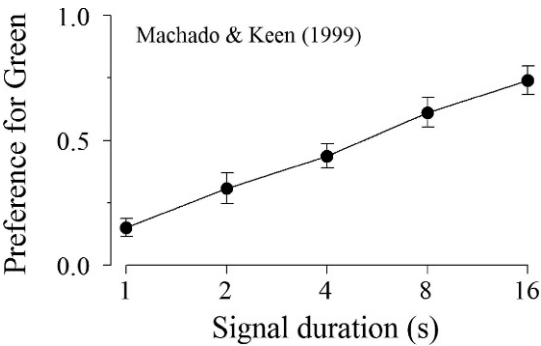


Fig. 2. Proportion of responses to Green, given a choice between Green and Blue, as a function of sample duration. The Green and Blue comparisons had been reinforced following 4-s samples (see Figure 1). The data are from Machado and Keen (1999); N=8 pigeons; the bars show the SEM.

and Blue were reinforced following the same sample duration, the context in which such reinforcement occurred differed. Choosing Green was reinforced in a context in which a shorter sample signaled extinction for Green, whereas choosing Blue was reinforced in a context in which a longer duration signaled extinction for Blue. As a series of studies using different versions of the double bisection procedure have shown, the context effect is quite robust and general (Arantes, 2008; Arantes & Machado, 2008; Machado & Arantes, 2006; Machado & Keen, 1999; Machado & Pata, 2005; Oliveira & Machado, 2008, 2009).

The context effect is theoretically important because it differentiates timing models. Consider the leading model, Scalar Expectancy Theory (SET; e.g., Church, 2003; Gibbon, 1977; Gibbon, Church, & Meck, 1984). According to SET's standard account for temporal bisec-

tion (e.g., Gibbon, 1981, 1991), the animal represents the reinforced sample durations in distinct memory stores. Therefore, in the double bisection task, the animal forms four stores, but only two are critical for the context effect, the store representing 4-s samples and associated with the Green comparison and the store representing 4-s samples associated with the Blue comparison. Moreover, according to SET, the contents of each store depend exclusively on its corresponding sample duration—they are context independent. Hence, the stores associated with Green and Blue will have identical representations causing SET to predict that the preference for Green over Blue will not vary with the sample duration. In short, because temporal memories are assumed context-independent, SET fails to predict the context effect.

In contrast with SET, the Learning-to-Time or LeT model (Machado, 1997) or its newer version (Machado, Malheiro & Erlhagen, 2009) is a context-dependent timing model. It assumes that, in a temporal discrimination task, the sample stimulus activates a set of states serially such that, on the average, the active state at the end of the sample varies directly with the sample duration (see also Killeen & Fetterman, 1988). Moreover, each state is coupled with the comparison stimuli and the degree of the coupling changes with reinforcement and extinction. Reinforcement strengthens and extinction weakens the coupling between the active state and the chosen comparison. Finally, the probability of choosing one or the other comparison depends on two factors, which state is active at the end of the sample (a function of sample duration) and which of the two couplings is stronger (a function of the reinforcement contingencies). Because the couplings between the active states and the comparisons depend on the sample durations and on which comparison is reinforced and which is extinguished following each sample, the model is context-dependent.

Before we explain in greater detail how LeT accounts for the context effect, we present a qualitative account of that effect that does not rely on a specific model. The account conceives of a simultaneous discrimination as two successive discriminations operating simultaneously. In the double bisection task, this means that pecking Green and Blue are

conceived of as operants controlled by two sample durations, an  $S^+$  and an  $S^-$ . For Green, the  $S^+$  and  $S^-$  are the 4-s and 1-s samples, respectively; for Blue, the  $S^+$  and  $S^-$  are the 4-s and 16-s samples, respectively. In addition, the account assumes that choice between the two comparison stimuli may be predicted from the generalization gradients induced by the discriminative training (Honig, 1962; Spence, 1937).

Machado and Pata (2005) suggested that the context effect could be the result of a peak-shift-like phenomenon in the temporal domain. Peak shift refers to a shift in the peak of a generalization gradient after intradimensional discrimination training, a shift in the direction opposite to the negative stimulus ( $S^-$ ). If we assume that the effect of the  $S^-$  is to shift the peak of the generalization gradient away from  $S^+$  (Hanson, 1959; see also Bloomfield, 1967; Purtle, 1973; Spence, 1937), then, in the double bisection task, the gradient for Green will have its maximum at a value greater than 4 s, whereas the gradient for Blue will have its maximum at a value less than 4 s. The net effect of these two shifts is that the gradient for Blue will be above the gradient for Green for durations less than 4 s, but the gradient for Green will be above the gradient for Blue for durations greater than 4 s. The relative positions of the two gradients could explain why preference for Green over Blue increases with sample duration.

A few studies have examined peak shift in postdiscrimination temporal-generalization gradients. Mellgren, Mays, and Haddad (1983) with rats and Spetch and Cheng (1998) with pigeons obtained generalization gradients resembling step functions (a high response rate in the presence of the  $S^+$  and higher values, and the same low response rate in the presence of the  $S^-$  and lower values); no peak shift was found. In Elsmore's (1971) study, 2 pigeons showed a peak shift effect and the other 2 responded approximately at the same rate to the  $S^+$  and to the values on the far side of it. More recently, Russell and Kirkpatrick (2007) found, with rats, gradients with a clear peak shift and gradients that peaked at a duration away from the  $S^+$  in the direction opposite the  $S^-$ , but did not decrease significantly for the remaining durations. Also recently, Bizo and McMahon (2007) reported a reliable peak shift effect in humans. To

Table 1

How couplings between the behavioral states and the choice comparisons change during training, according to LeT.

Comparison stimulus	Behavioral state		
	S1	S4	S16
Green	+ → 0	+ → ++	+
Blue	+	+ → ++	+ → 0

Note. S1, S4 and S16 represent the most likely active states at the end of the 1-s, 4-s, and 16-s samples, respectively. The arrows show the direction of change. “0”, “+”, and “++” stand for weak, moderate, and strong couplings.

summarize, the evidence for peak shift in the temporal domain is mixed, with one study reporting positive evidence, two studies reporting negative evidence and two others reporting, at least for some subjects, positive evidence.

Peak shift, however, can be seen as a special case of the area shift phenomenon. An area shift occurs when more than 50% of the area of the gradient lies on the side of S<sup>+</sup> opposite the S<sup>−</sup> (Rilling, 1977; see e.g. Cheng, Spetch, & Johnston, 1997, for an area shift effect in the spatial domain). It follows from the definition that whenever a peak shift occurs, an area shift also occurs, but the reverse is not necessarily the case. In fact, whereas area shift is reliably found in the temporal domain following intradimensional discrimination training, including in all the studies reported above, peak shift does not always occur. Hence, in what follows we focus our attention in the more general phenomenon of area shift.

The area shift account of the context effect would work as follows. If we assume that a) because the choice of Green is extinguished after 1-s samples, the area of the gradient for Green will be shifted to the right (i.e., toward durations longer than 4 s); b) because choice of Blue is extinguished after 16-s samples, the area of the gradient for Blue will be shifted to the left (i.e., toward durations shorter than 4 s); c) because of (a) and (b), the generalization gradient for Green will be below that for Blue at 1 s, approximately equal to it at 4 s, and above it at 16 s; d) if choice following a t-s sample depends on the relative heights of the two gradients at t s, then the proportion of Green choices should increase with sample

duration, the context effect. According to this account, the exact location of the peaks of the gradients is irrelevant.

