1   Choice Biases in No-sample and Delay Testing in Pigeons (*Columba livia*)

2   Carlos Pinto and Ana Sousa

3   University of Minho

4

5   University of Minho

6   School of Psychology

7   Campus de Gualtar

8   4710 Braga

9   Portugal

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

11   ORCID (Carlos Pinto): 0000-0002-4435-6880

12

16

17   Abstract

18   In experimental tasks that involve stimuli that vary along a quantitative continuum, some

19   choice biases are commonly found. Take, for instance, a matching-to-sample task where

20   animals must, following the presentation of sample stimuli (that differ in duration),

21   choose between two or more comparison stimuli. In tests where no sample is presented

22   there is usually a bias towards the comparison that is correct following the shortest

23   sample. To examine some aspects of these choice biases, pigeons were trained in a

24   symbolic matching-to-sample task with two durations of keylight as samples, where key

25   pecking had to be maintained during sample presentation. Firstly, even though animals

26   were required to attend to the sample, a preference for the "short" comparison in no-

27   sample testing was found. This result disproves an account where this effect was

28   hypothesized to happen due to non-programmed learning resulting from the animals

29   failing to attend to some trials. Secondly, even though a bias for "short" was found in

30   both no-sample and delay testing, the extent of the biases differed between tasks, thus

31   suggesting that forgetting the sample presented during a delay does not necessarily land

32   the animal in a state similar to presenting no sample at all to begin with.

33

34   Keywords: Pigeon, discrimination learning, symbolic matching-to-sample, delayed

35   matching-to-sample, choice, bias.

36

37

38

39   Declarations

45   Conflicts of interest/Competing interests: The authors declare that they have no conflict

46   of interest.

47   Ethics approval: The research was carried out in agreement with the European (Directive

48   2010/63/EU) and Portuguese law (Ordinance 1005/92 of October 23), being approved by

49   the Directorate-General for Food and Veterinary, the Portuguese national authority for

50   animal health (Authorization #024946).

51   Consent to participate: Not applicable

52   Consent for publication: Not applicable

53   Availability of data and material (data transparency): The datasets generated and/or

54   analysed during the current study are available from the corresponding author on

55   reasonable request.

56   Code availability: Not applicable

57

58   Choice Biases in No-sample and Delay Testing in Pigeons (*Columba livia*)

59

60   Relevant events in an environment may be signaled by a multitude of cues,

61 spanning several stimulus dimensions (e.g., visual, auditory, olfactory, or related to

62 location, duration or quantity). The way these stimulus dimensions are perceived can

63 vary in fundamental ways. For instance, while the color of a stimulus can be identified

64 immediately from its onset, its duration cannot. The same is true for a number whose

65 components are presented in succession: if a stimulus feature requires quantitative

66 estimation over time, be it time itself or number, identification is not immediate. These

67 features also fall along a quantitative continuum: a long duration (or a large number)

68 includes shorter durations (or smaller numbers). Such particularities may not only affect

69 how different stimulus dimensions are perceived, but also our study of them.

70   Take for instance a symbolic matching-to-sample task where, following the

71 presentation of a stimulus (known as sample), a choice must be made between two or

72 more options (known as comparisons). A version of this task uses stimulus duration as

73 samples, thus requiring the discrimination between two (or more) durations. Taking a

74 matching-to-sample task for pigeons as an example, on some trials the sample (e.g.,

75 houselight) is presented for 2 s (short sample) and the pigeons are rewarded for choosing

76 one comparison key (e.g., a pecking key illuminated with a green hue). On the remaining

77 trials the houselight is presented for 6 s (long sample) and the pigeons are rewarded for

78 choosing the other comparison key (e.g., a pecking key illuminated with a red hue).

