

**Number-of-responses matching in pigeons (*Columba livia*): choice biases following delay and no-sample tests**

Carlos Pinto\* and Márcio Mota

University of Minho

\* corresponding author

Address: University of Minho, School of Psychology, Gualtar Campus, 4710-057

Braga, Portugal.

E-mail address of corresponding author: cpinto@psi.uminho.pt

Abstract

In a symbolic matching-to-sample task, pigeons learned to discriminate between 5 and 15 key pecks (samples): different choices were correct following the smaller and the larger response requirements. Subsequently, accuracy was tested in delayed matching, with the delay spent in darkness, contrarily to previous studies, that used illuminated delays. On average, delayed choices reflected indifference between the choices, but individual analyses showed different biases, replicating previous findings. It has been suggested that the end result of a delay may be similar to presenting no sample to begin with, so we compared preferences following a delay and following trials where no pecks were required. Performance in the two situations differed and, on zero-peck trials, a bias towards the “small” choice was found. Finally, to assess if the “small” bias was due to stimulus generalization, we compared zero-peck trials and trials with small response requirements (ranging from one to

four) and found a discontinuity between zero and non-zero samples that may seem to be at odds with a generalization account.

Keywords: Discrimination learning, Symbolic matching-to-sample, Fixed-ratio schedule, Delay, Choice bias, Pigeon

## 1. Introduction

The delayed matching-to-sample task has been used to study a variety of processes, such as stimulus control, the coding of response rules, memory, or choice (see, for a review, Zentall and Smith, 2016). In this task, a trial initiates with the presentation of a sample stimulus that is then followed, after a delay, by a choice between (at least) two comparison stimuli. The comparison that should be chosen is contingent on the sample presented: for instance, if the samples are stimuli that differ in duration (say, a keylight in the operant chamber that is on for 2s or 10s), choices of one comparison (say, a green keylight) are only correct following the 2-s sample, and choices of the other comparison (say, a red keylight) are only correct following the 10-s sample.

When the task involves a temporal discrimination – like the previous example – choice biases may occur: in some cases, following the delay there is a tendency for choosing the comparison associated with the shortest sample, leading to a marked reduction in matching accuracy following long samples, but not following short samples. This result is known as the choose-short effect (e.g., Kelly & Spetch, 2000; Kraemer et al., 1985; Pinto & Machado, 2011; Spetch, 1987; Spetch & Grant, 1993;

Spetch & Wilkie, 1982). A similar bias is found on no-sample tests: when no sample stimulus is presented before the comparisons, there is a preference for the comparison associated with the shortest sample (e.g., Church, 1980; Pinto et al., 2017; Pinto & Machado, 2015, 2017; Pinto & Sousa, 2021; Spetch & Wilkie, 1983). The bias for short found in both delay and no-sample tests has led to hypothesizing that the end result of a delay following sample presentation is similar to presenting no sample to begin with (Kraemer et al., 1985; Sherburne, Zentall, & Kaiser, 1998; Spetch & Wilkie, 1983).

However, delay testing has also produced cases with no bias at all, or even with the opposite bias (that is, a bias for the comparison associated with the longest sample – a choose-long effect). In studies where a choose-short effect was found, the intertrial interval (ITI) and the delay were generally signaled in the same fashion: both were spent in darkness. This similarity could lead to instructional ambiguity (Fetterman & MacEwen, 1989; Zentall, 1997, 2006) – the intervals could be confused and if a delay was treated as an ITI, a trial with a delay could be seen as a no-sample trial (given that it would appear that the trial consisted solely of the presentation of the comparisons). To control for this potential confound, delay and ITI have been signaled differently: in some cases, this manipulation removed choice biases, with both samples showing similar retention functions (Sherburne et al., 1998; Spetch & Rusak, 1992), but in other cases a bias for short persisted (Kelly & Spetch, 2000; Pinto & Machado, 2011; Pinto & Sousa, 2021).

Finally, a choose-long effect has typically been found when, instead of filled intervals (where a stimulus is presented for the full duration to be timed), empty intervals (where no stimulus is presented during the to-be-timed duration, which is

signaled instead by “start” and “stop” time markers) are employed (Grant, 2001; Grant & Talarico, 2004; Santi et al., 2003; Santi et al., 1999). It is worth noting that a choose-long bias has also been found with filled intervals, in tasks where the ITI was houselight-illuminated and the sample duration was signaled by a keylight in a darkened chamber (Pinto & Machado, 2015, 2017). However, in these studies the animals appeared, in some trials, to make use of the ITI illumination as a “start” time marker – ignoring the nominal sample – and thus it could be considered that they were timing an empty interval (between the termination of the ITI and the onset of the comparisons). In this case, in delay testing, the animals would add sample and delay durations (as they would both fall within the empty interval being timed), thus yielding a choose-long bias. In conclusion, the variety of results in delay testing underlines that the effect of a delay depends significantly on task characteristics, namely how different intervals are signaled.

Tasks where animals must discriminate between numbers of responses (e.g., Fetterman, 1993; Rilling, 1967) are closely related to the temporal discriminations described thus far: given that both stimulus dimensions are extended over time, a sample cannot be identified immediately, requiring integration over time. Additionally, number and duration tend to correlate: a high number of responses usually takes longer to complete than a low number of responses. It has been considered that timing and counting may share a common mechanism, and timing models have accommodated both types of discriminations (Meck et al., 1985; Roberts, 1997). Despite these similarities, numerosity discriminations have not been as widely studied as timing, but choice biases have also been found in tasks where animals must discriminate between numbers of responses.

For instance, Maki, Moe, and Bierley (1977, Experiment 3) used a matching-to-sample task where one comparison was correct following 10 pecks – Fixed Ratio (FR) 10 – and the other comparison following 20 pecks – FR 20. Following training they introduced a delay between sample and comparisons that could be either illuminated by the houselight or spent in darkness (the ITI was always dark). The delay was progressively increased, and matching accuracy decreased with delay, more so when the houselight was on. Interestingly, when the delay was illuminated (that is, when the ITI and delay were signaled differently) there was a tendency for choosing the comparison associated with the larger FR, a choose-large effect. When both ITI and delay were spent in darkness there was no choice bias. A similar procedure was employed by Fetterman and MacEwen (1989), who trained pigeons to discriminate between FR 10 and FR 40 before introducing delays, which could be houselight-illuminated or dark. However, in this case not only did the animals show a bias towards the comparison associated with the smaller FR (a choose-small effect), but this effect was stronger when the delay was spent in darkness (the chamber was also dark during the ITI). One difference that could explain the disparity in results between these studies is that, contrarily to Maki, Moe, and Bierley (1977), where a single delay was used in a given session, for several sessions in a row, Fetterman and MacEwen (1989) intermixed different delays in every test session: Extended training with a given delay has been found to affect choice (Spetch & Wilkie, 1983).