The area shift hypothesis is consistent with how the LeT model accounts for the context effect. We present the general argument qualitatively and defer until the Discussion a more quantitative treatment (for additional details, see Machado et al., 2009). Table 1 helps to understand the argument. Let S1, S4, and S16 represent the behavioral states most likely to be active at the end of 1-s, 4-s and 16-s samples, respectively. Initially, these states are equally coupled with the two critical comparisons, the Green and Blue keys, and the degree of the coupling is represented by the “+” symbol in the table. During Type 1 trials, choices of Green will be reinforced after 4-s samples and extinguished after 1-s samples. Hence, the coupling of S4 with Green will increase to “++” but the coupling of S1 with Green will decrease to “0”. Similarly, during Type 2 trials, choices of Blue will be reinforced after 4-s samples and extinguished after 16-s samples and consequently the coupling of S4 with Blue will increase to “++” but the coupling of S16 with Blue will decrease to “0”. Two consequences follow from the coupling profiles acquired during training. First, if we run generalization tests by presenting either the Green or the Blue key following samples ranging from 1 s to 16 s, then we should obtain the two area-shifted generalization gradients mentioned above, a gradient shifted to the right (Green) and a gradient shifted to the left (Blue). Second, if the animal is given a choice between Green and Blue following samples ranging from 1 s to 16 s, then it should prefer Blue following 1-s samples, be indifferent following 4-s samples, and prefer Green following 16-s samples—the context effect.

The foregoing account also justifies the designation of a context effect. Although the animal learns to respond to Green and Blue following the same 4-s samples, the context in which such learning takes place differs: Responses to Green are extinguished following 1-s samples, whereas responses to Blue are extinguished following 16-s samples. According to LeT, the difference in the learning context explains why preference for Green over Blue increases with sample duration.

To summarize, the LeT model instantiates the area shift hypothesis. The coupling profiles acquired during training induce area-shifted generalization gradients which combined produce the context effect.

The main purpose of the present study was to test directly this generalization-based account of the context effect. To that end, we simplified the double bisection task by retaining only the elements central to the account, the two operants and their reinforcement contingencies—reinforcement after 4-s samples for both operants; extinction after 1-s samples for one operant; and extinction after 16-s samples for the other operant. Specifically, in the new task, the pigeons started by learning one of the two basic discriminations, to peck A after 4-s samples, the  $S^+$ , and not to peck A after 1-s samples, the  $S^-$ . Then, we varied the sample duration from 1 to 16 s to obtain the temporal generalization gradient for pecking A. Next, the pigeons learned the second basic discrimination, to peck B after 4 s ( $S^+$ ), and not to peck B after 16 s ( $S^-$ ) and then we varied the sample duration to obtain the gradient for pecking B. Finally, we presented the pigeons with samples ranging from 1 to 16 s and gave them a choice between the two operants, peck A and peck B. According to our hypothesis, the preference of A over B should increase with sample duration, and, in addition, the two generalization gradients, with their areas shifted in opposite directions, should predict the preference data.

One potential difficulty with the task described above is that it involves successive choice during the basic discriminations (e.g., peck A following 4-s samples; do not peck A following 1-s samples), but simultaneous choice during the final test phase (peck A or B). The novelty of simultaneous choice during the final test phase could mask the effects of the generalization gradients induced by the two basic discriminations. To eliminate this potential difficulty, we introduced a “dummy” alternative during the basic discriminations. Specifically, after 1-s and 4-s samples, the pigeon was given a choice between pecking A (e.g., a Green key) and pecking a key with a vertical bar. Choice of A was reinforced after the 4-s samples, but not after the 1-s samples, and choice of the vertical bar was never reinforced. Similarly, after 4-s and 16-s samples, the pigeon was given a choice between pecking B (e.g., a Red key) and pecking a key

with a vertical bar. Choice of B was reinforced after the 4-s samples, but not after the 16-s samples, and again choice of the vertical bar was never reinforced. The “dummy” alternative gave the pigeons experience with simultaneous choice from the beginning of the experiment.

## METHOD

### *Subjects*

Six adult pigeons (*Columba livia*) participated in the experiment. The birds had previous experience with the time-left procedure, but not with matching-to-sample tasks. They were maintained at 80% of their free-feeding body weights throughout the experiment and were housed in individual home cages with water and grit continuously available. A 13:11 h light/dark cycle, beginning at 8:00 am, was in effect in the pigeon colony.

### *Apparatus*

Three identical Lehigh Valley® operant chambers were used. Each chamber was 34 cm high, 35 cm long and 31 cm wide. The walls and ceiling were made of aluminum and the floor was wire mesh. The response panel contained three circular keys, 2.5 cm in diameter, arranged in a horizontal row, 22.5 cm above the floor, and 9 cm apart, center to center. The keys could be illuminated with yellow, green and red lights and with a vertical white bar on a dark background. On the back wall of the chamber, 4 cm below the ceiling, a 7.5-W houselight provided general illumination. Reinforcement consisted of mixed grain delivered by a hopper that was accessible through a 6×5-cm opening, centered on the response panel 8.5 cm above the floor. A 7.5-W white light illuminated the opening whenever a reinforcer was available. An outer box enclosed the operant chamber. The box was equipped with a ventilation fan that circulated air through the chamber and provided masking noise. A personal computer controlled all experimental events and recorded the data.

### *Procedure*

The pigeons learned two temporal discriminations, 1 s versus 4 s and 4 s versus 16 s, referred to as Type 1 and Type 2, respectively.



Table 2

Comparison Stimuli, Sample duration, and the number of reinforced and non-reinforced trials per session.

Phase	Condition	Comparison Stimuli		Sample duration (s)				
				1	2	4	8	16
Type 1	Training	Bar	Green	30–		30+		
	Pretesting	Bar	Green	30–		24+,6–		
	Testing	Bar	Green	30–	2–	24+	2–	2–
Type 2	Training	Bar	Red			30+		30–
	Pretesting	Bar	Red			24+,6–		30–
	Testing	Bar	Red	2–	2–	24+	2–	30–
Type 1 + Type 2 (Combined)	Training	Bar	Green	20–		20+		
		Bar	Red			20+		20–
	Pretesting	Bar	Green	20–		14+,6–		
		Bar	Red			14+,6–		20–
	Testing	Bar	Green	20–		14+		
		Bar	Red			14+		20–
		Green*	Red	6–				6–
		Green*	Red		4–	4–	4–	
		Green**	Red	6–			6–	
		Green**	Red		4–	4–		4–

Note. N+ means that N trials per session were reinforced (provided the choice was correct) and N- means that N trials per session were in extinction.  
\* For pigeons P890, P948, P178, and P841  
\*\* For pigeons P170 and PG12

Half of the birds learned the Type 1 discrimination first and the Type 2 discrimination next, whereas the other half learned the discriminations in the opposite order. For 3 pigeons (P170, P178, and P841) a Green key was associated with the 4-s samples from Type 1 trials and a Red key was associated with the 4-s samples from the Type 2 trials; for the other 3 pigeons, the reverse assignment was in effect. However, for clarity we describe the procedure and the experimental results as if all of the birds had learned the Type 1 discrimination first and had the Green key assigned to the 4-s sample of the Type 1 discrimination.

Table 2 summarizes the procedure. It consisted of three phases, one in which only the Type 1 discrimination was trained, another in which only the Type 2 discrimination was trained, and yet another in which both Type 1 and Type 2 discriminations were trained. In addition, each phase consisted of three conditions: Training, during which the discrimination was learned and all correct choices were reinforced; Pretesting, during which some correct choices were extinguished to adapt the pigeons to the intermittent reinforcement that would be in effect during the next condition; and Testing, during which new sample durations were introduced to obtain

the generalization gradients (Phases 1 and 2); or new sample durations and a new pair of choice keys, Green and Red, were introduced to examine the context effect (Phase 3).