79   A task focused on stimulus duration brings with it some unavoidable

80   asymmetries. In our example, all trials will, in a way, involve the shorter duration: a trial

81   will either end after 2 s, or will go through the shorter duration to reach 6 s. The same is

82   not true for the longer duration. Additionally, the short sample will, by definition, be

83   present less time than the long sample; if an animal temporarily disengages from the task,

84   it is more likely to miss a short sample than a long sample. Along a similar line, the

85   exposure to a short stimulus does not last the same as to a long stimulus, and the time

86   from the beginning of a trial until a reinforcer is collected may differ between short-

87   sample and long-sample trials. The preceding examples can also apply to discriminations

88   based on number, be it of stimuli or responses: if the components of a number are

89   presented sequentially, a smaller number of stimuli may take less time to present than a

90   larger one; similarly, a smaller number of responses will generally take less time to

91   produce than a larger number of responses. These differences may lead to an asymmetry

92   in how responding to the samples is learned, and ultimately to potential choice biases.

93   Some of these choice biases can be found in tasks where a delay is introduced

94   before a response, a strategy that has a long tradition as a means to study memory – some

95   records of this practice are over a century old (Hunter 1913). In a delayed symbolic

96   matching-to-sample task (e.g., Blough 1959; D'Amato 1973; Wixted 1989), the delay is

97   inserted between sample and comparisons, and is a period during which the information

98   required to select the correct comparison must be maintained – hence this delay also

99   being known as a retention interval. To that end, during the delay the stimulus dimension

100   that was trained is not presented. For example, in a visual discrimination, the delay may

be a period of darkness so that no other visual stimuli interfere with performance.

However, in temporal discrimination tasks, such a control is not attainable: it is not

possible to remove the relevant stimulus dimension (time) from the delay. Thus, a delay

may be interpreted as a new temporal stimulus instead of a retention interval, which can

make the interpretation of delay tests challenging. In fact, in some cases the delay

appears to be added to the sample duration (e.g., Grant 2001; Grant and Talarico 2004;

Pinto and Machado 2015, 2017; Santi et al. 2003; Santi et al. 1999).

A choice bias found in (temporal) delayed matching-to-sample tasks is known as

the choose-short effect (e.g., Grant and Spetch 1991, 1993; Kelly and Spetch 2000;

Kraemer et al. 1985; Pinto and Machado 2011; Spetch 1987; Spetch and Grant 1993;

Spetch and Wilkie 1982): as the delay increases, accuracy on long-sample trials

decreases, i.e., animals tend to instead choose the "short" comparison (the comparison

correct following short samples). Interestingly, accuracy on short-sample trials remains at

a higher level through all the delays. A similar bias has been found when the samples

consisted of number of responses: when one comparison is correct following a small

number of responses, and the other comparison following a large number of responses,

there is a preference for the "small" comparison in delay testing (Fetterman and

MacEwen 1989, 2003).

Several explanations for the choose-short effect have been put forth, positing

different possibilities to what happens during a delay: a gradual shortening of a memory

trace (Spetch and Wilkie 1983), a sudden loss of a code for the sample (Kraemer et al.

1985), or confusion with the inter-trial interval (Sherburne et al. 1998). An assumption

123   common to all these explanations is that following a (long enough) delay, the animal

124   would be in a situation akin to having no sample presented to begin with. Additionally,

125   when no sample is presented, the animals are predicted to choose the comparison

126   associated with the shortest sample duration, due to stimulus generalization (the absence

127   of a sample is considered a 0-s sample, which is closest to the shortest of the trained

128   samples). Hence, the preference for the "short" comparison in delay testing.

129   When no-sample tests are run (as the name indicates, in these tests the

130   presentation of the comparisons is not preceded by a sample), a preference for the

131   comparison associated with shortest sample is consistently found (Church 1980; Pinto et

132   al. 2017; Pinto and Machado, 2015, 2017; Spetch and Wilkie 1983). The preference for

133   the "short" comparison in no-sample tests aligns with how the aforementioned theories

134   (Kraemer et al. 1985; Sherburne et al. 1998; Spetch and Wilkie 1983) explain the choose-

135   short effect. However, performance on no-sample trials is not always in accord with

136   performance in delay testing. Church (1980, Experiment 3A) found that rats, after

137   training in a 2 s vs. 8 s discrimination task, showed a preference for the "short"