Fetterman and MacEwen (2003) replicated their findings in a task where samples could differ in hue (red or green) or FR requirement (FR 5 or FR 20): for a group of birds hue was relevant in the trained discrimination (and FR was irrelevant), for another group the opposite was true, and for a third group both hue and FR were

relevant. For the groups where FR was relevant, a choose-small effect was found in delay testing (both ITI and delays were spent in darkness).

The results of these FR discriminations indicate, once more, that performance on delay testing seems to depend on how the components of the task are signaled: when ITI and delay were similar, either no bias or a choose-small bias occurred; when they were different, both choose-small and choose-large biases were found, as well as no bias at all. It is worth noting that when ITI and delay were differentiated (Fetterman & MacEwen, 1989; Maki et al., 1977), it was by keeping the ITI dark and illuminating the delay. During a delay, presenting stimuli (Zentall, 1973) or illuminating the chamber (D'Amato & O'Neill, 1971; Etkin, 1972; Grant & Roberts, 1976) has been found to reduce matching accuracy in visual matching tasks. In discriminations based on the number of responses, Maki, Moe, and Bierley (1977, Experiment 4) found that houselight illumination decreased accuracy, and that the longer the period of illumination, the stronger this effect. Given that delay illumination may disrupt matching performance, the first goal of the present experiment was to run a delay test following a FR discrimination (FR 5 vs. FR 15), where the ITI and delay were differentiated, but with houselight-illuminated ITIs and delays spent in darkness. The disparity in results in previous FR studies (choose-small, choose-large, or no biases at all have been found) could perhaps be due, at least in part, to the effects of delay illumination, so the present experiment could help clarify this issue.

As aforementioned, in timing tasks, a bias towards the “short” comparison is commonly found in no-sample tests, and no-sample test performance has been equated to delay performance. In Fetterman and MacEwen (1989)’s FR

discrimination experiment, a bias for the comparison associated with the smaller ratio was found in no-sample tests – similarly to what is typically found in timing tasks with no response requirement. A second goal of this study was to replicate the no-sample tests, but to allow a better comparison between choices in no-sample and delay trials, both tests were run in the same daily sessions.

The bias for the “short/small” comparison on no-sample tests has been explained by stimulus generalization – the process by which stimuli similar along one dimension evoke similar responses. According to this view, if no-sample trials are treated as 0-pecks/0-s samples (and this sample is part of the same continuum as non-zero samples), the “short/small” comparison should be preferred because it is the one numerically closest to zero. Finally, we obtained a generalization gradient with new FR requirements as samples (1, 2, 3, 4, 9, 25), to assess whether performance on no-sample tests would be similar to the smaller FR requirements, in accord with a generalization account.

## 2. Method

*2.1. Subjects.* The subjects were 8 pigeons (*Columba livia*), maintained at 85% of their free-feeding body weight. The birds were housed individually, with water and grit freely available. In the colony room, a 13:11-h light/dark cycle was maintained, with lights on at 08:00, and the temperature was kept between 20 and 22° C. The experiment was conducted once a day, at approximately the same time for each pigeon, six days a week. All pigeons had experience with temporal discriminations in matching-to-sample tasks. Additionally, two of the birds (P918 and P449) had already run numerical discrimination tasks.

2.2. *Apparatus.* Four operant chambers were used: three LVE (Lehigh Valley Electronics) chambers, and one homemade chamber. The LVE chambers measured 34 x 35 x 31 cm (height x length x width), and were equipped with three circular response keys, arranged horizontally, 2.5 cm in diameter, and 9 cm apart, center to center. These keys were localized 22.5 cm above the wire mesh floor and each was equipped with a 12-stimulus IEE (Industrial Electronics Engineers) in-line projector (each stimulus was illuminated with a 28-V, 0.1-A lamp). The reinforcement, mixed grain, was delivered through a LVE (Lehigh Valley Electronics) food hopper, with a 6 x 5 cm (w x h) opening, centered horizontally on the response panel, 8.5 cm above the floor. When the hopper was raised to allow access to food, a 28-V, 0.04-A light illuminated the opening. A 28-V, 0.1-A houselight, on the wall opposite the response panel, provided general illumination to the chamber. An exhaust fan that circulated air through the chamber and masked outside noises.

The homemade chamber measured 31 x 33 x 33 cm (h x l x w). The diameter of the three response keys was 2.5 cm, and they were 9 cm apart (center to center). The bottom edge of the keys was 21 cm above the wire mesh floor. In each key, stimuli were presented through a 12-stimulus IEE in-line projector – a 28-V, 0.1-A lamp illuminated each stimulus. The opening of the LVE food hopper, centered horizontally on the response panel, 6.5 cm above the floor, was 6-cm wide x 4.5-cm high. When the hopper was activated, a 28-V, 0.04-A light illuminated its opening. On the wall opposite the intelligence panel, 27.5 cm above the floor, there was a 28-V, 0.1-A houselight. The operant chamber was enclosed by a PVC sound-attenuating cubicle (Med Associates, ENV-018V) equipped with an exhaust fan. On the left and right keys red or green hues were presented, and on the center key only a white hue



was presented. Computers running the ABET II software (Lafayette Instrument Company) controlled the experiment and recorded the data.

### 2.3. Procedure

*2.3.1. Training.* The pigeons began by learning a symbolic matching-to-sample task. Every session started with a 30-s inter-trial interval (ITI) illuminated by the houselight. At trial onset, the houselight was turned off, and the center key was illuminated with a white hue. The pigeons then would have to peck five (Fixed Ratio of five pecks – FR 5) or 15 times (Fixed Ratio of 15 pecks – FR 15) on the illuminated key. When the pecking requirement was met, the center-key light was turned off and the two side keys were illuminated, one with a green hue and the other with a red hue (comparisons). Choice of one comparison was correct following the FR 5 and the other following the FR 15. After a peck to either comparison, the side key lights were turned off and two outcomes could follow. If the choice was correct, reinforcement was delivered and, after that, the ITI started. If the choice was incorrect, the ITI started immediately. A correction procedure was in effect: in case of an incorrect response, the trial was repeated up to a maximum of three times; at the fourth repetition, only the correct comparison was presented. Reinforcement duration was adjusted individually to minimize feeding outside the experimental task and varied between 1.5 and 4 s.

Without considering trials repeated due to the correction procedure, each session comprised 64 trials divided in two blocks of 32 trials each. On each block,

there were 16 trials (randomly distributed) of each FR sample. The position of the comparisons was counterbalanced within each block, so that each comparison was presented the same number of times in each side key. Dividing the session in two blocks ensured a more balanced distribution of trials throughout the session. The correct comparison following each FR requirement was counterbalanced across subjects.