*Phase 1.* The general structure of a Type-1 Training Condition trial was as follows. After a dark, 20-s ITI, the houselight was turned on and the center key was illuminated with yellow light. After the sample duration elapsed (1 s or 4 s) the center keylight was turned off and the side keys were illuminated, one with a green light and the other with a white vertical bar on a dark background. The two side keys remained illuminated for at least 6 s. If the S<sup>+</sup>, 4-s sample had been presented, the first peck at a choice key after 6 s turned all keylights and the houselight off, and if the response was correct (pecking Green), it activated the food hopper. The hopper duration varied across birds from 3 s to 9 s in order to maintain body weight with minimal extra session feeding. After food, the ITI followed. If the response was incorrect (pecking the vertical bar), the ITI started immediately and the trial was repeated. If the bird made three consecutive errors, only the Green key was presented (correction procedure). If the S<sup>-</sup>, 1-s sample had been presented, the side keys and the houselight were turned off after the 6-s period regardless

of the animal's behavior and then the ITI started.

Learning was assessed by the relative response rate during the 6-s choice period, that is, by the discrimination ratio " $\#S^+ / (\#S^+ + \#S^-)$ ", where  $\#S^+$  is the total number of responses to Green after the 4-s samples, and  $\#S^-$  is the total number of responses to Green after the 1-s samples. Initially, the learning criterion was set at a discrimination ratio of .85 or above, excluding repeated trials, for five consecutive sessions. For some subjects, the criterion had to be reduced to .80 because the initial criterion proved to be too hard to reach.

Sessions 1 to 12 consisted of 60 trials, 30 with the  $S^+$  and 30 with the  $S^-$  (see first row of Table 2). Within each set of 30 trials, 15 had Green on the left key and 15 had Green on the right key. However, as some birds continued to peck Green during the  $S^-$  trials, in subsequent sessions we increased the proportion of  $S^-$  trials: 20  $S^+$  and 40  $S^-$  trials for pigeons P170, P178, P890, and PG12, and 16  $S^+$  and 44  $S^-$  trials for pigeons P841 and P948. For the last 2 pigeons we also reduced to 3 s the duration of the choice period in order to make their choice responses more contiguous with reinforcement and extinction. Before advancing to the next condition, though, all pigeons returned to the original session structure with 30  $S^+$  and 30  $S^-$  trials and a 6-s choice period. The Training Condition lasted 42 sessions on average (range: 32 to 69).

In the Pretesting Condition (see second row of Table 2), extinction trials were introduced. In addition to not ending with food, even when a choice was correct, extinction trials also were not repeated when the choice was incorrect. Sessions consisted of 60 trials, 30  $S^+$  (24 reinforced and 6 unreinforced, 12 and 3 for each left/right assignment, respectively) and 30  $S^-$  (15 for each left/right assignment). Pretesting lasted until the learning criterion was met ( $M = 7$  sessions; range: 5 to 12).

In the Testing Condition (see third row of Table 2), each session consisted of 60 trials, 24  $S^+$  and 30  $S^-$  (12 and 15 for each left/right assignment, respectively) and 6 test trials. On test trials the sample duration equaled 2, 8 or 16 s, and each sample was presented twice per session (once for each left/right assignment). If during any test session a bird's discrimination ratio decreased by more than .05 from its criterion (either .85 or .80), the bird was

returned to training for at least two sessions and until its performance met the criterion again. Testing continued until 10 test sessions were completed ( $M = 12$  sessions; range: 10 to 19).

*Phase 2.* After the pigeons completed the three conditions of Phase 1, they were exposed to exactly the same conditions but with the Type 2 discrimination (see Table 2). Training lasted from 16 to 53 sessions ( $M = 30$ ), Pretesting from 5 to 17 ( $M = 7$ ) and Testing from 12 to 26 ( $M = 15$ ).

*Phase 3.* Initially, the Type 1 and Type 2 discriminations alternated across sessions for a minimum of five sessions and until the pigeons reached the criterion. Next, both discriminations were presented within the same session (see Table 2, "Type 1 + Type 2"). Each session consisted of 80 trials, 40 of Type 1 and 40 of Type 2. For each type, there were 20  $S^+$  and 20  $S^-$  trials (10 for each left/right assignment). Training lasted from 6 to 31 sessions ( $M = 16$ ).

During the Pretesting Condition, the session structure remained the same except that a few  $S^+$  trials were conducted in extinction (see Table 2). Pretesting lasted from 5 to 18 sessions ( $M = 7$ ).

During the Testing Condition, each session consisted of 68 training trials and 12 test trials for a total of 80 trials. During the test trials, Green and Red were presented together for the first time. In addition, the sample duration on test trials equaled 1 s or 16 s during the first 6 sessions and 2 s, 4 s or 8 s during the next nine sessions (see Table 2, 3<sup>rd</sup> and 4<sup>th</sup> rows from bottom). Due to a programming error, for pigeons P170 and PG12, the sample durations equaled 1 s and 8 s during the first six sessions and 2 s, 4 s and 16 s during the next nine sessions (Table 2, last two rows). Each test trial was presented six times per session during the first six sessions, and four times per session during the last nine sessions.

If during any test session a bird's discrimination ratio for Type 1 or Type 2 trials decreased by more than .05 from its current criterion (either .80 or .85), the bird returned to training for at least two sessions and until its performance recovered. However, for pigeons P948 and P841, testing disrupted the Type 1 discrimination to such extent that the criterion had to be lowered. Pigeon P948 recovered partly and completed the tests with the

criterion set at .75 for the Type 1 discrimination; for Type 2 trials, the criterion was maintained at .80. Pigeon P841 did not recover the Type 1 discrimination, even after 20 sessions of Training and Pretesting. Because its ratio for Type 2 trials remained above .80 and the ratio for Type 1 trials stabilized around .65, testing was resumed. The Testing Condition, with intercalated training sessions when required, lasted from 15 to 42 sessions across birds ( $M = 22$ ).

## RESULTS

*Training.* By the 10th session of training 5 of the 6 pigeons were rarely pecking the key with the vertical bar (fewer than two pecks per session on average); P841 was the exception, for this pigeon maintained some responding to the vertical bar throughout the experiment (from 4.5 to 30.8 pecks per session).

All subjects learned the two basic discriminations. They required a similar number of sessions to learn the Type 1 and Type 2 discriminations ( $M = 35$  and 37 sessions, respectively,  $t(5) = 0.29$ , ns). And though the discrimination learned in the second place (Type 1 for half of the pigeons and Type 2 for the other half) required fewer sessions than the discrimination learned in the first place ( $M = 30$  and 42, respectively), the difference was not statistically significant ( $t(5) = 1.83$ ,  $p = .13$ ). During the last five training sessions, the discrimination ratio averaged .93 on Type 1 trials (range: .84–.99) and .92 on Type 2 trials (range: .89–.97). When the two discriminations were combined into the same session, the overall discrimination ratio averaged .90 (range: .82–.99).

The total number of sessions required to learn the two discriminations (i.e., Type 1 alone + Type 2 alone + Types 1 and 2 combined) ranged from 93 to 158 ( $M = 118$ ). This number is significantly higher than the number obtained in previous studies with a simultaneous double bisection procedure (e.g., 29 to 34 in Machado & Keen, 1999; 27 to 65 in Machado & Pata, 2005; 26 to 45 in Machado & Arantes, 2006; 28 to 68 in Oliveira & Machado, 2008; and 26 to 38 in Oliveira & Machado, 2009). However, the number is close to that obtained by Arantes (2008) with a successive double bisection procedure ( $M = 103$ , range: 51–133). These results (see also

Arantes & Grace, 2008) show that simultaneous double bisection tasks are easier to learn than successive ones, presumably because learning to not respond following an  $S^-$  is easier if the same  $S^-$  is an  $S^+$  for responding on another key.

