

## CODING IN PIGEONS: MULTIPLE-CODING VERSUS SINGLE-CODE/DEFAULT STRATEGIES

CARLOS PINTO AND ARMANDO MACHADO

UNIVERSITY OF MINHO

To investigate the coding strategies that pigeons may use in a temporal discrimination tasks, pigeons were trained on a matching-to-sample procedure with three sample durations (2s, 6s and 18s) and two comparisons (red and green hues). One comparison was correct following 2-s samples and the other was correct following both 6-s and 18-s samples. Tests were then run to contrast the predictions of two hypotheses concerning the pigeons' coding strategies, the multiple-coding and the single-code/default. According to the multiple-coding hypothesis, three response rules are acquired, one for each sample. According to the single-code/default hypothesis, only two response rules are acquired, one for the 2-s sample and a "default" rule for any other duration. In retention interval tests, pigeons preferred the "default" key, a result predicted by the single-code/default hypothesis. In no-sample tests, pigeons preferred the key associated with the 2-s sample, a result predicted by multiple-coding. Finally, in generalization tests, when the sample duration equaled 3.5s, the geometric mean of 2s and 6s, pigeons preferred the key associated with the 6-s and 18-s samples, a result predicted by the single-code/default hypothesis. The pattern of results suggests the need for models that take into account multiple sources of stimulus control.

*Key words:* timing, delayed matching-to-sample, many-to-one matching, pigeons

The ability to learn to behave according to a rule is often referred to as coding (e.g., prospective or retrospective; analogical or non-analogical; for a review, see Grant, Spetch, & Kelly, 1997), and this ability is believed to have been achieved through the evolution of general processes of learning focused on the antecedents and consequents of action (Skinner, 1984). One important goal of research in the field of learning is to understand why and how behavior accords to different rules in different circumstances.

To study coding, researchers have frequently used a delayed matching-to-sample task. In its simplest version, one of two stimuli (samples S1 and S2) is presented and then removed. Afterwards, two other stimuli (the comparisons, C1 and C2) are presented simultaneously. If the animal chooses C1 following S1, or C2 following S2, it receives a reward. Hence, learning the task may be conceived of as learning two conditional discriminations, "If S1, choose C1" and "If S2, choose C2". Of particular interest to test theories and models of coding is the effect on choice accuracy of introducing retention intervals between the samples and the comparisons.

When the samples are stimuli varying in duration (e.g., a light lasting 2s, S1, or 10s, S2, the retention functions following S1 and S2 typically diverge. That is, as the retention interval increases, choice following the short sample remains accurate (i.e., the animal continues to prefer C1 over C2), but choice following the long sample becomes increasingly inaccurate (i.e., choice of C2 decreases or, equivalently, choice of C1 increases with the retention interval). This result is known as the choose-short effect (Spetch & Wilkie, 1982; see also Spetch, 1987; Spetch & Wilkie, 1983).

One account of the choose-short effect is the coding model put forth by Kraemer, Mazmanian and Roberts (1985). The model makes four assumptions. First, the animal learns a specific response code for each sample (e.g., "If the light lasts 2s choose the red key"). Second, during the retention interval, the code is increasingly likely to be forgotten, perhaps because of stimulus interference during the interval and, in some cases, the similarity between the retention interval and the intertrial interval (see Dorrance, Kaiser, & Zentall, 2000; Kelly & Spetch, 2000; Sherburne, Zentall, & Kaiser, 1998). Third, the absence of a code is functionally equivalent to a no-sample or 0-s sample. And fourth, in the absence of a code at the moment of choice, the animal chooses the comparison associated with the sample closest to 0s.

Correspondence concerning this article should be addressed to Carlos Pinto, School of Psychology, University of Minho, Campus de Gualtar, 4710 Braga, Portugal (e-mail: carlos.arop@gmail.com).

doi: 10.1002/jeab.153

The model predicts that, as the retention interval increases, the code is increasingly likely to be lost, and therefore the animal is increasingly likely to prefer the comparison associated with the shortest sample (for alternative accounts of the choose-short effect, see, e.g., Gaitan & Wixted, 2000; Grant, 2009; Sherburne et al., 1998; Spetch & Wilkie, 1983). Because Kraemer et al.'s (1985) model assumes one code for each sample, it will be referred to as the multiple-coding hypothesis.

In more complex tasks, the multiple-coding strategy may not be the most economical or easiest to learn. Suppose that three samples, S1, S2, and S3, are mapped onto two comparisons, C1 and C2, such that C1 is correct following S1, and C2 is correct following both S2 and S3. In this many-to-one task, instead of learning three codes, the animal could learn a single code for S1 ("If S1, choose C1") and a default rule to be applied following any other sample or even no sample (i.e., "If not S1, choose C2"). This seemingly more economical strategy is known as the single-code/default.

To test the single-code/default hypothesis, Singer, Klein, and Zentall (2006, Experiment 2) rewarded pigeons for choosing comparison C1 following 8-s samples, and comparison C2 following 2-s and 32-s samples. After the pigeons learned the task, retention intervals ranging from 1 to 30 s separated the samples from the comparisons (see also Clement and Zentall, 2000; Gaitan & Wixted, 2000). To predict the retention functions, the authors made the following three assumptions. First, in this many-to-one mapping task pigeons learn a single code for 8-s samples and a default rule ("If 8 s, choose C1; otherwise, choose C2"). Second, during the retention interval, the sample code is increasingly likely to be forgotten. And third, in the absence of a sample code at the moment of choice, the pigeon behaves according to the default rule and chooses C2, the comparison associated with the 2-s and 32-s samples. Therefore, the single-code/default hypothesis predicted that, as the retention interval increased, choice following the 2-s and 32-s samples should remain accurate, but choice following the 8-s samples should become increasingly inaccurate. The results confirmed these predictions.