Training lasted a minimum of 15 sessions, and continued until the pigeon met a criterion of at least 80% correct responses to each sample in a session (excluding correction trials), for three consecutive sessions, or until a maximum of 35 sessions was completed.

*2.3.2. No-sample & Delay Test.* Each session included 80 trials, 56 training trials (28 FR 5 and 28 FR 15 trials, identical to the Training phase), 16 delay test trials and 8 no-sample test trials. A delay test trial began similarly to a training trial, but following completion of the FR requirement, a delay (lasting 2.5, 5, 10 or 20 s) was introduced before comparison presentation. The delay was spent in darkness. Reinforcement on delay trials was delivered pseudo-randomly, on half of the FR 5 and half of the FR 15 trials. On no-sample test trials, the comparisons were presented immediately following the 30-s ITI. That is, the white keylight was not presented on the center key, and the birds were not required to peck. Reinforcement on no-sample test trials was non-differential, with half of the trials ending with reinforcement.

The trials could be presented in three types of blocks: In a training block, there were eight training trials (four FR 5 and four FR 15). In a delay block, there were 24 training trials (12 FR 5 and 12 FR 15) and 16 delay trials. Of the 16 delay trials, eight began with an FR 5 (two trials per delay) and eight began with an FR 15 (two trials

per delay). Finally, a no-sample block comprised 16 training trials (eight FR 5 and eight FR 15) and eight no-sample trials. Within each block trial order was randomized. In a daily session there were four blocks, the training block was presented twice and the delay and no-sample blocks were presented once each. The first and the third blocks were always training blocks, so that the block order could either be training/delay/training/no-sample or training/no-sample/training/delay. Testing lasted 10 sessions, half of the sessions with each block order. Block sequences were alternated across days, with half the pigeons starting with one block order and the other half with the other block order.

*2.3.3. Generalization Test.* Each session included 80 trials: 56 training (28 FR 5 and 28 FR 15) and 24 generalization trials. A generalization trial was identical to a training trial, with the exception that the FR requirement could be 1, 2, 3, 4, 9, or 25 pecks. The FR parameters below the trained FR 5 allowed us to evaluate choice behaviour for values close to zero; the FR 9 was chosen because it is approximately the geometric mean between the two trained FRs, so it should be equally discriminable from both (Weber's law). The FR 25 was chosen so that the ratio between the FR15 and FR25 was the same as between FR9 and FR15. Half of the generalization trials were reinforced, randomly. The trials were presented in two blocks of 40, each block containing 28 training (14 FR 5 and 14 FR 15) and 12 generalization trials (two per FR), randomly distributed.

*2.4. Statistics.* We analyzed the time taken to complete each FR requirement and also choice behavior between comparison stimuli. Given that these variables met the required assumptions, parametric tests were conducted in all our statistical analyses – using IBM SPSS Statistics for Windows, Version 28 – with the Type-1 error rate set

at 0.05. Time to complete the FR schedules was separated between trials that ended with correct and incorrect choices – these two groups of data were compared via two-tailed, paired-samples t-tests. Regarding delay testing, a repeated-measures ANOVA (with sample and delay duration as factors) was performed to assess the effect of delay on choice. In cases where the Mauchly's sphericity test was significant, Greenhouse-Geisser corrections were applied. 95% Confidence Intervals were calculated for choice following delays, to allow comparisons with chance levels. On no-sample tests, choice performance was contrasted to chance levels via a two-tailed, one-sample t-test. Performance following no-sample and long-sample trials was compared via a two-tailed, paired-samples t-test. For the Generalization Test, FR 9 was compared with chance levels via a two-tailed, one-sample t test, and the smaller FR schedules (FR 1 to FR 5) were compared via a repeated-measures ANOVA (with FR requirement as factor). Finally, small-FR and no-sample performance was compared via a two-tailed, paired-samples t-test. For measures of effect size, for t tests the standardized mean difference effect size was used ( $d_z$ , see, e.g., Cohen, 1988, p. 48; Lakens, 2013), and for ANOVA the generalized eta square was used ( $\eta_G^2$ , e.g., Bakeman, 2005; Olejnik & Algina, 2003).

### 3. Results

*3.1. Training.* Seven birds met the learning criterion, taking an average of 20.3 sessions (range: 15 – 29,  $SD = 4.9$ ). The eighth bird, PG47, failed to meet the learning criterion after the maximum of 30 sessions and did not progress to testing. By the last three sessions of training, average matching accuracy was .91 (range: .84 – .97,  $SD = .04$ ) for the FR 5 and .89 (range: .82 – .95,  $SD = .04$ ) for the FR 15. In those same

three sessions, the average time the pigeons took to meet the FR parameters was 1.11s (range: 0.64 – 1.51,  $SD = 0.26$ ) for the FR 5 and 3.97s (range: 2.18 – 5.06,  $SD = 0.96$ ) for the FR 15.

Given that a FR 5 took less time to complete than a FR15, the pigeons could have learned the task, not by counting the number of pecks, but by timing how long they were pecking or how long the center keylight was on. One way to attempt to disentangle number from time is to do an error analysis: if animals were timing, they would have learned to associate the “small” comparison with shorter durations, so it would be expected that, among the FR 5 trials, errors would be more likely to happen when the animals took longer to complete the FR (time  $FR5_{\text{incorrect}} > \text{time } FR5_{\text{correct}}$ ). Following the same reasoning, it would be expected that the FR 15 trials that ended in incorrect responses would be the ones where the FR was completed fastest (time  $FR15_{\text{correct}} > \text{time } FR15_{\text{incorrect}}$ ). At the group level, pigeons took, on average, 1.10s (range: 0.65 – 1.50,  $SD = 0.25$ ) to complete the FR 5 on trials in which a correct choice was made and 1.27s (range: 0.59 – 1.74,  $SD = 0.36$ ) on trials in which an incorrect choice was made. A paired-samples t-test confirmed that this difference was statistically significant,  $t(6) = 2.50$ ,  $p = .023$ ,  $d_z = 0.94$ . On the other hand, on FR 15 trials, pigeons took, on average, 3.99s (range: 2.19 – 5.13,  $SD = 0.96$ ) to complete the FR on trials in which a correct choice was made, and 3.89s (range: 2.08 – 4.83,  $SD = 1.01$ ) on trials in which an incorrect choice was made, and a paired-samples t-test revealed that this difference was not statistically significant,  $t(6) = 0.83$ ,  $p = .218$ ,  $d_z = 0.08$ . At the individual level, only three of the seven pigeons showed the aforementioned pattern (time  $FR5_{\text{incorrect}} > \text{time } FR5_{\text{correct}}$  and time  $FR15_{\text{correct}} > \text{time } FR15_{\text{incorrect}}$ ).