138   comparison in a no-sample test. However, the same animals' preferences approached

139   indifference in delay testing. In another example, pigeons learned a matching-to-sample

140   task with three samples (2, 6, and 18 s) and two comparisons, where choices of one

141   comparison were correct following the shortest sample, and choices of the other

142   comparison were correct following the two longer samples (Pinto and Machado 2015,

143   2017). In delay testing, the birds showed a preference for the "long" comparison, but in a

144   no-sample test a preference for the "short" comparison was, once again, found. In sum,

145   on no-sample trials there is a robust preference for the comparison associated with the

146   shortest sample, but this preference does not necessarily track the preferences shown in

147   delay testing. These results suggest that performance in the two tasks may be unrelated –

148   the effect of a delay may not be the same as having no sample presented.

149   As mentioned, performance on no-sample trials has been explained by stimulus

150   generalization. If sample durations are distributed along a continuum, the absence of a

151   sample (a 0-s sample) would be closest to the shortest of the samples. Thus, in a no-

152   sample test, choices should be made to the comparison that is correct following the

153   sample closest to 0s, the short sample. Nevertheless, some data appears to go against this

154   hypothesis. In the 3-sample, 2-comparison procedure mentioned in the previous

155   paragraph, Pinto and Machado (2015) ran a generalization test where new sample

156   durations were presented. If responding on no-sample (0 s) trials were controlled by

157   stimulus generalization, performance on 0-s, 1-s, and 2-s trials should be similar.

158   However, while following 1-s and 2-s samples the pigeons preferred the "short"

159   comparison on, respectively, 97% and 95% of the trials, this preference was markedly

160   different on 0-s trials, dropping to 72% – a result hard to reconcile with a generalization

161   account. Perhaps the difference between no sample and a non-zero sample is more of a

162   qualitative nature (presence versus absence) rather than a quantitative one (different

163   points along the same continuum).

164   Given that a generalization account does not appear to fully explain the

165   preference for "short" on no-sample tests, other processes may be at play. Due to some of

166   the idiosyncrasies of temporal stimuli mentioned earlier in the text, this preference for

167   "short" could be the result of learning a non-programmed contingency. While learning

168   the task, during the inter-trial interval the animals may engage in several activities, such

169   as grooming, or exploring the experimental chamber. When a new trial begins, these

170   activities may preclude the animals from attending to the intelligence panel quickly

171   enough to see the sample. Since the short sample is the one that, by definition, is

172   presented for the smallest amount of time, it is the most likely to be missed. Therefore,

173   animals may have learned that, when they did not see a sample, choices of the "short"

174   comparison were most often the ones reinforced.

175   In the present work, pigeons were trained in a matching-to-sample task with two

176   durations of white keylight (2 and 6 s) as samples. To initiate a trial the birds had to peck

177   once in a key illuminated with a cross, which would then turn to white (the sample

178   stimulus). The animals were required to maintain pecking (at least one peck every 1.5 s)

179   on this illuminated key during sample presentation. If animals stopped pecking during

180   the sample, the trial would be cancelled. This requirement aimed to prevent the animals

181   from disengaging from the task and make a choice between the comparisons without

182   attending to the sample. After learning this task, the pigeons were first exposed to a no-

183   sample test, to assess whether the typical preference for the "short" comparison would

184   emerge. Secondly, a delay test was run, with delays ranging from 2.5 to 20 s. We were

185   interested to see if, in this variant of the matching-to-sample task with a peck

186   requirement, delay and no-sample testing would yield similar results, thus providing a

187   test of one of the main assumptions of all models of the choose-short effect.