However, Singer et al.'s (2006) findings are not conclusive regarding the animal's coding strategy because they can also be explained by

the multiple-coding hypothesis. According to the latter, during training the pigeons learned a specific code for each of the three samples, "If 2 s, choose C2", "If 8 s, choose C1", and "If 32 s, choose C2". During the retention intervals, the codes were increasingly likely to be forgotten and, on those occasions, pigeons chose the comparison associated with the shortest sample, C2. Because C2 was correct following both 2-s and 32-s samples, the accuracy for those samples remained high and the accuracy for the 8-s sample decreased. Therefore, both the single-code/default and the multiple-coding hypotheses account for Singer et al.'s (2006) results.

The present experiment was designed to contrast the two coding hypotheses. To that end, pigeons were exposed to a many-to-one matching task with three samples and two comparisons. Specifically, pigeons learned to choose C1 following 2-s samples, and to choose C2 following 6-s and 18-s samples. With this mapping, the two hypotheses predict different retention function patterns.

According to the multiple-coding hypothesis, forgetting during the retention interval should lead the animal to act as if in a no-sample trial and consequently to choose C1, the comparison associated with the sample closest to 0 s. In terms of retention functions, correct responses following 2-s samples should remain high, whereas correct responses following 6-s and 18-s samples should decrease. The left panel of Figure 1 shows these predictions.

In contrast, if a single-code/default strategy is used, the pigeons should learn a single code for the 2-s samples and apply a default rule in the absence of the 2-s-sample code (i.e., "If 2 s, choose C1, otherwise choose C2"). Hence, as the retention interval increases, the pigeons are more likely to forget the sample code and consequently more likely to choose C2. In terms of retention functions, correct responses following 2-s samples should decrease, whereas correct responses following 6-s and 18-s samples should remain high. The right panel of Figure 1 shows these predictions. Note that, according to both hypotheses, the functions following the 6-s and 18-s samples should not differ.

Another goal of the present study was to test an assumption common to both accounts, namely, that if during the retention interval the memory for the sample (or its code) becomes unavailable, the animal behaves as if in a no-sample trial. To test this assumption, we included test trials without

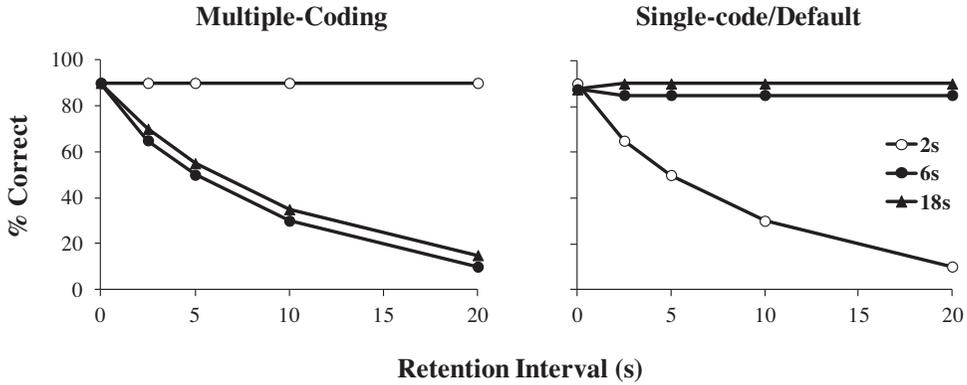


Fig. 1. Retention functions predicted by the multiple-coding hypothesis (left panel) and the single-code/default hypothesis (right panel) in a delayed matching-to-sample task with 2-s samples associated with one comparison and 6-s and 18-s samples associated with other comparison.

a sample—the comparisons were presented either immediately after the intertrial interval or after a “retention” interval (with no sample, a retention interval is defined procedurally as an interval similar to that included when the trial includes a sample). The two coding hypotheses predict different choice patterns on no-sample trials. Whereas the multiple-coding hypothesis predicts a preference for C1, the comparison associated with the shortest sample, the single-code/default hypothesis predicts a preference for C2, the comparison associated with the default rule. Moreover, the preference for C1 (multiple coding) or for C2 (single-code/default) should not vary with the retention interval.

Another reason to manipulate the retention interval on no-sample trials is that the resulting retention function may reveal sources of control over choice other than the sample duration. To illustrate, if choice on no-sample trials without a retention interval differs from the two model predictions, one would conclude that the offset of the intertrial interval also affects choice. Similarly, if choice on no-sample trials changes with the retention interval, then the stimulus conditions introduced by the retention interval (e.g., Dorrance et al., 2000; Grant, 2006), including the passage of time during the interval, also affect choice. More generally, knowledge of the retention function obtained on no-sample trials may help us to isolate the specific effects of the samples on choice.

The final goal of the present experiment was to obtain in this many-to-one task a psychometric function relating choice proportion to

sample duration. We were particularly interested in the bisection point and the overall shape of the psychometric function for their potential implications for the two coding hypotheses.

Consider the bisection point. Previous research has shown that, in one-to-one matching-to-sample tasks with two sample durations and two comparisons, the bisection point occurs at the geometric mean of the two sample durations (e.g., Catania, 1970; Church & Deluty, 1977; Stubbs, 1968, 1976). Although the present task is not one-to-one, it is conceivable that bisection will be determined solely by the 2-s samples (associated with C1) and the 6-s samples (associated with C2); the 18-s samples (also associated with C2) would not affect the bisection point. In this case, the pigeons should be indifferent between C1 and C2 at 3.5 s. But other factors may bias choice, and one of them is the default rule: If 3.5-s samples are sufficiently distinct from 2-s samples to activate the single code, the default rule will be triggered and choice following 3.5-s samples will be biased toward C2. If pigeons prefer C2 at the geometric mean of 2 s and 6 s, then they will be indifferent at durations shorter than the geometric mean.

Consider now the shape of the psychometric function. According to the multiple-code hypothesis, when a sample code is lost the animal chooses the comparison associated with the sample closest to 0 s, in our case, the 2-s sample. It follows that preference for C1 should be strong and roughly constant for samples in the range of 0 to 2 s. Moreover, if we make the

reasonable assumption that each code generalizes to sample durations around the trained sample duration, then the multiple code hypothesis predicts a typical psychometric function, with a bisection point at the geometric mean of 3.5 s. The results of the temporal generalization tests may help us decide among these different possibilities.