3.2. *No-sample & Delay Test.* Figure 1 presents the group results of the no-sample and delay tests: the data are presented as proportion of “small” choices, that is, choices of the comparison that was correct following the FR 5 in training. With the introduction of delays, matching accuracy decreased, approaching chance (dashed horizontal line): choice proportion following delays did not differ significantly from 0.50 for both samples; for FR 5, 95% Confidence Interval (CI) [0.43, 0.51], for FR 15, 95% CI [0.43, 0.52]. Moreover, delay performance was similar for both samples and did not vary much with delay duration as can be observed by the overlap of the data points. A repeated-measures ANOVA with sample (two levels) and non-zero delays (four levels) as factors confirmed that preferences following the two FR schedules did not differ and that the delay duration did not affect choices: there was no main effect of sample,  $F(1, 6) = 0.04$ ,  $p = .847$ ,  $\eta_G^2 = .0003$ , delay,  $F(1.95, 11.69) = 0.68$ ,  $p = .520$ ,  $\eta_G^2 = .015$ , nor factor interaction,  $F(3, 18) = 1.19$ ,  $p = .343$ ,  $\eta_G^2 = .009$ . That is, irrespective of the sample presented, when a trial involved a delay (of any duration) the birds appeared to be indifferent between the comparisons.

On no-sample tests (rightmost data point in Figure 1), average proportion of “small” choices was .61 (range: .50 – .75,  $SD = .10$ ), a value shown by a one sample t-test to be significantly different from chance (0.50):  $t(6) = 2.88$ ,  $p = .028$ ,  $d_z = 1.09$ . A result that illustrates the extent of this preference for “small” is that, out of the 70 test sessions (10 sessions x 7 birds), in only two did a pigeon choose “large” more often than “small”. In the other direction, choices of “small” were more frequent than “large” in 36 sessions. One of the goals of this study was to compare choices on no-sample trials with choices following long (20 s) delays, and a paired-samples t-test showed that performance differed between these two situations:  $t(6) = 2.99$ ,  $p = .024$ ,

$d_z = 1.13$  (for this analysis, FR 5 and FR 15 trials were averaged, given that choices did not differ between the two samples at the 20-s delay; paired-samples t-test:  $t(6) = 0.42$ ,  $p = .689$ ,  $d_z = 0.16$ ). Despite the difference in choices following long delays and on no-sample trials, they appeared to be correlated, but this relation was non-significant,  $r(5) = .60$ ,  $p = .153$ .

Figure 2 depicts the individual results of the no-sample and delay tests. The individual data show that, on trials with a delay, the FR 5 and FR 15 functions approached superimposition and tended to be flat, not changing much with delay duration. In other words, performance following a delay did not appear to be affected by the sample presented or the delay duration. While the average plot (Figure 1) showed the functions superimposing around chance level, Figure 2 reveals that that average is a mixture of patterns. Two of the birds (P449 and P724, bottom row of Figure 2) chose “small” significantly below chance: combining both samples, the 95% CI for “small” choices on delay trials was [0.14, 0.27] for P449 and [0.29, 0.44] for P724. Two birds (P918 and PG39, second row from the bottom) showed the opposite result, choices of “small” significantly above chance: for P918 95% CI [0.52, 0.67], and for PG39 95% CI [0.58, 0.73]. The remaining three birds (PG30, PG40, PG29, top two rows) did not differ significantly from 0.50: for PG30 95% CI [0.37, 0.53], for PG40 95% CI [0.41, 0.56], and for PG29 95% CI [0.47, 0.63].

*3.3. Generalization Test.* A typical generalization gradient was obtained (Figure 3): choices following the new, smaller, FRs (FR 1 to FR 4) were similar to those following the FR 5, and following the new, larger, FR (FR 25) were similar to those following the FR 15. Following the FR 9, situated near the geometric mean of the two trained FR requirements (8.7), there was indifference between the

comparisons: choices were shown to be not different from chance levels by a one-sample t-test,  $t(6) = 0.50$ ,  $p = .637$ ,  $d_z = 0.16$ . Of special interest to us was the comparison between the smaller FR requirements (1 to 5) and the no-sample test (gray data point on Figure 3, plotted as “0” on the x axis). Firstly, we confirmed that performance was similar across all the smaller FRs: a repeated-measures ANOVA showed no main effect of FR requirement,  $F(2.10, 12.58) = 1.77$ ,  $p = .210$ ,  $\eta_G^2 = .108$ . However, when performance following the smaller FR (average proportion of “small” choices of .91) was compared with no-sample tests (.61), a paired-samples t-test revealed the difference to be statistically significant,  $t(6) = 7.44$ ,  $p < .001$ ,  $d_z = 2.81$ . That is, pigeons did not treat the absence of a sample – a “FR 0”, if you will – similarly to the other smaller FR requirements.

#### 4. Discussion

Pigeons learned, in a symbolic matching-to-sample task, to discriminate between small (5) and large (15) Fixed Ratio (FR) schedules. Following that, performance in delay and no-sample tests, ran in the same daily session, was compared. The delays were spent in darkness, due to the possibility that delay illumination could disrupt matching performance. Finally, untrained FR requirements were presented in a generalization test.

Starting with the question of whether the pigeons were responding on the basis of number of responses or timing, at the group level, the data followed the pattern predicted by timing ( $\text{time FR5}_{\text{incorrect}} > \text{time FR5}_{\text{correct}}$  and  $\text{time FR15}_{\text{correct}} > \text{time FR15}_{\text{incorrect}}$ ), but these differences were only significant for FR 5 trials. Individually, only three birds showed the pattern for both samples. Even though this



error analysis does not allow us to confirm that the animals were timing, that possibility should not be discounted. In fact, it has been found that time and number are not easily disentangled: When number and duration are correlated, both stimulus dimensions seem to be taken into account (e.g., Fetterman et al., 1986; Meck & Church, 1983; Meck et al., 1985; Roberts & Mitchell, 1994; Roberts & Boisvert, 1998). Regarding tasks involving FR discriminations, Rilling (1967) ran an error analysis to assess if number of responses or run time controlled choices, and found that number of responses was the best predictor of choice. However, Fetterman (1993) found that both run time and number of responses played a role in the animals' choices.

In any case, whether the discrimination is temporal (e.g., Spetch, 1987; Spetch & Wilkie, 1982) or numerical (e.g., Hope & Santi, 2004; Roberts et al., 1995; Santi & Hope, 2001), when inter-trial intervals (ITI) and delays are signaled similarly (for instance, both spent in darkness), choice biases towards the “short/small” comparison are commonly found. When ITI and delays are signaled differently, choice biases tend to weaken and sometimes disappear, and we found that to be true in our procedure, where the average retention function approached indifference between the comparisons (Figure 1). Even though the absence of choice biases in the average retention functions is consistent with the instructional ambiguity account (Fetterman & MacEwen, 1989; Zentall, 1997, 2006), the way these curves quickly reach indifference is at odds with this account, seeing that, by reducing the ambiguity between ITI and delay, better retention of the samples would be expected.