188   Finally, a second type of no-sample test was run. In the first no-sample test, a test

189   trial consisted solely in the presentation of the comparisons (no peck was required to

190   initiate a trial). In this second no-sample test, the animals were required to peck once on a

191   key (illuminated with a cross) to initiate a trial, which was then immediately followed by

192   the presentation of the comparisons. Given that the pigeons were required to peck during

193   sample presentation, it is possible that, instead of sample duration, the animals were

194   responding on the basis of the duration spent pecking, or of the number of pecks

195   produced in each trial. In both cases, a no-sample trial that required a peck to be initiated

196   could functionally be a "very short" sample (the role of "no sample" would be reserved

197   exclusively for the no-sample tests with no initial peck). If the animals were basing their

198   choices on duration of the sample alone there should be no difference between the no-

199   sample tests (in both, the sample was not presented), but if timing was initiated by

200   pecking (or if the number of pecks was a discriminative stimulus), the two no-sample

201   tests could yield different results.

202

203

Method

Subjects

7 pigeons (*Columba livia*) were kept at approximately 85% of their free-feeding weight

for the duration of the experiment. The birds were housed in individual home cages,

where water and grit were available at all times. The colony room was maintained in a

13:11-h light/dark cycle (with lights on at 8 AM) where temperatures oscillated between

20 and 22 °C. The experiment was conducted once a day, at approximately the same time

for each pigeon, 6 days a week. All animals had experience with timing tasks; the

experience varied between birds (Fixed Interval or Variable Interval schedules,

Midsession Reversal Task, Temporal Bissection, Matching to sample in a long

experimental chamber), but PG18 and P088 took part in procedures that were closest to

the present experiment (symbolic matching-to-sample tasks with similar sample and

comparison stimuli, although three samples and two comparisons were employed, and

pecking during the sample was not required). PG18, P088, and PG49 had experience

with number-discrimination tasks.

Apparatus

Two identical operant chambers (Med Associates, ENV-007), measuring 28.5 cm

(height) × 30 cm (length) × 24 cm (width), were used. On the response panel, three keys

(2.5 cm in diameter) were horizontally arranged, 8 cm center-to-center apart from one

another. The bottom of each key was 20.5 cm above the steel grid floor, and every key

was equipped with a 12-stimulus projector (Med Associates, ENV-130M). The food,

226   mixed grain, was delivered via a LVE (Lehigh Valley Electronics) food hopper, and was

227   accessible through a 6.5-cm wide × 5-cm high opening, centered horizontally on the

228   response panel, 4 cm above the floor. When the hopper was raised, a 28-V, 0.04-A light

229   illuminated its opening. On the wall opposite to the response panel (26 cm above the

230   floor) a 28-V, 0.1-A houselight provided general illumination. The operant chamber was

231   inside a PVC sound attenuating cubicle (Med Associates, ENV-018V), equipped with an

232   exhaust fan that circulated air through the chamber and masked outside noises.

233   In this experiment, red or green hues were presented on the side keys and a white

234   hue or a white cross (+) on a black background were presented on the center key.

235   Personal computers running the ABET II software (Lafayette Instrument Company)

236   controlled the experimental events and recorded the data.

237

238   *Procedure*

239   *Pre-training*. To ensure that the birds were pecking reliably, two pre-training

240   sessions were run. The first of those sessions was a 60-trial session of a Fixed Ratio (FR)

241   5 schedule: in each trial, the animals had to peck five times on the center key to receive 5

242   s of access to food. In half of the trials the key was illuminated with a white hue, and in

243   the other half with a white cross (+) on a black background. These stimuli would both be

244   used in the experiment proper. A 30-s, houselight-illuminated, intertrial interval (ITI)

245   separated the trials. The houselight was turned off during the trials. In the second pre-

246   training session, reinforcement was made available under a Variable Ratio 15 schedule.

247   The 60-trial session was divided in two 30-trial blocks (from a programming standpoint;

248   from the point of view of the animals nothing signaled a block separation). Within a

249   block, the response requirement varied between 1 and 29, in increments of one, each

250   presented once (with the exception of 15, that was presented twice), randomly. Aside

251   from the reinforcement requirement, the trial structure was the same as in the FR 5

252   session.