The two comparison keys can be labeled in two different ways: as a function of their associated sample duration (the correct key following 2-s samples [C1] is the "short" key and the correct key following 6-s and 18-s samples [C2] is the "long" key) or as a function of their meaning according to the single-code/default hypothesis (the correct key following 2-s samples [C1] is the "single-code" key and the correct key following 6-s and 18-s samples [C2] is the "default" key). These two nomenclatures will be used hereafter.

## Method

### Subjects

Six pigeons (*Columba livia*) maintained at approximately 80% of their free-feeding body weight served as subjects. Water and grit were freely available in their home cages. The pigeon room was maintained in a 13:11 hour light/dark cycle, with the lights on at 08:00, and its temperature was kept between 20°-22°C. The experiment was conducted once a day, 7 days a week, at approximately the same time of day for each pigeon.

Three of the pigeons (P463, P501, and P536) had previously participated on a timing experiment, although with comparison stimuli different from the ones used in this experiment, and the three other pigeons (P785, P917, P973) were experimentally naive.

### Apparatus

Six operant chambers were used: Five LVE (Lehigh Valley Electronics) chambers, and a homemade chamber. The LVE chambers measured 34 × 35 × 31 cm (h × l × w). Three circular response keys, 2.5 cm in diameter, were arranged horizontally on the response panel. The bottom edge of each key was 22.5 cm above the wire mesh floor, and the keys were 9 cm apart, center to center. Each key was equipped with a 12-stimulus IEE (Industrial Electronics Engineers) in-line projector. The food hopper

was accessible through a 6-cm wide × 5-cm high opening that was centered horizontally on the response panel, 8.5 cm above the floor. When the hopper was raised, a 28-V, 0.04-A light illuminated its opening and grain became accessible to the pigeon. On the wall opposite the response panel, 30 cm above the floor, a 28-V, 0.1-A houselight provided general illumination. The operant chamber was enclosed in an outer box equipped with an exhaust fan. The fan circulated air through the chamber and masked outside noises.

The homemade chamber measured 31 × 33 × 33 cm (h × l × w). Three circular response keys, 2.5 cm in diameter, were arranged horizontally on the response panel. The bottom edge of each key was 21 cm above the wire mesh floor, and the keys were 9 cm apart, center to center. Each key was equipped with a 12-stimulus IEE in-line projector. A LVE food hopper was accessible through a 6-cm wide × 4.5-cm high opening that was centered horizontally on the response panel, 6.5 cm above the floor. When the hopper was raised, a 28-V, 0.04-A light illuminated its opening and grain became accessible to the pigeon. On the wall opposite the response panel, 27.5 cm above the floor, a 28-V, 0.1-A houselight provided general illumination. The operant chamber was enclosed by a PVC sound attenuating cubicle (Med Associates, ENV-018V) equipped with an exhaust fan.

In this experiment, the side keys were illuminated with red or green hues and the central key was illuminated with a white hue. Personal computers using the ABET II (Lafayette Instrument Company) software controlled the experimental events and recorded the data.

### Procedure

**Training.** The birds were trained in a symbolic matching-to-sample task. Following the presentation of a white hue on the center key for 2, 6 or 18 s (sample stimulus), each of the side keys was illuminated with either a red or a green hue (comparison stimuli). One of the comparisons was correct following the 2-s sample, and the other comparison was correct following the 6-s and 18-s samples (the correct comparison for each sample was counterbalanced across pigeons). After a response, the comparison keys were turned off and, if the response was correct, reinforcement was delivered and the 30-s intertrial interval (ITI) started. If the response

was incorrect, the ITI started immediately. To avoid confusion between the ITI and the dark retention interval used in subsequent tests, the houselight was illuminated during the ITI but was turned off at sample onset. A correction procedure was in effect: Following an incorrect response, the trial was repeated; after three consecutive incorrect responses, only the correct comparison key was presented. The birds began each session at approximately 80% of their free-feeding weight and the reinforcement duration was adjusted for each bird to minimize feeding outside the experimental session. It varied from 1.5 to 4.5 s across animals.

Each session consisted of sixty-four trials (excluding correction trials), thirty-two 2-s sample trials, sixteen 6-s sample trials, and sixteen 18-s sample trials. Across trials, the location of the comparison stimuli varied pseudorandomly with the constraint that each comparison stimulus was presented the same number of times on each of the side keys. Training continued until the pigeon met a criterion of at least 80% correct responses to each sample in a session (excluding correction trials), for five consecutive sessions, or until 40 sessions were completed.

**Retention Test.** After the training phase, a retention interval was introduced between the sample and comparison stimuli. The retention interval—spent in darkness—could be 2.5, 5, 10 or 20 s long. Each session consisted of 80 trials, 48 regular training trials ( $24 \times 2$  s,  $12 \times 6$  s,  $12 \times 18$  s) and 32 retention-interval test trials ( $16 \times 2$  s,  $8 \times 6$  s,  $8 \times 18$  s). Irrespective of trial type, correct responses were reinforced, but the correction procedure was in effect only on regular trials (with no retention interval). To minimize feeding outside the experimental session, the reinforcement durations were recalculated and varied from 1 to 4 s across animals. Testing continued for 30 sessions.

**Retraining I.** The birds returned to a training phase until they made at least 80% correct responses to each sample (excluding correction trials) for 5 consecutive sessions or until 15 sessions were completed.

**No-sample Test.** In this test, no-sample trials were interspersed among the regular training trials. In a no-sample trial, the comparison keys were presented immediately after the ITI. Each session consisted of 72 trials, 64 regular training trials ( $32 \times 2$  s,  $16 \times 6$  s,  $16 \times 18$  s) and 8 no-sample test trials. Responses on no-sample trials

were never reinforced. To minimize feeding outside the experimental session, the reinforcement durations were recalculated and varied from 1.5 to 5 s across animals. Testing continued for 10 sessions.