Focusing now on tasks with FR discriminations, our (average) results are in accord with Fetterman and MacEwen (1989)'s finding that when ITI and delay were

disambiguated choice biases tended to be weak or nonexistent. However, as previously mentioned, not all FR studies with differentiated ITI and delays return the same findings: while Fetterman and MacEwen (1989) found a weak bias for the “small” comparison or no bias at all, Maki, Moe, and Bierley (1977, Experiment 3) found a bias for the “large” comparison. At the individual level, we found all three cases (Figure 2): two “small” biases, two “large” biases, and three pigeons showing no bias. If delayed testing may lead to indifference between the comparisons, we could observe preferences moving away from chance levels either due to random variations or due to other biases that may be expressed then – a color bias, for instance. In this case, the disparity between Fetterman and MacEwen (1989) and Maki, Moe, and Bierley (1977) could be due to their reduced samples, four and three pigeons, respectively. That is, smaller samples could be more vulnerable to chance variations or the expression of punctual biases. Additionally, perhaps FR tasks are more fertile territory for diverging strategies – of learning the task or interpreting a delay – to emerge; given that two dimensions (time and number) co-vary, there are more degrees of freedom that could be expressed in a greater variety of outcomes.

In a timing task, Kelly and Spetch (2000) found that, while a dark ITI and an illuminated delay lead to no biases, an illuminated ITI and a dark delay yielded a choose-short effect. In addition to different ITI-delay arrangements leading to different delay results, delay illumination has been found to disrupt matching performance (D'Amato & O'Neill, 1971; Etkin, 1972; Grant & Roberts, 1976; Parker & Glover, 1987). The previous FR studies that disambiguated ITI and delay did so by illuminating the delay, so one of our goals was to assess if the opposite ITI-delay illumination arrangement would lead to different results. Given that our results seem

consistent with previous findings, it appears that, in those cases, the delay illumination did not bring any disruption. At least not additional disruption to the one that the introduction of delays brought: in our results the duration of the delays appeared to be irrelevant; the introduction of a delay, any delay, lead to a similar decrease in matching accuracy. Additionally, the functions plotting choices following the two samples tended to superimpose (Figure 2). That is, choices in delay testing did not appear to depend much on the sample or delay presented.

The step-like pattern of the functions, not showing a gradual decrease with delay duration, may suggest that the delay test was not measuring memory (where more forgetting would be more likely following longer delays), but instead causing a generalized disruption in matching, a disruption that was similar to all samples and all delays (see Ward & Odum, 2007, for an example of how stimuli may function as disruptors that induce choice biases in timing tasks). A disruption of this sort could limit the expression of memory-related choice biases, but similar step-like patterns have been obtained previously – Fetterman and MacEwen (1989) found it on three of their four birds –, but do not occur always – Fetterman and MacEwen (2003)'s retention functions followed a gradual shape. Similar results can be seen on timing tasks, with retention tests sometimes yielding step-like retention functions (e.g., Grant & Talarico, 2004; Pinto & Machado, 2015, 2017; Pinto & Sousa, 2021; Sherburne et al., 1998; Spetch & Rusak, 1992). It is not clear which conditions lead to different retention-function patterns, but the type of training could be an important factor: following matching-to-sample training with no delay, Spetch (1987) found step-like functions: accuracy was similar following 5- and 10-s delays (Phase 1b). However, after training with delays (and then presenting longer delays in testing), the effect of

the delay was gradual, with long-sample accuracy decaying more following a longer delay (Phases 2b and 3b). Training without delays may not always lead to step-like retention functions, but training with delays could preclude them from happening. However, in slightly different tasks, such as the discrimination of empty intervals, this may not always be the case: after training pigeons with a delay, Grant and Talarico (2004, Experiment 1) observed step-like retention functions in delay tests. The type of training may be another example of how task characteristics affect performance, and that, besides the direction of the preferences, the shape of the retention functions could be informative of processes ongoing in a delay test.

Regarding the no-sample test, firstly we replicated the “small” bias found by Fetterman and MacEwen (1989). One point of interest is the extent of this preference: In Fetterman and MacEwen (1989), proportion of “small” choices was .89, while in the present study it was not as pronounced, .61. In a study where animals had to learn a discrimination based on the number of flashes of the feeder light, Santi and Hope (2001) found, on a no-sample test, a proportion of choices to the “small” comparison of .78. We call attention to these values because a comparable bias (for the “short” comparison) is also commonly found in temporal discriminations (e.g., Church, 1980; Pinto et al., 2017; Pinto & Machado, 2015, 2017; Spetch & Wilkie, 1983), and this bias has been bounded between .60 and .80 (for a more detailed analysis of these results please refer to Pinto and Sousa, 2021). Taking into account all these studies, Fetterman and MacEwen (1989) found the strongest bias, which could be suggestive that there could be something different in no-sample tests in FR tasks (where both number and time co-vary), but our results did not confirm that. Perhaps future tests of no-sample matching following FR discriminations may shed light on this possibility.

Additionally, we were interested in comparing choices following a long delay and in a no-sample trial, to assess if the effect of a (long enough) delay between sample and comparisons was equivalent to presenting no sample to begin with. To allow for a closer comparison, we ran both tests in the same daily sessions, and found that they differed: while there was the aforementioned bias for “small” in no-sample tests, preference for “small” was around chance level (.46) following 20-s delays. Another result suggesting that no-sample and delay tests are not equivalent is that no-sample preferences were more consistent than preferences following a delay: At the individual level, in no-sample tests all pigeons were either indifferent or showed a “small” bias, while in delay testing, as we saw earlier, there was indifference, “small”, and “large” biases.

Finally, our last goal was to assess how well a trial where no sample pecking occurred (a no-sample trial) would fit on a pecking-requirement generalization gradient. For non-zero FR requirements, the gradient followed the typical pattern (Figure 3): performance was similar among the smaller requirements (from FR 1 to FR 5), with high proportions of “small” choices, and also similar between the larger requirements, FR 15 and FR 25, with almost no “small” choices occurring. When we extend the analysis to no-sample trials (plotted under “0” on the x axis of Figure 3), there is a discontinuity in the gradient; even though, numerically, a FR 0 is close to the smaller FR, there is a considerable discrepancy in choices between them: while proportion of “small” choices was .91 for the small FR requirements, it decreased to .61 in the no-sample trials. Choices following FR 0 did show a “small” bias (as would be expected by generalization), but the extent of this bias – resulting in a large discrepancy between FR 0 and the smaller FR – appears to be at odds with a

generalization account. A similar discontinuity has been found in temporal discriminations both when pigeons were required to maintain pecking during sample (Pinto & Sousa, 2021) and when they were not (Pinto & Machado, 2015).