*Stimulus generalization testing.* Figure 3 shows the stimulus generalization gradients for the two temporal discriminations. The response rate on the Green or Red keys during the 6-s choice period is plotted against sample duration. All pigeons exhibited similar generalization gradients: Response rate was lowest at the  $S^-$ , intermediate between the  $S^-$  and the  $S^+$ , and highest at the  $S^+$  and durations further away from the  $S^-$ . For pigeon P890, the gradient for Type 2 trials decreased for the shortest sample duration of 1 s.

Figure 4 shows that the average gradient was similar for the two discriminations. Response rate increased from the  $S^-$  to the  $S^+$  and then it remained high (Type 1) or decreased slightly (Type 2). A two-way repeated-measures ANOVA, with sample duration and type of discrimination as factors revealed significant effects of sample duration [ $F(4, 20) = 38.95$ ,  $p < .001$ ], type of discrimination [ $F(1, 5) = 15.10$ ,  $p < .05$ ], and their interaction [ $F(4, 20) = 59.72$ ,  $p < .001$ ].

To analyze the symmetry of the two generalization gradients, we reflected the curve for Type 2 trials around the line  $t = 4$ . The bottom panel of Figure 4 shows the result. The (reflected) gradient for Type 2 trials was below the gradient for Type 1 trials, particularly at the geometric means of the training durations (i.e.,  $t = 2$  s for Type 1 and  $t = 8$  s for Type 2) and at the extreme durations on the opposite side of  $S^-$  ( $t = 16$  s for Type 1 and  $t = 1$  s for Type 2). A two-way, repeated-measures ANOVA with sample duration and type of discrimination as factors revealed that all effects were significant (type of discrimination,  $F(1, 5) = 15.10$ ,  $p < .05$ , sample duration,  $F(4, 20) = 68.66$ ,  $p < .001$ , and their interaction  $F(4, 20) = 6.17$ ,  $p < .005$ ).

Paired  $t$ -tests conducted for the  $t = 2$  s,  $t = 8$  s and  $t = 16$  s samples (defined by the Type 1 discrimination in the bottom panel of Figure 4) revealed a significant difference between response rate on  $t = 2$  s and  $t = 16$  s [ $t(5) = 5.42$ ,  $p < .05$  and  $t(5) = 3.66$ ,  $p < .05$ , respectively]; on  $t = 8$  s the difference approached significance [ $t(5) = 2.27$ ,  $p = .07$ ].



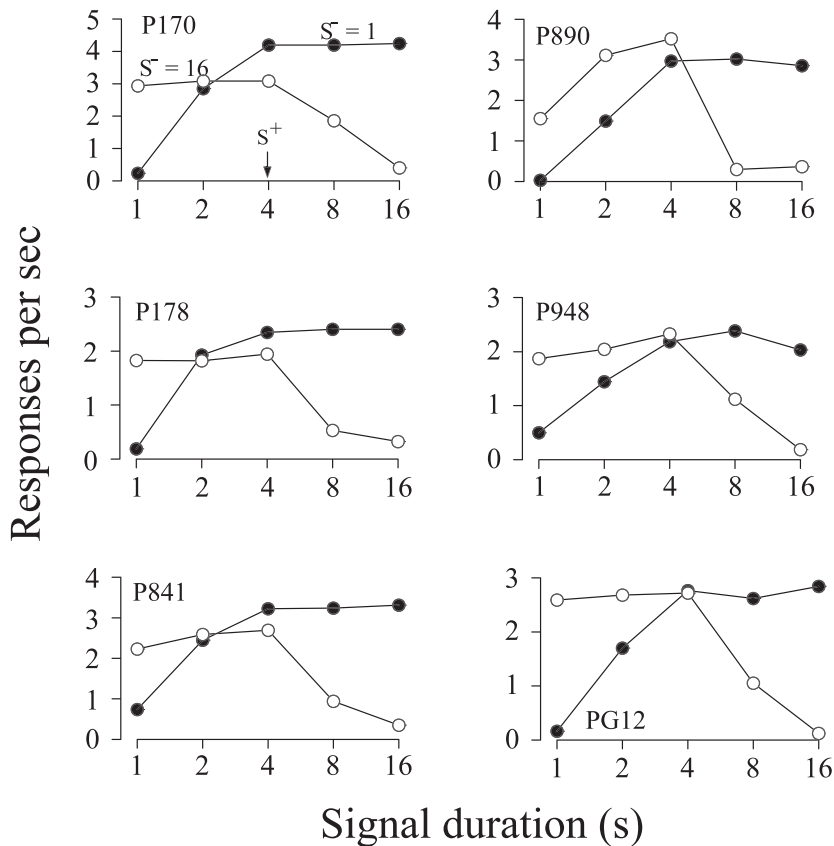


Fig. 3. Generalization gradients obtained after the Type 1 ( $S^+ = 4s$ ,  $S^- = 1s$ ) and Type 2 ( $S^+ = 4s$ ,  $S^- = 16s$ ) discriminations. Note the logarithmic scale on the x-axis.

*Stimulus-response generalization testing.* The data analysis from the choice tests addresses an important question: Does the context effect hold even when only two responses are reinforced, that is, when pecking Green and Red are reinforced following 4-s samples but not following 1-s and 16-s samples, respectively? Figure 5 shows the individual and average results. The filled circles show that the preference for Green—defined by the ratio between the total number of responses to Green and the total number of responses to Green plus Red—tended to increase with sample duration. A repeated-measures ANOVA revealed a significant effect of sample duration,  $F(4,20) = 13.79$ ,  $p < .001$ .

The filled circles in the bottom panel show that, on the average, preference for Green increased monotonically from about .1 to about .7, thus reproducing the context effect previously found with other variants of the

double bisection task (compare Figures 2 and 5; see also, Arantes, 2008; Arantes and Machado, 2008; Machado and Arantes, 2006; Machado and Keen, 1999; Machado and Pata, 2005; Oliveira and Machado, 2008, 2009).

The individual curves show some variability across birds. Pigeons P890, P948 and P178 displayed steep, monotonically increasing preference functions; pigeon P841 also presented a monotonically increasing function, but it never exceeded .50; for pigeons PG12 and P170, preference for Green also tended to increase with sample duration, although not monotonically. Three subjects revealed a strong overall bias for one comparison, Green in the case of PG12 and Red in the case of P170 and P841.

The causes of the individual differences in the strength of the context effect remain unclear. On the one hand, the relatively weaker effect revealed by pigeon P841 could

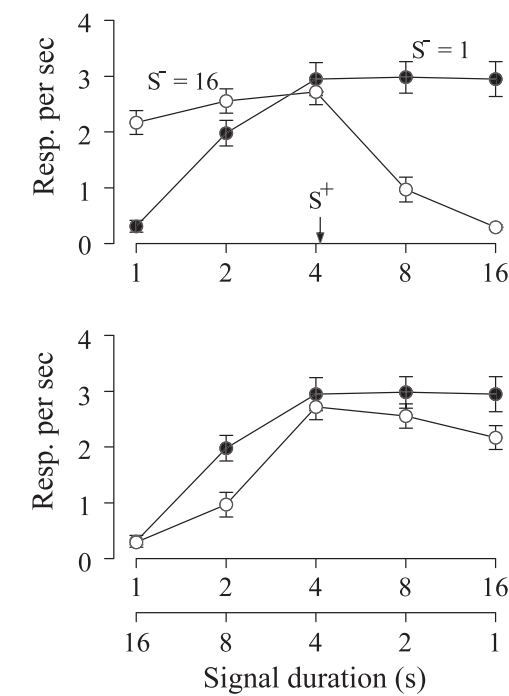


Fig. 4. Top panel. Average generalization gradients obtained after the Type 1 ( $S^+=4s$ ,  $S^-=1s$ ) and Type 2 ( $S^+=4s$ ,  $S^-=16s$ ) discriminations. Bottom panel. The gradient for Type 2 trials is reflected along the vertical line  $t=4$  (the inverted x-axis shows the original sample durations). The bars show the SEM. Note the logarithmic scale on the x-axis.

be due to its poor discrimination between the 1-s and 4-s samples during Phase 3, but the same cannot be said of pigeons P170 and PG12 because their discrimination ratios never fell below .80. On the other hand, pigeons P170 and PG12 started the choice test with sample durations of 1 s and 8 s, whereas the remaining pigeons started them with sample durations of 1 s and 16 s, but it is unclear how this difference could explain the differences in the strength of the context effect.