**Retraining II.** This phase was exactly the same as Retraining I.

**Generalization Test.** Two ranges of untrained sample durations were presented. The first range included samples of 1, 3.5, 10.4 and 36 s. Two of these values (1s and 36s) were outside the training range. The other two (3.5 s and 10.4 s) were inside the training range and were equally discriminable from the adjacent training durations (3.5 s is the geometric mean of 2 s and 6 s, and 10.4 s is the geometric mean of 6 s and 18 s). The second range included samples of 3, 4, 5 and 10.4 s. The range spanned the interval with higher variability during the first test; the duration of 10.4 s was common to both test ranges and allowed a direct comparison between them.

Each session consisted of 96 trials, 56 regular training trials ( $28 \times 2$  s,  $14 \times 6$  s,  $14 \times 18$  s) and 40 generalization trials (10 trials for each test sample). The session was divided into six blocks of 16 trials each. The first block included only regular trials ( $8 \times 2$  s,  $4 \times 6$  s,  $4 \times 18$  s). The next five blocks included eight regular trials ( $4 \times 2$  s,  $2 \times 6$  s,  $2 \times 18$  s) and eight generalization trials (two trials per test sample). Responses following test samples were not reinforced. To minimize feeding outside the experimental session, the reinforcement durations on regular trials were recalculated and varied from 2 to 6 s across animals. Testing continued for 10 sessions, 5 for each range.

**Retraining III.** This phase was exactly the same as Retraining I except that the maximum number of session was reduced to 10.

**No-sample Retention Test.** This test was similar to the No-sample Test, with the exception that, on some of the no-sample trials, a retention interval of 2.5, 5, 10 or 20 s was introduced. Therefore, at the end of the ITI the houselight was turned off for the duration of the retention interval and then the comparison keys were illuminated. Each session consisted of 78 trials, 48 regular training trials ( $24 \times 2$  s,  $12 \times 6$  s,  $12 \times 18$  s) and 30 no-sample trials. Of the no-sample trials, 6 had no retention interval and 24 had a retention interval (6 trials for each of the 4 retention interval durations). No-sample trials were never reinforced. To

minimize feeding outside the experimental session, the reinforcement durations on regular trials were recalculated and varied from 2 to 6 s across animals. Testing lasted 10 sessions.

## Results

**Training and Retraining.** Five of the six pigeons met the learning criterion in 13 to 29 sessions (average of 20 sessions). The exception, P917, although failing to reach criterion in 40 sessions, acquired the discrimination (its matching accuracy during the last five training sessions equaled 78% for 2-s samples, 84% for 6-s samples and 99% for 18-s samples).

Due to an equipment malfunction, when switching to the Retention Test, three of the birds were exposed to sessions where the ITI was not illuminated. P463 ran one session in such conditions, P501 ran three sessions and P536 ran seven sessions. These birds returned to the training phase until the criterion was reached again or a total of 40 training sessions was completed. P463 ran five additional training sessions, P536 ran six additional sessions, and P501 failed to reach the criterion, having run 20 additional sessions (until the maximum of 40 sessions was reached). P501's failure to reach the criterion was due to the 6-s samples: Its matching accuracy for the last five training sessions equaled 88% for 2-s samples, 76% for 6-s samples and 99% for 18-s samples. Although failing to reach the criterion, P501 continued to next phase of the experiment.

The birds needed from 7 to 15 sessions (average of 12 sessions) to complete Retraining I, from 5 to 15 sessions (average of 8 sessions) to complete Retraining II and from 5 to 10 sessions (average of 6 sessions) to complete Retraining III.

**Retention Test.** In the Retention Test there were two types of trials: training trials with no retention interval, and retention-test trials. On both trials, correct responses were reinforced. Therefore, reinforcement on test trials could have changed performance during testing. To assess whether performance changed during testing, we analyzed separately the data from the first test session and from all 30 test sessions.

Figure 2 shows the results of the first test session. Matching accuracy on 2-s trials decreased abruptly with the shortest retention interval and then stabilized, whereas matching accuracy on 6-s and 18-s trials did not decrease as much and as

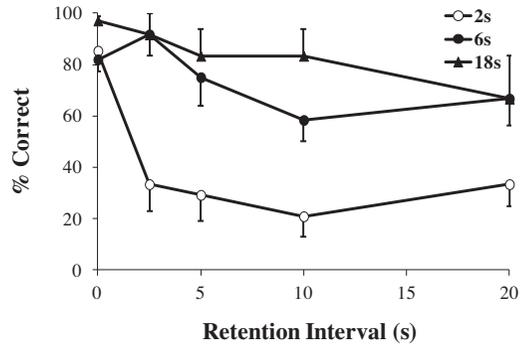


Fig. 2. Mean (with SEM) percent correct to each of the three sample durations as a function of retention interval duration. The data come from the first test session of the Retention Test.

abruptly. A two-way repeated-measures ANOVA with sample duration (three levels) and retention interval (five levels) as factors revealed a significant main effect of sample duration,  $F(2, 10) = 10.91$ ,  $MSE = .147$ ,  $p = .003$ , and of retention interval,  $F(4, 20) = 16.60$ ,  $MSE = .021$ ,  $p < .001$ . The interaction also was significant,  $F(8, 40) = 2.54$ ,  $MSE = .043$ ,  $p = .024$ , confirming that the retention interval did not affect matching accuracy equally following the three samples. This pattern of results is consistent with the single-code/default hypothesis.

Figure 2 also shows that percent correct following 6-s and 18-s samples decreased with retention interval (a repeated-measures ANOVA showed a significant effect of retention interval,  $F(4, 20) = 4.53$ ,  $MSE = .032$ ,  $p = .009$ ). Correct responses following 6-s samples seemed to be below correct responses following 18-s samples, but that difference was not significant:  $F(1, 5) = 1.29$ ,  $MSE = .109$ ,  $p = .301$ ). Moreover, the decrease in percent correct following the 2-s samples occurred abruptly from no retention interval to the 2.5-s retention interval, but it did not change with longer intervals (a repeated-measures ANOVA showed no significant effect of non-zero retention interval on 2-s samples,  $F(3, 15) = .909$ ,  $MSE = .023$ ,  $p = .460$ ). Finally, on retention trials, percent correct following 2-s samples was significantly below indifference (95% Confidence Interval = 20% - 38%).