This discrepancy may suggest that “zero” samples do not belong in the same continuum as “non-zero” samples. However, we must also take into consideration that, by definition, a no-sample trial – given the absence of a sample – is significantly different from “non-zero” sample trials and that could bring some degree of generalization decrement that might explain the discontinuity just described. Perhaps, future studies could attempt to reduce the difference between no-sample and regular training trials to find if that would bring choices following “zero” and “close to zero” samples closer together.

#### Acknowledgments

Funding: The present work was conducted at the Psychology Research Centre (PSI/01662), University of Minho, and was supported by the Portuguese Foundation for Science and Technology and the Portuguese Ministry of Science, Technology and Higher Education through national funds. It was also co-financed by FEDER – through COMPETE2020 – under the PT2020 Partnership Agreement (POCI-01-0145-FEDER-007653).

Compliance with ethical standards: The research was carried out in agreement with the European (Directive 2010/63/EU) and Portuguese law (Ordinance 1005/92 of October 23), being approved by the Directorate-General for Food and Veterinary, the Portuguese national authority for animal health (Authorization #024946).

CRedit authorship contribution statement: Carlos Pinto conceptualized the research and programmed the experiment. The experiment was conducted, the data analyzed, and the manuscript written by Carlos Pinto and Márcio Mota.

Conflict of Interest: The authors declare that they have no conflicts of interest.

#### References

- Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, *37*, 379–384.  
<https://doi.org/10.3758/BF03192707>
- Church, R. M. (1980). Short-term memory for time intervals. *Learning and Motivation*, *11*, 208–219. [https://doi.org/10.1016/0023-9690\(80\)90013-2](https://doi.org/10.1016/0023-9690(80)90013-2)
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Earlbaum Associates.
- D'Amato, M. R., & O'Neill, W. (1971). Effect of delay-interval illumination on matching behavior in the capuchin monkey. *Journal of the Experimental Analysis of Behavior*, *15*, 327–333. <https://doi.org/10.1901/jeab.1971.15-327>
- Etkin, M. W. (1972). Light produced interference in a delayed matching task with capuchin monkeys. *Learning and Motivation*, *3*, 313–324.  
[https://doi.org/10.1016/0023-9690\(72\)90027-6](https://doi.org/10.1016/0023-9690(72)90027-6)
- Fetterman, J. G. (1993). Numerosity discrimination: Both time and number matter. *Journal of Experimental Psychology: Animal Behavior Processes*, *19*, 149–164.  
<https://doi.org/10.1037/0097-7403.19.2.149>

- Fetterman, J. G., & MacEwen, D. (1989). Short-term memory for responses: the “choose-small” effect. *Journal of the Experimental Analysis of Behavior*, *52*, 311–324. <https://doi.org/10.1901/jeab.1989.52-311>
- Fetterman, J. G., & MacEwen, D. (2003). Acquisition and retention in compound matching with hue and peck number elements. *Learning and Motivation*, *34*, 354–371. [https://doi.org/10.1016/S0023-9690\(03\)00034-1](https://doi.org/10.1016/S0023-9690(03)00034-1)
- Fetterman, J. G., Stubbs, D. A., & Dreyfus, L. R. (1986). Scaling of events spaced in time. *Behavioural Processes*, *13*, 53–68. [https://doi.org/10.1016/0376-6357\(86\)90016-1](https://doi.org/10.1016/0376-6357(86)90016-1)
- Grant, D. S. (2001). Memory for empty time intervals in pigeons. *Animal Learning & Behavior*, *29*, 293–301. <https://doi.org/10.3758/BF03192896>
- Grant, D. S., & Talarico, D. C. (2004). Processing of empty and filled time intervals in pigeons. *Learning & Behavior*, *32*, 477–490. <https://doi.org/10.3758/BF03196043>
- Grant, D. S., & Roberts, W. A. (1976). Sources of retroactive inhibition in pigeon short-term memory. *Journal of Experimental Psychology: Animal Behavior Processes*, *2*. <https://doi.org/10.1037/0097-7403.2.1.1>
- Hope, C., & Santi, A. (2004). Pigeons’ memory for number of events: Effects of intertrial interval and delay interval illumination. *Learning and Motivation*, *35*, 348–370. <https://doi.org/10.1016/j.lmot.2004.04.005>
- 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 B: Comparative and Physiological Psychology*, *53*, 309–323. <https://doi.org/10.1080/713932737>



- 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.  
<https://doi.org/10.3758/BF03208009>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 1–12. <https://doi.org/10.3389/fpsyg.2013.00863>
- Maki, W. S., Moe, J. C., & Bierley, C. M. (1977). Short-term memory for stimuli, responses, and reinforcers. *Journal of Experimental Psychology: Animal Behavior Processes*, *3*, 156–177. <https://doi.org/10.1037/0097-7403.3.2.156>
- Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, *9*, 320–334. <https://doi.org/10.1037/0097-7403.9.3.320>
- Meck, W. H., Church, R. M., & Gibbon, J. (1985). Temporal integration in duration and number discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, *11*, 591–597. <https://doi.org/10.1037/0097-7403.11.4.591>
- Olejnik, S., & Algina, J. (2003). Generalized Eta and Omega Squared Statistics: Measures of Effect Size for Some Common Research Designs. *Psychological Methods*, *8*, 434–447. <https://doi.org/10.1037/1082-989X.8.4.434>
- Parker, B. K., Glover, R. L. (1987). Event duration memory: The effects of delay-interval illumination and instructional cuing. *Animal Learning & Behavior*, *15*, 241–248. <https://doi.org/10.3758/BF03205018>

- Pinto, C., Fortes, I., & Machado, A. (2017). Joint stimulus control in a temporal discrimination task. *Animal Cognition*, *20*, 1129–1136.  
<https://doi.org/10.1007/s10071-017-1130-6>
- 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. <https://doi.org/10.1016/j.lmot.2010.05.001>
- Pinto, C., & Machado, A. (2015). Coding in pigeons: Multiple-coding versus single-code/default strategies. *Journal of the Experimental Analysis of Behavior*, *103*, 472–483. <https://doi.org/10.1002/jeab.153>
- Pinto, C., & Machado, A. (2017). Unraveling sources of stimulus control in a temporal discrimination task. *Learning & Behavior*, *45*, 20–28.  
<https://doi.org/10.3758/s13420-016-0233-2>
- Pinto, C., & Sousa, A. (2021). Choice biases in no-sample and delay testing in pigeons (*Columba livia*). *Animal Cognition*, *24*, 593–603.  
<https://doi.org/10.1007/s10071-020-01457-1>
- Rilling, M. (1967). Number of responses as a stimulus in fixed interval and fixed ratio schedules. *Journal of Comparative and Physiological Psychology*, *63*, 60–65.  
<https://doi.org/10.1037/h0024164>
- Roberts, W. A. (1997). Does a common mechanism account for timing and counting phenomena in the pigeon? In C. M. Bradshaw & E. Szabadi (Eds.), *Time and behaviour: Psychological and neurobehavioural analyses*. (pp. 185–215) Amsterdam: Elsevier.
- Roberts, W. A., & Boisvert, M. J. (1998). Using the peak procedure to measure timing and counting processes in pigeons. *Journal of Experimental Psychology*:

- Animal Behavior Processes*, 24, 416–430. <https://doi.org/10.1037/0097-7403.24.4.416>
- Roberts, W. A., Macuda, T., & Brodbeck, D. R. (1995). Memory for number of light flashes in the pigeon. *Animal Learning & Behavior*, 23, 182–188. <https://doi.org/10.3758/BF03199933>
- Roberts, W. A., & Mitchell, S. (1994). Can a pigeon simultaneously process temporal and numerical information? *Journal of Experimental Psychology: Animal Behavior Processes*, 20, 66–78. <https://doi.org/10.1037/0097-7403.20.1.66>
- Santi, A., & Hope, C. (2001). Errors in pigeons' memory for number of events. *Animal Learning & Behavior*, 29, 208–220. <https://doi.org/10.3758/BF03192888>
- Santi, A., Hornyak, S., & Miki, A. (2003). Pigeons' memory for empty and filled time intervals signaled by light. *Learning and Motivation*, 34, 282–302. [https://doi.org/10.1016/S0023-9690\(03\)00021-3](https://doi.org/10.1016/S0023-9690(03)00021-3)
- Santi, A., Ross, L., Coppa, R., & Coyle, J. (1999). Pigeons' memory for empty time intervals marked by visual or auditory stimuli. *Animal Learning & Behavior*, 27, 190–205. <https://doi.org/10.3758/BF03199675>
- 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 intertrial intervals. *Psychonomic Bulletin & Review*, 5, 516–522. <https://doi.org/10.3758/BF03208831>
- Spetch, M. L. (1987). Systematic errors in pigeons' memory for event duration: Interaction between training and test delay. *Animal Learning & Behavior*, 15, 1–5. <https://doi.org/10.3758/BF03204897>

- Spetch, M. L., & Grant, D. S. (1993). Pigeons' Memory for Event Duration in Choice and Successive Matching-to-Sample Tasks. *Learning and Motivation*, 24, 156–174. <https://doi.org/10.1006/lmot.1993.1010>
- Spetch, M., & Rusak, B. (1992). Time present and time past. In W. Honig & J. Fetterman (Eds.), *Cognitive aspects of stimulus control* (pp. 47–67). Hillsdale, NJ: Erlbaum.
- Spetch, M., & Wilkie, D. (1982). A systematic bias in pigeons' memory for food and light durations. *Behaviour 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. <https://doi.org/10.1037/0097-7403.9.1.14>
- Ward, R., & Odum, A. (2007). Disruption of temporal discrimination and the choose-short effect. *Learning & Behavior*, 35, 60–70. <https://doi.org/10.3758/BF03196075>
- Zentall, T. R. (1973). Memory in the pigeon: Retroactive inhibition in a delayed matching task. *Bulletin of the Psychonomic Society*, 1, 126–128. <https://doi.org/10.3758/BF03334317>
- Zentall, T. R. (1997). Animal Memory: The Role of “Instructions.” *Learning and Motivation*, 28, 280–308. <https://doi.org/10.1006/lmot.1996.0968>
- Zentall, T. R. (2006). Timing, memory for intervals, and memory for untimed stimuli: The role of instructional ambiguity. *Behavioural Processes*, 71, 88–97. <https://doi.org/10.1016/j.beproc.2005.07.008>

Zentall, T. R., & Smith, A. P. (2016). Delayed matching-to-sample: A tool to assess memory and other cognitive processes in pigeons. *Behavioural Processes*, *123*, 26–42. <https://doi.org/10.1016/j.beproc.2015.07.002>