253   *Training*. Pigeons learned a symbolic matching-to-sample task. A trial began with

254   the illumination of the center key with a white cross on a black background. A single

255   peck at this key turned off the cross and initiated the sample by illuminating the center

256   key with a white hue. Pigeons were required to peck – at least once every 1.5 s – for the

257   duration of the sample, which could be 2 or 6 s. If the rate of pecking was not met, the

258   trial would be cancelled, the ITI would begin and the trial would then be repeated. If

259   pecking met the required rate, at the end of the sample the white center key was turned

260   off and the two side keys were illuminated, one with a red hue and the other with a green

261   hue (comparison stimuli). One comparison was correct for the 2-s sample and the other

262   comparison was correct for 6-s sample. The correct comparison for each sample was

263   counterbalanced across animals. After a single peck to either comparison, both

264   comparison keylights were turned off. If the comparison was chosen correctly,

265   reinforcement was delivered and then the ITI started; if the comparison was chosen

266   incorrectly, no reinforcement was delivered and the ITI started immediately. The ITI had

267   a fixed duration of 30 s, during which only the houselight was lit (aside from the ITI, the

268   houselight was always turned off). To minimize feeding outside the experimental

269 session, reinforcement duration was adjusted individually and varied from 2 s to 4 s

270 across birds.

271 A correction procedure was in effect: for the first training sessions, following one

272 incorrect choice, the trial repeated and only the correct comparison was presented in the

273 repeated trial. When the birds were able to conclude two full sessions without stalling,

274 the number of consecutive errors required to initiate a correction procedure was increased

275 to two. When two sessions were completed without stalling under these conditions, the

276 number of repetitions was increased to the final value of three (that is, after three

277 consecutive errors, in the next repetition of the trial only the correct comparison was

278 presented).

279 Each session comprised 60 trials (excluding correction trials), 30 2-s sample

280 trials, and 30 6-s sample trials. The trials were arranged in three 20-trial blocks (ten 2-s

281 trials and ten 6-s trials in each block). Across trials, the location of the comparisons

282 varied randomly with the constraint that each comparison was presented the same

283 number of times on each side key. Training lasted a minimum of 15 sessions and

284 continued until the pigeon met a criterion of at least 80% correct responses to each

285 sample in a session (excluding correction trials), for three consecutive sessions.

286 *No-sample test without initial peck (I)*. In this test there were two types of trials:

287 regular training trials and no-sample test trials. The regular training trials remained

288 exactly as during Training. On no-sample trials the center key was never illuminated

289 (neither with a cross nor a white hue), and a trial consisted solely in the presentation of

290 the comparisons. The position of the comparisons was counterbalanced so that each

291   comparison appeared the same number of times on each side key. On a no-sample trial,

292   responses were never reinforced. Each session comprised a total of 56 trials, 48 regular

293   training trials – 24 2-s sample trials, and 24 6-s sample trials, – and 8 no-sample trials.

294   Trials were organized in two 28-trial blocks (each block composed of 12 2-s trials, 12 6-s

295   trials, and 4 no-sample trials) and were presented randomly inside each block. This test

296   lasted for five sessions. To minimize feeding outside the operant chamber during this

297   test, reinforcement durations were readjusted and varied from 2 s to 4.5 s across animals.

298   *Delay Test*. The overall structure of the session was maintained from the previous

299   test, with the no-sample test trials replaced by delay test trials. A delay trial began

300   similarly to training trials: following a single peck on the key illuminated with a cross,

301   the sample was presented, during which pecking had to be maintained at a minimum rate

302   of 1.5 pecks per second. After sample offset, the experimental box was in total darkness

303   for a delay lasting 2.5, 5, 10 or 20 s. After the delay, the 2 comparison keys were

304   illuminated. Correct choices were reinforced and then followed by the ITI, and incorrect

305   choices initiated the ITI immediately. Each session comprised 56 trials, 48 training trials

306   (24 x 2-s, 24 x 6-s) and 8 delay test trials (2 trials for each delay). This test was run for

307   five sessions.

308   *No-sample test with initial peck (I)*. This test was the same as the no-sample test

309   without initial peck, the only exception being how a no-sample test trial initiated. On a

310   test trial, the center key was illuminated with a white cross on a black background and,

311   after a single peck on this key, the comparisons were presented.