Data from all test sessions were divided into six 5-session blocks. Figure 3 shows data from the first and the last blocks. The results from the first block were similar to the first session (compare the left panel of Fig. 3 with Fig. 2): A two-way

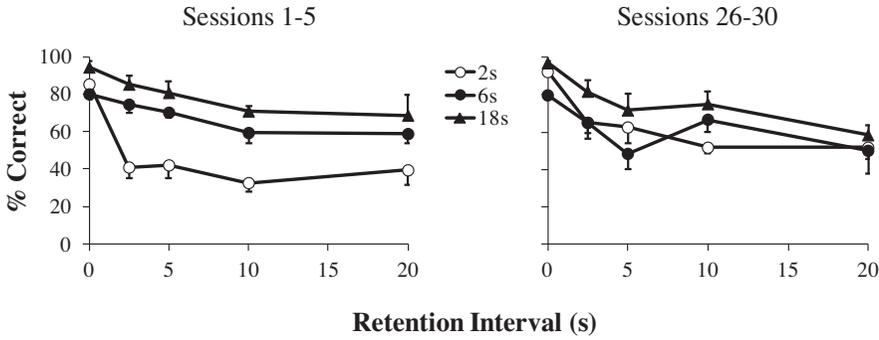


Fig. 3. Mean (with SEM) percent correct to each of the three sample durations as a function of retention interval in the first five sessions (left panel) and last five sessions (right panel) of the Retention Test.

repeated-measures ANOVA with sample duration (three levels) and retention interval (five levels) as factors revealed a significant interaction,  $F(8, 40) = 2.65$ ,  $MSE = .017$ ,  $p = .020$ , confirming that, during retention testing, matching accuracy to the three samples evolved differently. Similarly to the first session, percent correct following 6-s and 18-s samples decreased with retention interval (a repeated-measures ANOVA showed a significant effect of retention interval,  $F(4,20) = 9.98$ ,  $MSE = .013$ ,  $p < .001$ ). Additionally, matching accuracy to 6-s samples was lower than to 18-s samples (significant main effect of sample:  $F(1,5) = 13.29$ ,  $MSE = .014$ ,  $p = .015$ ). However, as the test progressed, the response pattern changed and the three curves approached each other (right panel, Fig. 3). By the end of testing, the curves for all three sample durations approached indifference as the retention interval increased.

To understand the evolution of responding during testing, Figure 4 shows percent correct on test trials (trials with a retention interval) across the six blocks, with sample duration as the parameter. As testing progressed, there was an increase in correct responses following the 2-s samples, and a slight decrease in correct responses following the 6-s. The net result was the maintenance of average percent correct across blocks (see dotted line). A two-way repeated-measures ANOVA with block (six levels) and sample duration (three levels) as factors confirmed this interpretation: No main effect of block was found,  $F(5, 25) = .794$ ,  $MSE = .021$ ,  $p = .564$ , but there was a significant main effect of sample duration,  $F(2, 10) = 9.48$ ,  $MSE = .177$ ,  $p = .005$ , and of the interaction,  $F(10, 50) = 2.95$ ,

$MSE = .028$ ,  $p = .005$ . Accuracy on test trials did not evolve similarly following each sample.

**No-sample Test.** Figure 5 shows the results for the No-sample Tests, with and without retention intervals. In tests with no (or 0-s) retention intervals, most choices were to the “short” key (%Long = 28% and 32% in the two phases, a nonsignificant difference,  $F(1, 5) = .47$ ,  $MSE = .012$ ,  $p = .525$ ). Moreover, a  $t$ -test showed that %Long differed significantly from 50%,  $t(5) = 5.420$ ,  $p = .003$ . Finally, at the individual level, a two-tailed normal approximation to the binomial showed that preference for “Short” was below chance in five of the six birds ( $\alpha = .05$ ). These results are consistent with the multiple-coding hypothesis.

The data from the No-sample Test with retention intervals show that, as the retention

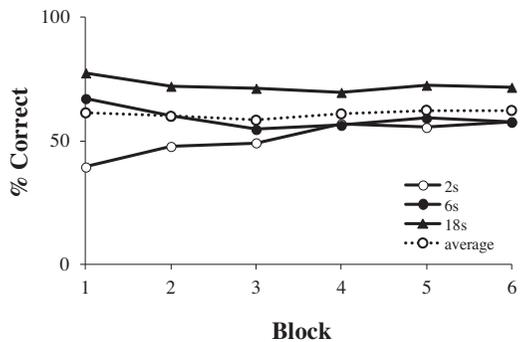


Fig. 4. Mean percent correct on test trials of the Retention Test to each of the three sample durations as a function of testing sessions (each block is composed of five sessions). The dotted line is the average of the three samples.

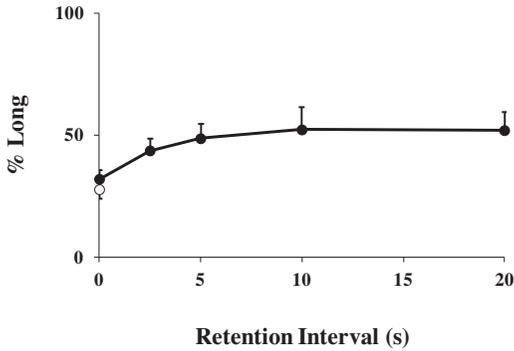


Fig. 5. Mean (with SEM) percent of choices to the "long" key (key associated with 6-s and 18-s samples) as a function of retention interval on the No-sample Test. The white dot is the result of the first No-sample Test.

interval increased, the percentage of "long" responses converged to chance. At the longest retention interval (20 s), overall preference for the "long" key had clearly increased: A two-tailed normal approximation to the binomial ( $\alpha=.05$ ) showed that, while two birds continued to prefer the "short" key, two preferred the "long" key, and two did not differ significantly from chance. Neither the multiple-coding nor the single-code/default hypotheses predicted this result.