Another potential source of the individual differences may be the number of choice trials on which the pigeon pecked neither the Green nor the Red keys. Table 3 shows the number of these “empty” trials, out of 36, for each pigeon and sample duration. The number of empty trials was particularly high following the 8-s and 16-s samples. However, there was no clear relation between that number and the preference for Green. Consider the 16-s samples: Of the 3 pigeons that

Table 3  
Number of empty trials (out of 36) following each sample duration.

Pigeon	Sample duration				
	1	2	4	8	16
P170	0	0	1	13	33
P178	21	7	0	11	22
P841	0	0	0	7	4
P890	18	24	1	26	30
P948	2	1	0	6	1
PG12	7	10	0	20	27
Avg	8	7	0	14	20
$r_s$	-.03	.32	.42	.43	-.20
$p$	.48	.27	.20	.20	.35

Note.  $r_s$  is Spearman’s rank-order correlation between the number of empty trials and the preference for Green.  $p$  is the corresponding p-value under  $H_0$ .

showed a strong preference for Green, 2 had a large number of empty trials (P890 and P178) but 1 had only one empty trial (P948). The case was similar following 1-s samples. Spearman’s rank-order correlations between the number of empty trials and the preference for Green were not statistically significant (see Table 3, bottom lines).

Although variation in the number of empty trials does not seem to account for the differences in the context effect, the large number of empty trials following some sample durations raises an interpretative problem related to the measurement of preference. Consider the data for pigeon P170. The 33 empty trials following the 16-s samples mean that the pigeon responded on 3 trials only. On two of them, it chose Green, and on one of them it chose Red, which suggests a preference for Green. However, preference for Green measured by relative response rate was only .27 (see Figure 5) because response rate on Red was higher than on Green. Differences in response rate may have contaminated the preference measure.

This difficulty suggested a second way to measure preference. For each sample, we counted the number of trials on which the pigeon pecked Green at least once (even if it also pecked Red) and then divided it by the total number of trials on which the pigeon pecked at least one key. This “all-or-none” preference measure is not affected by differences in response rate.

The unfilled circles in Figure 5 show the “all-or-none” measure. It is clear that both

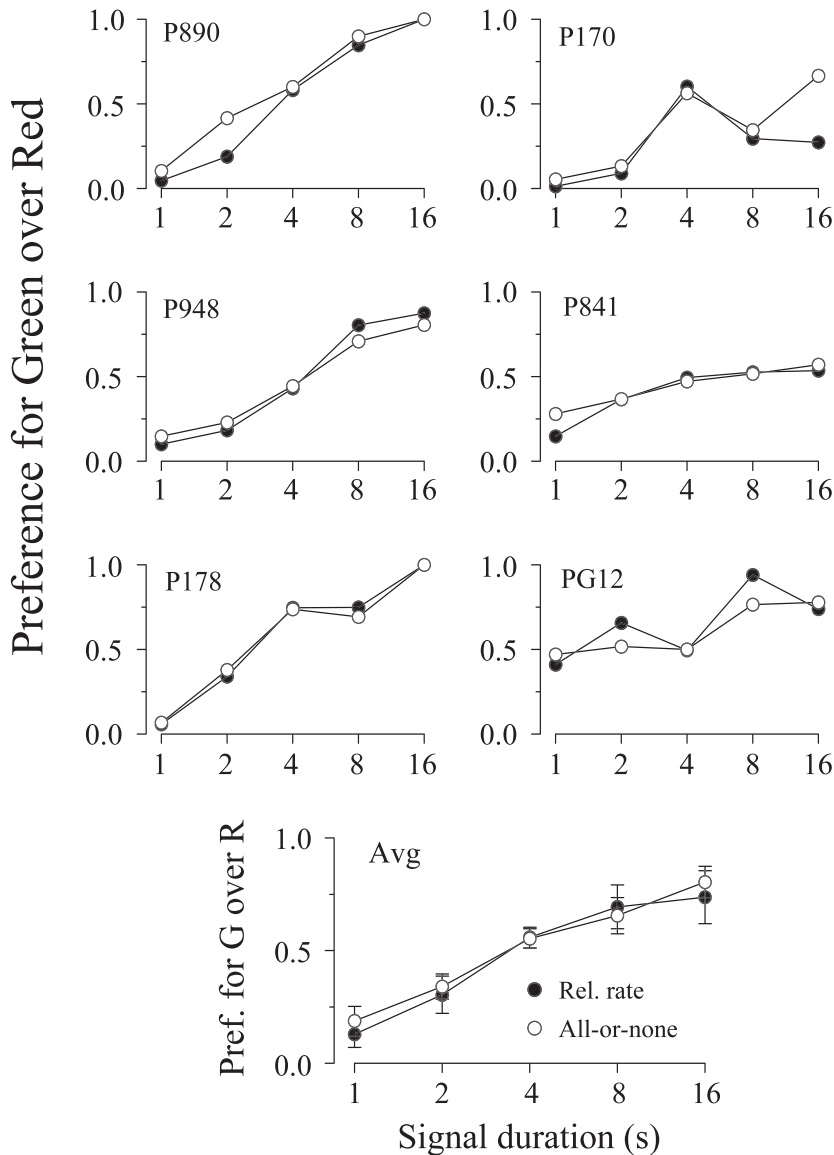


Fig. 5. The top panels show the individual functions relating preference for Green over Red to sample duration. Filled and unfilled circles correspond to different measures of preference, relative response rate or "all-or-none". The bottom panel shows the average results. The vertical bars show the SEM. Note the logarithmic scale on the x-axis.

preference measures yielded similar results. The correlations between them were strong for 4 pigeons ( $.91 \leq r^2 \leq .99$  for P178, P841, P890, and P948), and moderate for the remaining 2 ( $.60 \leq r^2 \leq .75$ ). The averages of the two measures (see bottom panel) were positively correlated ( $r^2 = .97$ ). We conclude that, with few exceptions, the two measures of preference were consistent and, therefore,

that relative response rate is a reliable measure of preference in the present task.