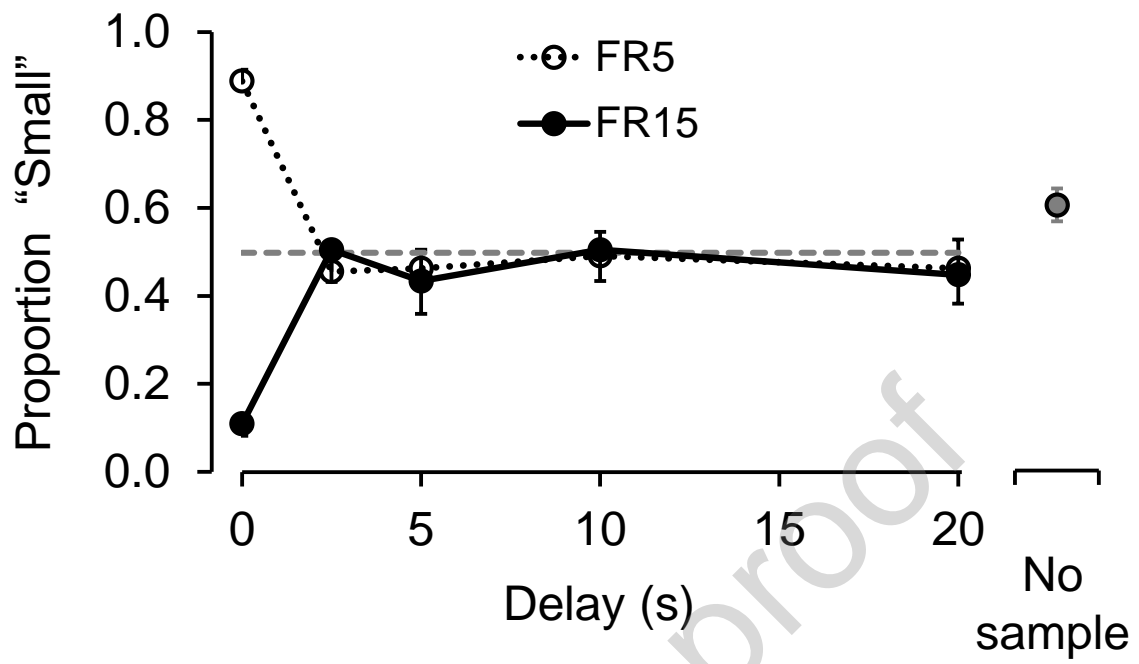
Journal Pre-proof

## Figure Captions

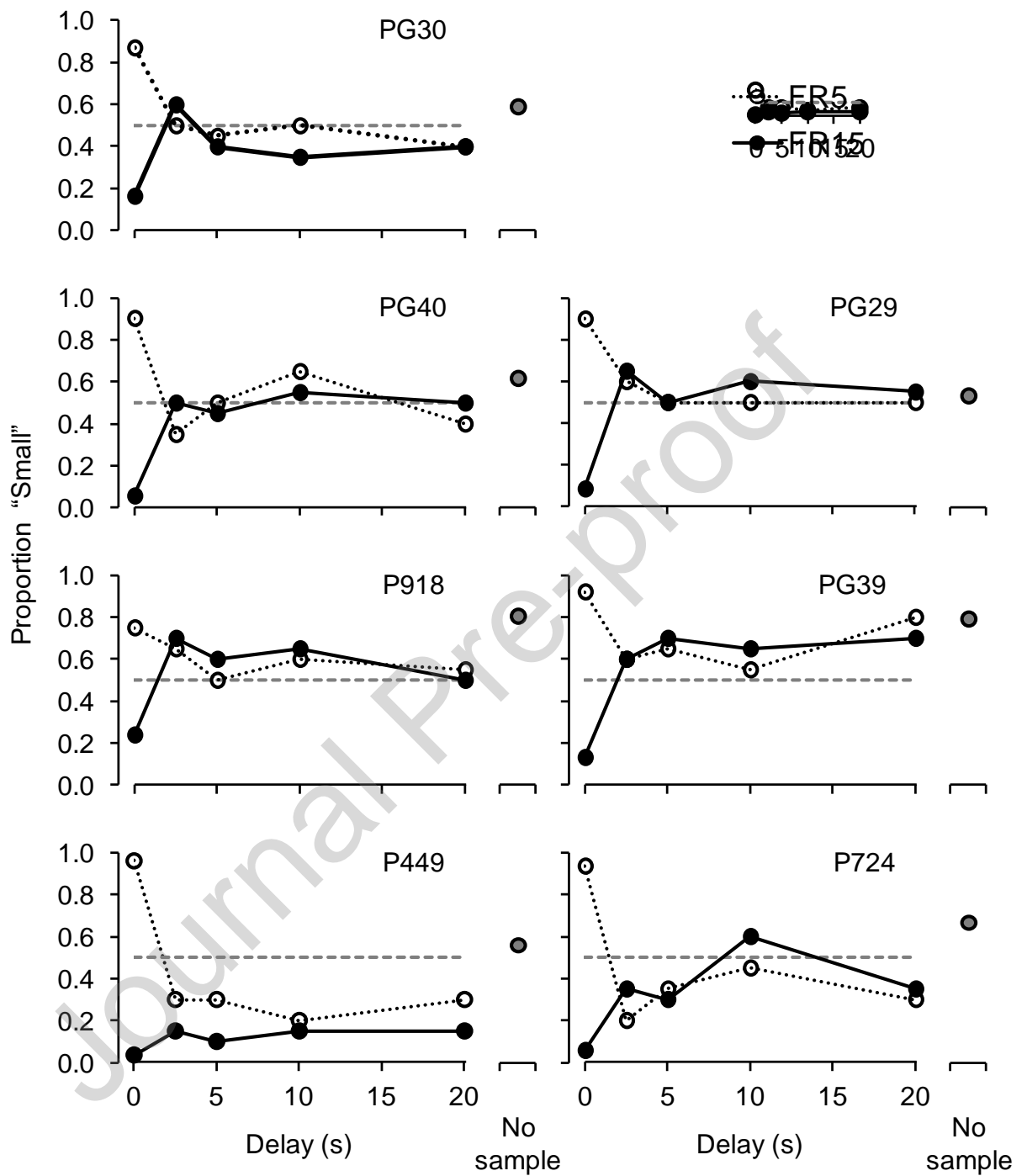
Figure 1 – Mean (with SEM) proportion of choices to the comparison associated with the small (FR5) sample following FR5 (empty data points connected by dotted line) and FR15 (filled data points connected by solid line) samples as a function of delay duration and on no-sample trials (rightmost data point, colored gray). The horizontal dashed line represents a proportion of 0.5.

Figure 2 – Mean proportion of choices to the comparison associated with the small (FR5) sample following FR5 (empty data points connected by dotted lines) and FR15 (filled data points connected by solid lines) samples as a function of delay duration and on no-sample trials (rightmost data point in each panel, colored gray). The horizontal dashed lines represent a proportion of 0.5. Each panel refers to one pigeon.

Figure 3 – Mean (with SEM) proportion of choices to the “small” comparison (comparison associated with FR5) as a function of FR requirement on the Generalization Test. The empty data points identify the previously-trained FR (FR5 and FR15) and the gray dot is the result of the No-sample Test.

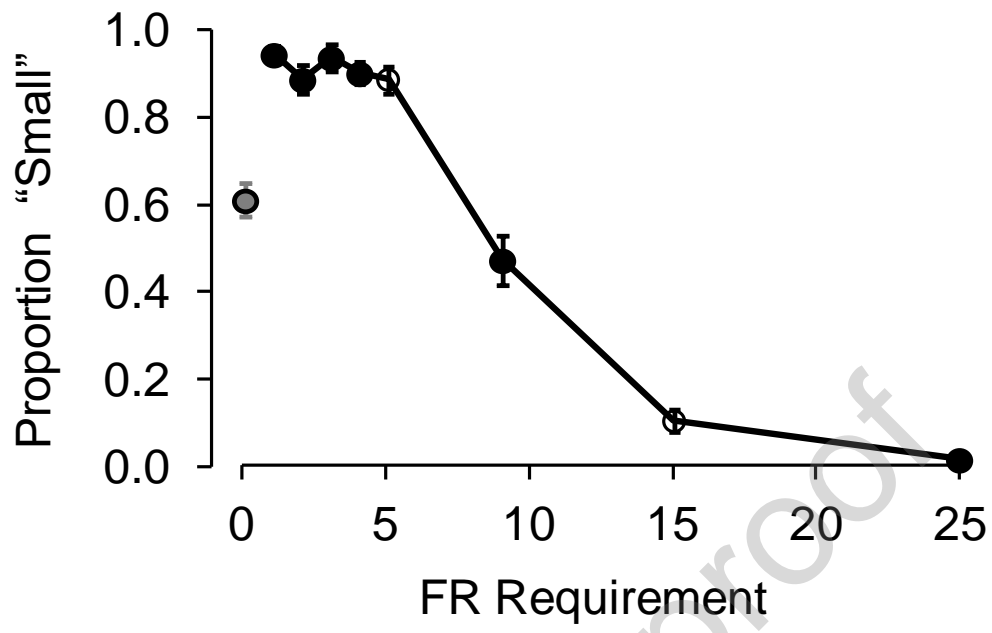


Pinto &amp; Mota - Figure 1



Pinto &amp; Mota - Figure 2





Pinto &amp; Mota - Figure 3

### Highlights

- Pigeons were trained to discriminate number of responses (Fixed Ratio, FR)
- A retention test revealed different choice biases
- Delay performance did not match no-sample (FR0) performance
- FR0 preferences did not appear to conform to a FR-requirement generalization gradient