**Generalization Test.** Two ranges of samples were presented. To test whether performance differed between the two ranges, performance on samples common to both ranges (2, 6, 10.4 and 18 s) was compared via a two-way repeated-measures ANOVA with test range (two levels) and sample duration (four levels) as factors. There were no significant effects of test range ( $F(1, 5) = .76$ ,  $MSE = .001$ ,  $p = .424$ ) or its interaction with sample duration ( $F(3, 15) = 1.37$ ,  $MSE = .001$ ,  $p = .290$ ). Therefore, we combined the results from the two ranges and averaged the data from the common samples.

Figure 6 shows the mean percent of "long" or "default" choices as a function of sample duration. The general pattern of the psychometric function was the one expected on the basis of temporal generalization: When a sample was presented, preference for the "long" key increased with sample duration according an ogive function. As for choice percentage following the 3.5-s samples (the geometric mean of 2 s and 6 s), the multiple-coding hypothesis predicted indifference, whereas the single-code/default hypothesis

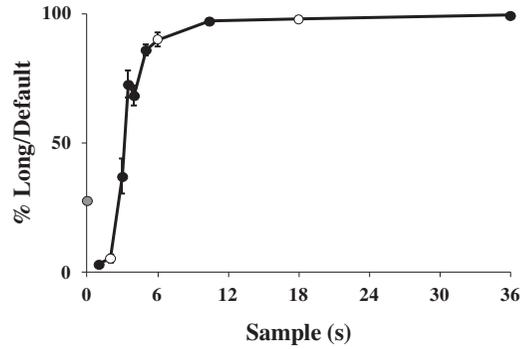


Fig. 6. Mean (with SEM) percent of choices to the "long" or "default" key (key associated with 6-s and 18-s samples) as a function of sample duration on the Generalization Test. The white dots identify the previously-trained durations (2 s, 6 s and 18 s) and the gray dot is the result of the first No-sample Test.

predicted a preference for the "default" key. The result, 73% choices for the "long" or "default" key, was consistent with the latter (a  $t$ -test showed that choices differed significantly from 50%,  $t(5) = 3.96$ ,  $p = .011$ ). At the individual level, a two-tailed normal approximation to the binomial showed that the preference for the "long" or "default" key was significantly above chance in four of the six birds.

Finally, it is also worth noting that preference following 0-s samples was significantly different from preference following the durations nearest to it: 1-s samples ( $F(1, 5) = 43.60$ ,  $MSE = .004$ ,  $p = .001$ ) and 2-s samples ( $F(1, 5) = 33.57$ ,  $MSE = .005$ ,  $p = .002$ ).

## Discussion

This study attempted to uncover the coding strategies adopted by pigeons in a matching-to-sample task where one comparison was correct after one sample and another comparison was correct following two samples. The pigeons could adopt one of two strategies, use three codes, one for each sample (multiple-coding hypothesis), or use two codes, one specific to the 2-s sample, and a default code triggered by any other sample (single-code/default hypothesis).

We examined the two hypotheses in the light of three types of tests, a retention-interval test, a no-sample test (with and without a retention interval), and a generalization test. Trials with a retention interval tested whether, when sample

information is lost, pigeons prefer the “short” key, as the multiple-coding hypothesis proposes, or the “default” key, as the single-code/default hypothesis proposes. Trials without a sample tested whether such trials are functionally equivalent to losing sample information during a retention interval, an assumption shared by both hypotheses. Generalization trials tested the predictions of each hypothesis concerning the location of the bisection point, at the geometric mean according to the multiple-code hypothesis, or below the geometric mean according to the single-code/default hypothesis.

The general pattern of results found in the Retention Test supported the single-code/default hypothesis: When retention intervals were introduced, the birds showed a preference for the “default” key. However, some results are hard to reconcile with the hypothesis. First, percent correct following the 6-s and 18-s samples decreased with retention interval (see Fig. 2 and left panel of Fig. 3). Since on these trials a 2-s sample was not presented, according to the hypothesis, the pigeons should have chosen the “default” key, both on trials with and on trials without a retention interval. Hence, percent correct following the 6-s and 18-s samples should not decrease, as it did. Second, percent correct on the 6-s samples was generally below percent correct on the 18-s samples. Since 6-s and 18-s samples share the same “default” response, the single-code/default hypothesis does not predict a difference between these two functions. Third, also not predicted by the single-code/default hypothesis was the abrupt decrease in accuracy on the 2-s sample trials following the shortest retention interval, accuracy that then remained relatively stable with longer intervals. A progressive decrease would be expected given that the retention interval should increase the probability of losing the sample code, and therefore increase the probability of choosing the “default” key. These discrepancies suggest that the single-code/default hypothesis needs to be elaborated with additional principles to account for the full range of effects of the retention interval.

Pinto and Machado (2011) suggested that multiple effects may be present in delayed matching-to-sample tasks. One of them is the disruption of timing and consequent random responding produced by the retention interval.

Specifically, the longer the retention interval, the more likely the disruption and the closer to indifference choice should be. This effect could explain the decrease in matching accuracy following the 6-s and 18-s samples (Fig. 2 and Fig. 3). Further evidence consistent with the effect can be seen in the No-sample Tests (Fig. 5): As the retention interval increased, choice also approached chance.

To further clarify the retention interval effect, Figure 7 re-plots the percentage of “long” choices as a function of sample duration (including no, or 0-s, samples), with the retention interval as a parameter. The data for the 2-s, 6-s, and 18-s samples come from the Retention Test trials, and the data for the 0-s samples come from the No-sample Test trials. The curve for trials without a retention interval (filled triangles) shows a preference for the “short” key following 2-s samples, and a preference for the “long” key following 6-s and 18-s samples. This result is expected because these were the choices the birds learned during training. However, when retention intervals were introduced (circles), the curve flattened and approached indifference. In fact, the longer the retention interval, the closer the curve came to indifference (contrast the empty and filled circles).