*Predicting preference functions from stimulus generalization gradients.* The major goal of the present study was to assess how well the context effect found in double bisection tasks could be predicted from the stimulus generalization gradients. To that end we carried out the following analysis. For each pigeon and

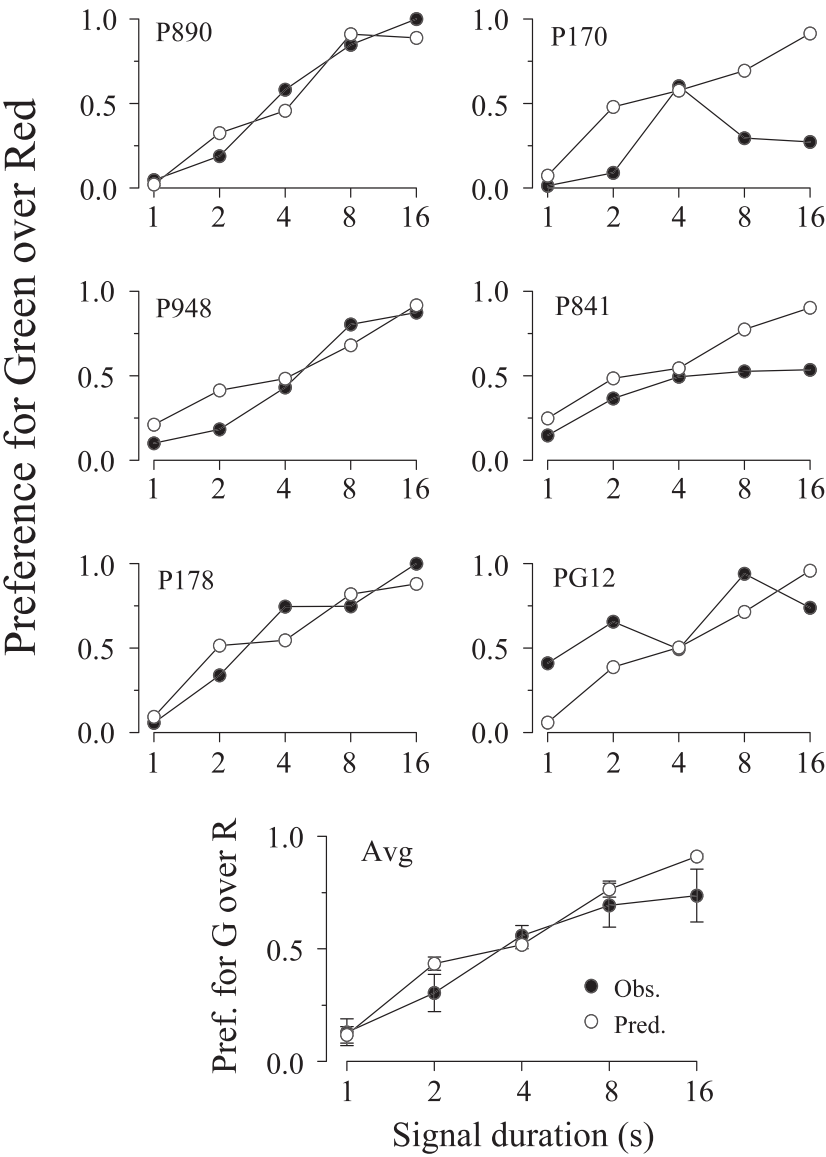


Fig. 6. Filled circles show the preference for Green obtained in the Testing Condition of Phase 3 with samples ranging from 1 to 16 seconds and Green and Red as comparisons. The unfilled circles show the preference for Green predicted from the gradients obtained during the stimulus generalization tests of Phases 1 and 2. The vertical bars show the SEM. Note the logarithmic scale on the x-axis.

sample duration, we used the response rates from Phases 1 and 2 displayed in Figure 3 to compute the predicted preference for Green. Specifically, we divided the absolute response rate to Green by the sum of the absolute response rates to Green and Red. Finally, we plotted the preference function predicted from the generalization gradients and compared it

with the preference function obtained during the choice test. Figure 6 shows the results. All six predicted functions (unfilled circles) expressed the context effect: The preference for Green increased monotonically with sample duration. In addition, for pigeons P890, P948, and P178, the predicted and obtained functions were similar ( $r^2 = .93, .90,$  and  $.85,$

respectively); for pigeon P841 the two functions differed appreciably at 8- and 16-s sample durations ( $r^2 = .81$ ); and for pigeons P170 and PG12, they differed considerably ( $r^2 = .54$  and  $.27$ , respectively). Interestingly, the pigeons that showed the greatest overlap between the predicted and the obtained preference functions also showed the strongest context effect.

The average of the predicted functions (see bottom panel) increased monotonically from about .1 to .9. Although more extreme than the average of the obtained preference functions, it had approximately the same shape. The two average functions were strongly correlated ( $r^2 = .93$ ).

### DISCUSSION

Six pigeons were exposed to a simplified version of the temporal double bisection procedure. On Type 1 trials they learned to peck a Green key following 4-s samples but not to peck it following 1-s samples; on Type 2 trials they learned to peck a Red key following 4-s samples but not following 16-s samples. After each discrimination was learned, a stimulus generalization gradient was obtained by varying sample duration from 1 to 16 s. Next, the two trial types were included in the same session and a stimulus-response generalization test was conducted in which the samples ranged from 1 to 16 s and the comparison stimuli were the Green and Red keys. The study had two interrelated goals, to investigate whether the two temporal generalization gradients could predict the context effect that was expected to occur during the choice trials, and to determine whether the generalization gradients as well as the context effect were consistent with the LeT model.

The choice test showed that, for 4 out of 6 pigeons, preference for Green over Red increased monotonically with the sample duration (Figure 5). For the other 2, the effect was weaker but the overall positive trend was observed. The context effect was obtained regardless of whether preference was assessed by relative response rate or by an "all-or-none" measure that is not influenced by differences in response rate between the two keys. And in spite of considerable procedural changes, the magnitude of the average effect was similar to that obtained in previous studies. Replicating the context effect in a situation where only two

responses were reinforced opened the possibility to test a generalization-based account of it.

The results showed that the average context effect could be predicted from the generalization gradients. However, the accuracy of the predictions varied across pigeons. For 3 of them (see Figure 6, left panels), the generalization gradients predicted well not only the positive trend, but also the specific values of the preference functions; for the other 3, the gradients predicted well only the positive trend of the preference functions (Figure 6, right panels). The mismatches between the predicted and observed function values revealed that the generalization gradients tended to overestimate the strength of the context effect (cf. the slopes of the two functions in the right panels of Figure 6).

The reasons for the mismatches remain unclear. They could stem from generalization decrement because the stimulus conditions during the training and generalization trials (i.e., only one of the keys, Green or Red, was present) differed from the stimulus conditions during the choice trials (i.e., both the Green and Red keys were present). Furthermore, some pigeons may have learned that the distinctive combination of Green and Red keys signaled extinction and, as a consequence, they did not respond on a large proportion of choice trials (Table 3). (The large number of empty trials, particularly following the extreme sample durations of 1 s and 16 s, may have been caused also by the preceding stimulus generalization tests, which also were conducted in extinction.). A small number of non-empty trials could have distorted the measure of preference. However, note that there was no significant correlation between the number of empty trials and the degree of preference for the Green or Red keys. Also, how well the generalization gradients predicted the context effect did not seem to depend on how much the basic discriminations were disrupted during testing. In any event, the nuisance of empty trials may be reduced in future studies by using a two-group design, with only one group exposed to generalization tests, and by using choice tests with partial but nondifferential reinforcement.

To account for the context effect in terms of generalization gradients, we advanced one qualitative account relying on area shift, and one quantitative account based on the LeT model. We analyze them in this order.



*Qualitative account.* According to the area shift account, gradients obtained after intradimensional training are asymmetrical, displaying a higher number of responses in the side of  $S^+$  opposite the  $S^-$ . Then, the area of the generalization gradient on Type 1 trials should be shifted toward durations longer than 4 s, and the area of the generalization gradient on Type 2 trials should be shifted toward durations shorter than 4 s.

The generalization gradients obtained in Phases 1 and 2 matched these predictions. All 12 gradients showed an area shift effect. The gradient produced by pigeon P890 during Type 2 trials (see the top right panel in Figure 3), in contrast with the other gradients, showed a clear reduction in response rate for stimulus values moved away from the  $S^+$  in the direction opposite to the  $S^-$ . It is conceivable that a closer spacing of the test stimuli in the vicinity of  $S^+$  (4 s) might have revealed a peak around  $t = 3$  s. In the other 11 gradients, response rate increased as the stimulus duration changed from the  $S^-$  to the  $S^+$ , and then it remained high as the stimulus duration continued to change past the  $S^+$  (see Figures 3 and 4). The shape of these gradients resembles a step function, with a low response rate at the  $S^-$ , an intermediate response rate between the  $S^-$  and the  $S^+$ , and a high response rate at and past the  $S^+$ . Because the gradients obtained in the present experiment revealed an area shift effect and could predict the context effect, we conclude that the area shift account works well in explaining our findings.