Another effect that could account for some results at odds with the single-code/default hypothesis is stimulus generalization. This effect may explain why performance on 18-s samples was generally better than on 6-s

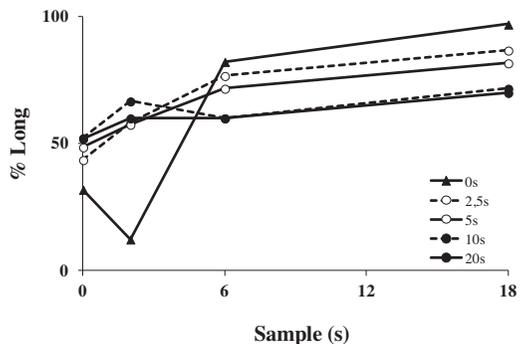


Fig. 7. Mean percent of choices to the “long” key (key associated with 6-s and 18-s samples) as a function of the sample duration presented. Each line refers to a retention interval. Data points for 2-s, 6-s, and 18-s samples come from the Retention Test and data points for 0-s samples come from the No-sample Test.

samples, even though they shared the same comparison key (see Fig. 3). Because the closer two stimuli are, the harder it is to discriminate between them, the discrimination between the 2-s and 6-s samples may have been more difficult than the discrimination between the 2-s and 18-s samples. Hence, percent correct following the 6-s samples was not as good as that following the 18-s samples.

Data from the Generalization Test trials lends further support to this hypothesis. The difference in correct choices between 6-s and 18-s samples (Fig. 6) suggests that the 6-s samples may have been coded as "short" slightly more often than the 18-s samples. If that was indeed the case, then it would follow that performance on the Retention Test trials would be slightly worse following the 6-s samples than following the 18-s samples. In fact, the average difference in correct choices between the 6-s and 18-s samples on the Generalization Test trials (Fig. 6) was of the same order (around 10%) than the average difference in performance between the 6-s and 18-s samples on the Retention Test trials (left panel of Fig. 3).

Finally, interpreting how matching accuracy following the 2-s samples changed with retention interval is challenging. According to the single-code/default hypothesis, we should expect a *gradual* decrease. The data showed an abrupt decrease, which did not seem to be affected by the duration of the retention interval (see Fig. 2 and left panel of Fig. 3). Perhaps generalization decrement due to stimulus changes brought about by the retention interval (dark period) rather than the duration of a retention interval causes the forgetting.

Yet another possibility to explain the pattern of responding following the 2-s samples combines two effects, a preference for the "default" key (as the single-code/default hypothesis predicts) and the aforementioned tendency for performance to approach chance with increasing retention intervals. The former effect would "push" the percent correct function towards 0%, while the latter would "pull" it towards 50%. Percent correct following 2-s samples would decrease until the two forces reached equilibrium and then it would stabilize around a value between 0% and 50%. The same two forces acting on performance following the 6-s and 18-s samples would "push" percent correct to 100% and "pull" it toward 50%. The

resulting force would maintain percent correct between these values, decreasing slightly because the pull to 50% would get stronger with the retention interval. In conclusion, the overall data set seems to require the integration of different effects.

The results of the No-sample Test were in the direction predicted by the multiple-coding model, that the pigeons would choose the comparison associated with the sample closest to 0 s. But if the preference for the "short" key following a 0-s sample were due to temporal generalization, then in the Generalization Test (Fig. 6), we would expect similar accuracies following the 0-s, 1-s, and 2-s samples. That was not the case. Following 1-s and 2-s samples, only 3% and 5%, respectively, of choices were to the "long" key, but following the 0-s samples, 28% of choices were to the "long" key. Although "short" remained the preferred key following the three samples, the preference following the 0-s sample was not as strong as expected from the generalization gradient alone. This result suggests a qualitative difference between zero and non-zero samples, perhaps the effect of generalization decrement due to stimulus changes rather than sample duration, an effect pushing performance toward indifference.

It is worth noting that a result based on generalization would not be incompatible with a single-code/default strategy, if we assume that a trial with a 0-s sample could be viewed as functionally similar to a trial with a 2-s sample. In that case, on 0-s trials, the birds would respond according to the "single-code" rule and choose the 2-s key.

Even though the no-sample test by itself may not be conclusive in telling us what coding strategy was in use, its results are informative in regard to an assumption shared by both models: The loss of sample information during a retention interval is functionally equivalent to a 0-s sample trial. If that were the case, we would expect similar preferences following a long retention interval and following a 0-s sample. The results were inconsistent with this prediction. In the Retention Test, the pigeons preferred the 6-s and 18-s key, but in the No-sample Test they preferred the 2-s key. Moreover, preference following the 0-s samples varied in an orderly fashion with the retention interval—as the interval increased, preference approached indifference. We conclude that, contrary to both models, the loss of sample

information during a retention interval is not equivalent to a 0-s sample.

A final piece of evidence consistent with the single-code/default hypothesis was the bisection point of the psychometric function obtained during the Generalization Test (Fig. 6). The hypothesis predicted the obtained result, a preference for “long” at the geometric mean of 2 s and 6 s or, equivalently, a bisection point slightly below the geometric mean. Hence, the default rule seems to determine choice following sample durations sufficiently away from 2 s.

In addition to the multiple-code and single-code/default, our results suggest a third strategy. The pigeons could have timed the interval from the end of the ITI to the beginning of the choice period, and then compared its duration with a threshold set between 2 s and 6 s; if the interval was below the threshold, they chose the “short” comparison; if above, they chose the “long” comparison. In this strategy, the effective time marker is not the onset of the center key light, the nominal time marker, but a more salient event, the offset of the houselight; and the effective sample is not the interval during which the center key is illuminated, the nominal sample, but the interval since the houselight was turned off.