The “opposed” generalization gradients displayed in Figure 3 and in the top panel of Figure 4 are similar to the gradients obtained by Boneau and Honig (1964). These authors were the first to examine generalization gradients based upon conditional discrimination training. As in the present study, their pigeons learned two conditional discriminations. When the response key was illuminated with a 550-nm light, a white vertical bar added to the key was the  $S^+$  and the absence of the bar was the  $S^-$ ; when the response key was illuminated with 570-nm light, the contingencies were reversed, the  $S^+$  and  $S^-$  were the absence and presence of the bar, respectively. During generalization tests, the authors varied the wavelength of the light from 540 to 580 nm with either the bar present or with the bar absent. When the bar was present, the gener-

alization gradient was high at 540 and 550 nm and low at 570 and 580. When the bar was absent, the generalization gradient was low at 540 and 550 nm, high at 570 nm, and low again at 580 nm. If we replace wavelength with sample duration and the presence/absence of the bar with Green/Red keylight color, then we can conceive of the present study as extending Boneau and Honig’s findings to the timing domain. In both studies, the areas of the two gradients shifted in opposite directions and no peak shift effect was observed.

*A quantitative (and integrative) account: the LeT model.* To see how well the LeT model reproduces the major trends in the data, we ran a simulation of the entire experiment, following the same phases and conditions as the pigeons, and for a similar number of sessions. Throughout, the model parameters remained constant and their values were similar to those used in previous studies (Machado et al., 2009).

The model behaves as follows. Consider a Type 1 training trial. During the sample, the behavioral states are activated serially—first state 1, then state 2, etc. Each state remains active for  $\lambda$  seconds, with  $\lambda$  sampled at trial onset from a normal distribution with mean  $\mu = 1.0$  s and standard deviation  $\sigma = 0.3$  s. At the end of the sample one state is active, say, state  $n^*$ . This state is coupled with the operant response (e.g., Green) and the degree of the coupling, always between 0 and 1, is represented by the variable  $WG(n^*)$ . When there is only one key, as during training and generalization trials<sup>1</sup>,  $WG(n^*)$  yields the probability of emitting a response. (To keep the simulation simple and focus on the shape of the predicted curves, we did not attempt to model absolute response rate.) When a response occurs,  $WG(n^*)$  changes, increasing if the response is reinforced (i.e.,  $\Delta WG(n^*) = \beta(1 - WG(n^*))$ ) and decreasing if the response is extinguished (i.e.,  $\Delta WG(n^*) = -\alpha \cdot WG(n^*)$ ), where  $\beta = 0.2$  and  $\alpha = 0.04$  are learning parameters.

In the present simulation, the coupling of each state was initialized at 0.8 (i.e.,  $WG(n) = 0.8$ , for  $n = 1, 2, 3, \dots$ ) ensuring that, during the first session, a response on the Green key

<sup>1</sup>All training and generalization trials included also a key with a vertical bar. However, because pecks on that key were never reinforced, and after the first few sessions the pigeons rarely pecked at it, the model ignores it.

was very likely to happen on each trial. After 40 sessions, with reinforcement following 4-s samples and extinction following 1-s samples, the  $WG(n)$  values changed and the result is the curve with filled circles displayed in the top panel of Figure 7<sup>2</sup>. The early states, active mostly after the 1-s samples, lost their coupling with Green, but subsequent states, active mostly after the 4-s samples, had their coupling strengthened. The remaining states, almost never active during Type 1 trials, retained their initial coupling of 0.8.

The simulation details for Type 2 trials were the same except that a different vector,  $WR(n)$ , coupled the states with the Red key. The curve with open circles shows how the  $WR(n)$  values changed with training from their initial value of 0.8. The early states, active mostly after the 4-s samples, either retained or strengthened their couplings with Red, whereas the states active mostly after the 16-s samples lost most of their coupling.

The middle panel shows the generalization gradients. Nothing in the model changed from training to testing. These gradients reflect the coupling strengths ( $W$ ) resulting from training. To illustrate, consider the curve for Type 1 trials (filled circles). The gradient at 1 s is close to zero because after 1-s samples a) the initial states ( $n = 1$  and 2) are the most likely to be active, and b) as the top panel shows, those states have close-to-zero couplings with Green. The generalization gradient reaches its maximum at 4 s because after 4-s samples the most active states are the intermediate states ( $n = 3$  to 6), which are strongly coupled with Green (see top panel). The response probability at 2 s occurs because the states most likely to be active after the 2-s samples, States 2 and 3, have weak and strong couplings with Green, respectively; the net effect is an intermediate response probability. Finally, the gradient decreases slightly after 4s because the couplings of later states ( $n > 6$ ), the states most likely to be active at the end of 8-s and 16-s samples, are not as strong as the coupling of the intermediate states.

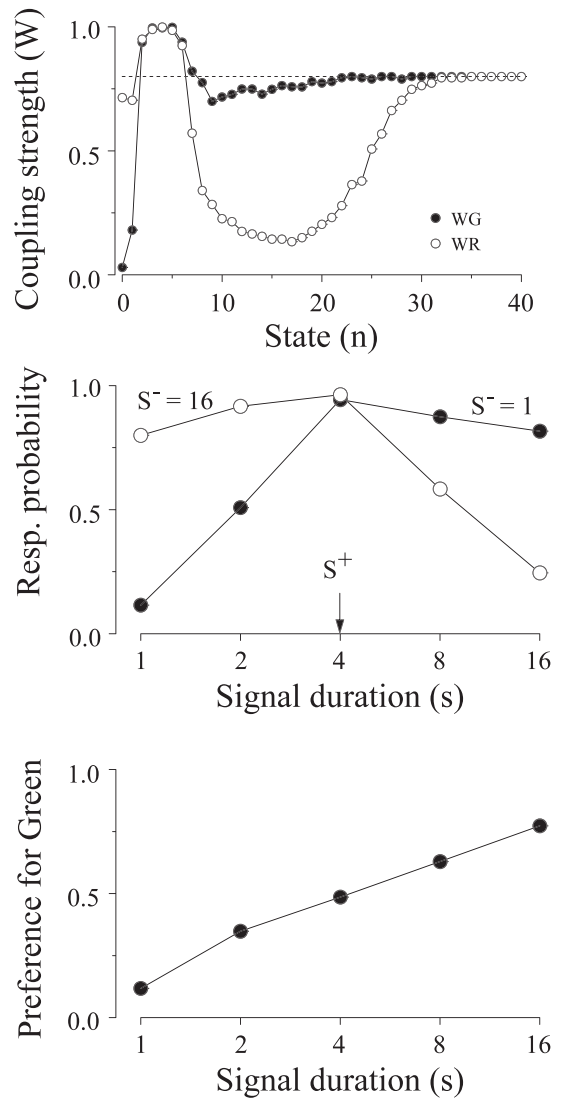


Fig. 7. Results from the simulation of the LeT model (Machado et al., 2009). Top panel. Strength of the couplings between the behavioral states and the operant responses, Green and Red, at the end of the Training Conditions of Phases 1 and 2 (see Table 2). Middle panel. Generalization gradients obtained during the Testing Conditions of Phases 1 and 2. Bottom panel. Preference for Green over Red obtained during the Testing Condition of Phase 3. Note the logarithmic scale on the x-axis in the middle and bottom panels.