This third strategy could explain the results of the Retention Tests. On those trials with both the nominal sample and a retention interval, the effective sample would almost always exceed the threshold, leading the pigeons to prefer the “long” comparison. Therefore, matching accuracy would not change with the retention interval following the 6-s and 18-s nominal samples, but it would decrease abruptly following the 2-s nominal samples—see Figure 2. The strategy could explain also the results of the Stimulus Generalization Tests. The sigmoid curve in Figure 6, with a steep slope at the indifference point, is consistent with a threshold-based account.

However, the strategy is hard to reconcile with the results of the no-sample tests. Pigeons should prefer the “short” comparison more strongly following 0-s nominal samples than 2-s nominal samples because the former should lead more than the second to effective samples below threshold. This result was not observed.

Similarly, if the pigeons followed the third strategy they should have preferred the “long” comparison when the 0-s nominal samples were followed by long retention intervals. This result also did not occur (e.g., in Fig. 5, pigeons were

indifferent between the comparisons at the 20-s retention interval).

Although this third strategy is plausible (see also Spetch & Rusak, 1989, 1992, for the effects of the ITI on matching-to-sample performance), it does not account for all of our main findings. Future work should explore the possibility that pigeons use multiple time markers (e.g., houselight offset, center keylight onset) and time multiple intervals.

In conclusion, of the two hypotheses put to test, the single-code/default provided more accurate predictions. However, neither hypothesis was consistent with the overall pattern of results. We argued that the pattern may result from a combination of different effects. Identifying the causal processes operating in matching-to-sample tasks and how these processes interact is fundamental to improving our understanding of how animals behave in environments where different coding strategies are possible.

### Acknowledgments

This work was supported by the Portuguese Foundation for Science and Technology (FCT) Grant (PTDC/MHC-PCN/3540/2012) to Armando Machado and an FCT Doctoral Grant (SFRH/BD/78566/2011) to Carlos Pinto.

### References

- 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, The theory of reinforcement schedules* (pp. 1–42). New York: Appleton-Century-Crofts.
- Church, R. M., & Deluty, M. Z. (1977). Bisection of temporal intervals. *Journal of Experimental Psychology: Animal Behavior Processes*, *3*, 216–228.
- Clement, T. S., & Zentall, T. R. (2000). Development of a single-code/default coding strategy in pigeons. *Psychological Science*, *11*, 261–264.
- Dorrance, B. R., Kaiser, D. H., & Zentall, T. R. (2000). Event-duration discrimination by pigeons: The choose-short effect may result from retention-test novelty. *Animal Learning & Behavior*, *28*, 344–353.
- Gaitan, S. C., & Wixted, J. T. (2000). The role of “nothing” in memory for event duration in pigeons. *Animal Learning & Behavior*, *28*, 147–161.
- Grant, D. S. (2006). Training delays reduce the choose-short effect with keylight duration samples in pigeons. *Behavioural Processes*, *72*, 139–148.
- Grant, D. S. (2009). Choose-long errors at delays shorter than the training delay persist when enhanced between-trial summation of duration memories are precluded in pigeons. *Behavioural Processes*, *82*, 335–339.

- Grant, D. S., Spetch, M. L., & Kelly, R. (1997). Pigeons' coding of event duration in delayed matching-to-sample. In C.M. Bradshaw C & E. Szabadi (Eds.), *Advances in psychology (Vol 120): Time and behavior: psychological and neuro-behavioral analyses* (pp. 217–264). Amsterdam: Elsevier / North Holland.
- Kelly, R., & Spetch, M. L. (2000). Choice biases in delayed matching-to-sample duration with pigeons: Manipulations of ITI and delay illumination. *The Quarterly Journal of Experimental Psychology*, *53B*, 309–323.
- Kraemer, P. J., Mazmanian, D. S., & Roberts, W. A. (1985). The choose-short effect in pigeon memory for stimulus duration: Subjective shortening versus coding models. *Animal Learning & Behavior*, *13*, 349–354.
- Pinto, C., & Machado, A. (2011). Short-term memory for temporal intervals: Contrasting explanations of the choose-short effect in pigeons. *Learning and Motivation*, *42*, 13–25.
- Sherburne, L. M., Zentall, T. R., & Kaiser, D. H. (1998). Timing in pigeons: The choose-short effect may result from pigeons' "confusion" between delay and inter-trial intervals. *Psychonomic Bulletin & Review*, *5*, 516–522.
- Singer, R. A., Klein, E. D., & Zentall, T. R. (2006). Use of a single-code/default strategy by pigeons to acquire duration sample discriminations. *Learning & Behavior*, *34* (4), 340–347.
- Skinner, B. F. (1984). The evolution of behavior. *Journal of the Experimental Analysis of Behavior*, *41*, 217–222.
- Spetch, M. (1987). Systematic errors in pigeons' memory for event duration: Interaction between training and test delay. *Animal Learning & Behavior*, *15*, 1–5.
- Spetch, M. L., & Rusak, B. (1989). Pigeons' memory for event duration: Intertrial interval and delay effects. *Animal Learning & Behavior*, *17*, 147–156.
- Spetch, M. L., & Rusak, B. (1992). Temporal context effects in pigeons' memory for event duration. *Learning and Motivation*, *23*, 117–144.
- Spetch, M. L., & Wilkie, D. M. (1982). A systematic bias in pigeons' memory for food and light durations. *Behavior Analysis Letters*, *2*, 267–274.
- Spetch, M. L., & Wilkie, D. M. (1983). Subjective shortening: A model of pigeons' memory for event duration. *Journal of Experimental Psychology: Animal Behavior Processes*, *9*, 14–30.
- Stubbs, A. (1968). The discrimination of stimulus duration by pigeons. *Journal of the Experimental Analysis of Behavior*, *11*, 223–238.
- Stubbs, A. (1976). Scaling of stimulus duration by pigeons. *Journal of the Experimental Analysis of Behavior*, *26*, 15–25.

Received: January 30, 2015  
 Final Acceptance: March 25, 2015