The two generalization gradients are opposed and both display an area shift. The gradient for Green is low at 1 s and high at 4 s and longer samples. The gradient for Red is low at 16 s and high at 4 s and shorter samples. The gradients are similar to those produced by

<sup>2</sup>The contribution of the Pretesting Condition (partial reinforcement) to the coupling strengths ( $W$ ) is negligible, thus, when the animal advances to testing, the response probabilities yielded by  $WG$  and  $WR$  would basically be the same.

our pigeons except that the model predicts symmetric gradients that decrease for either very long (Green) or very short (Red) durations. If the couplings had been initialized at a lower value (e.g., .2 instead of .8) the decrease for the extreme durations would have been more pronounced (see also Boneau & Honig, 1964, Figure 2).<sup>3</sup>

On choice trials, the Green and Red keys occur together. The model assumes that  $p$ , the probability of choosing the Green key, is given by the relative value of the couplings, that is,  $p = WG(n^*)/[WG(n^*) + WR(n^*)]$ , where  $n^*$  is the active state at the end of the sample. The bottom panel of Figure 7 shows the predicted preference function: As the sample duration ranges from 1 s to 16 s, preference for Green increases monotonically. The model predicts the context effect and reproduces well the pigeons' average preference function (compare the bottom panels of Figures 5 and 7).

In summary, the LeT model reproduced the major trends in the data. The shape of the simulated generalization gradients was consistent with the pigeons' gradients. The major discrepancy was the symmetry of the gradients (predicted by LeT but not observed reliably in the pigeons' data) and the decrease in response strength for durations significantly away from the  $S^+$  and  $S^-$  (also predicted by LeT but rarely observed in the pigeons' data). With respect to the context effect, the model's preference function also was consistent with the average of the pigeons' preference functions. However, the individual differences among the pigeons are beyond the scope of the model.

The present study makes two significant contributions to our understanding of temporal control, one related to the LeT model and the other to the broader subject of stimulus control. Concerning LeT, in previous studies with the double bisection task we showed that the model could predict the context effect, but

we had not tested directly the model's account of the effect, namely, that it depends on the coupling profiles learned during training. Temporal generalization tests with the two operants, pecking Green and pecking Red, are one way of revealing these coupling profiles and thereby of testing one of the model's hidden components. Because in LeT choice proportions depend on the coupling profiles, the latter are more fundamental than the former. Putting to test the model's assumptions about the coupling profiles is a significant advance over testing only the model's predictions about choice proportions.

Concerning stimulus control, the present study followed a strategy that proved fruitful in other domains, namely, to try to explain seemingly complex phenomena on the basis of stimulus generalization gradients and their interactions (e.g., transposition and peak shift from excitatory and inhibitory gradients; see Honig, 1962, Riley, 1968, and Rilling, 1977). One added benefit of this strategy is to bring to the forefront of timing research the concept of temporal-generalization gradient and highlight how little we know about it, its attributes, and the variables that affect it.

## 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.
- Arantes, J., & Grace, R. C. (2008). Failure to obtain value enhancement by within-trial contrast in simultaneous and successive discriminations. *Learning and Behavior*, 36, 1–11.
- Arantes, J., & Machado, A. (2008). Context effects in a temporal discrimination task: Further tests of the Scalar Expectancy Theory and Learning-to-Time models. *Journal of the Experimental Analysis of Behavior*, 90, 33–51.
- Bizo, L. A., & McMahon, C. V. (2007). Temporal generalization and peak shift in humans. *Learning and Behavior*, 35, 123–130.
- Bloomfield, T. M. (1967). A peak shift on a line-tilt continuum. *Journal of the Experimental Analysis of Behavior*, 10, 361–366.
- Boneau, C. A., & Honig, W. K. (1964). Opposed generalization gradients based upon conditional discrimination training. *Journal of Experimental Psychology*, 68, 89–93.
- 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.

<sup>3</sup> A different decision rule could solve the problem (but see the gradient for Type 2 trials of pigeon P890 in Figure 3): Assume that  $WG(n)$  and  $WR(n)$  are not probabilities but response strengths and that a response occurs only if its strength is above a threshold,  $\Theta$ . If the initial values of  $WG$  and  $WB$  are all greater than  $\Theta$ , the model predicts gradients that do not decrease at the extreme sample durations. This improved decision rule costs one extra parameter, the threshold  $\Theta$ , and it will not be discussed further here.

- Cheng, K., Spetch, M. L., & Johnston, M. (1997). Spatial peak shift and generalization in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 23, 469–481.
- Church, R. M. (2003). A concise introduction to scalar timing theory. In W. H. Meck (Ed.), *Functional and neural mechanisms of interval timing* (pp. 3–22). Boca Raton, FL: CRC Press.
- Church, R. M., & Deluty, M. Z. (1977). Bisection of temporal intervals. *Journal of Experimental Psychology: Animal Behavior Processes*, 3, 216–228.
- Elsmore, T. F. (1971). Control of responding by stimulus duration. *Journal of the Experimental Analysis of Behavior*, 16, 81–87.
- 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.
- 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.
- Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. In L. Allan, & J. Gibbon (Eds.), *Timing and time perception* (pp. 52–77). New York: Annals of the New York Academy of Sciences.
- Hanson, H. M. (1959). Effects of discrimination training on stimulus generalization. *Journal of Experimental Psychology*, 58, 321–334.
- Honig, W. K. (1962). Prediction of preference, transposition, and transposition reversal from the generalization gradient. *Journal of Experimental Psychology*, 64, 239–248.
- Killeen, P. R., & Fetterman, J. G. (1988). A behavioral theory of timing. *Psychological Review*, 95, 274–285.
- 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., & 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., Malheiro, M. T., & Erhagen, W. (2009). Learning to time: a perspective. *Journal of the Experimental Analysis of Behavior*, 92, 423–458.
- 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.
- Mellgren, R. L., Mays, M. Z., & Haddad, N. F. (1983). Discrimination and generalization by rats of temporal stimuli lasting for minutes. *Learning and Motivation*, 14, 75–91.
- Oliveira, L., & Machado, A. (2008). The effect of sample duration and cue on a double temporal discrimination. *Learning and Motivation*, 39, 71–94.
- Oliveira, L., & Machado, A. (2009). Context effect in a temporal bisection task with the choice keys available during the sample. *Behavioural Processes*, 81, 286–292.
- Platt, J. R., & Davis, E. R. (1983). Bisection of temporal intervals by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 9, 160–170.
- Purcell, R. B. (1973). Peak shift: a review. *Psychological Bulletin*, 80, 408–421.
- Richelle, M., & Lejeune, H. (1980). *Time in animal behavior*. New York: Pergamon.
- Riley, D. A. (1968). *Discrimination learning*. Boston: Allyn & Bacon.
- Rilling, M. (1977). Stimulus control and inhibitory processes. In W. K. Honig, & J. E. R. Staddon (Eds.), *Handbook of operant behavior* (pp. 432–480). Englewood Cliffs, NJ: Prentice Hall.
- Russell, R., & Kirkpatrick, K. (2007). The role of temporal generalization in a temporal discrimination task. *Behavioural Processes*, 74, 115–125.
- Shettleworth, S. J. (1998). *Cognition, evolution, and behavior*. New York: Oxford University Press.
- Spence, K. W. (1937). The differential response of animals to stimuli varying within a single dimension. *Psychological Review*, 44, 430–444.
- Spetch, M. L., & Cheng, K. (1998). A step function in pigeons' temporal generalization in the peak shift task. *Animal Learning and Behavior*, 26, 103–118.
- Stubbs, D. A. (1968). The discrimination of stimulus duration by pigeons. *Journal of the Experimental Analysis of Behavior*, 11, 223–238.

Received: August 1, 2011

Final Acceptance: March 7, 2012