312   Given that between the two no-sample tests there was a delay test, and to ensure

313   that any potential differences between no-sample tests were not due to this intermediate

314   test, the no-sample tests were replicated.

315   *No-sample test without initial peck (II)*. This test was an exact replication of the

316   first no-sample test without initial peck, and was run for five sessions.

317   *No-sample test with initial peck (II)*. This test was an exact replication of the first

318   no-sample test with initial peck, and was run for five sessions.

319

320   *Data Analysis*

321   We analyzed the number of responses that occurred during sample presentation

322   and also choice behavior between comparison stimuli. Parametric tests were conducted in

323   all our statistical analyses – using IBM SPSS Statistics for Windows, Version 17.0 – with

324   the Type-1 error rate set at 0.05. On no-sample tests, choice performance was contrasted

325   to chance levels as well as between tests via t tests, and a repeated-measures ANOVA

326   compared the two runs of the two tests. Responding during sample presentation was

327   separated between trials that ended with correct and incorrect choices – these two groups

328   of data were compared via paired-samples t-test. Regarding delay testing, a repeated-

329   measures ANOVA (with sample and delay durations as factors) was performed to assess

330   the effect of delay on choice. 95% Confidence Intervals were calculated for choice

331   following long delays, to allow comparisons with chance levels as well as with

332   performance on no-sample tests. For measures of effect size, the standardized mean

333   difference effect size was used for t tests (d_z, see, e.g., Cohen 1988, p. 48; Lakens 2013),

334   and the generalized eta square was used for ANOVA (η_G^2, e.g., Bakeman 2005; Olejnik

335   and Algina 2003).

336

337

338

339   Results

340   The pigeons took an average of 23 sessions (range: 18 – 35) to complete the

341   training. On the last three sessions of training, average matching accuracy was 88%

342   (range: 83% – 91%) for 2-s samples and 86% (82% – 91%) for 6-s samples.

343   Even though completing training successfully, pigeon P161 showed some

344   difficulty in maintain pecking at the required rate in the first days of testing, failing to

345   complete all daily sessions. Thus, for this bird to reach a number of test trials equivalent

346   to the remaining animals, its first no-sample test was run for seven sessions. All the

347   remaining tests were successfully completed in the planned five sessions.

348   In no-sample tests, due to an error in the trial-randomizing algorithm, following

349   the completion of the programmed 56 trials, sporadically one additional test trial was

350   presented. The error was corrected mid-experiment, and happened in 9 of the 142 no-

351   sample test sessions run. These additional test trials were included in the analyses.

352   The two no-sample tests (with and without a peck to initiate no-sample trials)

353   were run twice, and Figure 1 depicts the results of all four no-sample test runs. In all tests

354   there was a strong preference for the "short"/"small" comparison (the comparison correct

355   following the sample that is short, 2s, and also the sample during which the smaller

356   number of pecks is given): Across all seven birds and four tests, there was only one

357   single session where the "long"/"large" comparison was chosen more often than the

358   "short"/"small" comparison (note that the y-axis in Figure 1 begins at 50% preference for

359   "short"/"small"). For all four no-sample tests, preference for "short"/"small" was

360   statistically significant above chance levels of 50%: without peck requirement (first run,

361 $t(6) = 12.67$, $p < .001$, $d_z = 4.79$; second run, $t(6) = 4.73$, $p = .003$, $d_z = 1.79$), and with

362 peck requirement (first run, $t(6) = 26.05$, $p < .001$, $d_z = 9.84$; second run, $t(6) = 8.56$, p

363 $< .001$, $d_z = 3.23$).

364   The requirement of a peck to initiate no-sample trials appeared to modulate

365 preference. Starting by analyzing the first run of the no-sample tests (Figure 1, filled data

366 points), preference for the "short"/"small" comparison was 79% on the first test (where

367 no peck was required to initiate a no-sample trial), and increased to 94% on the second

368 test (where a single peck was required to initiate a no-sample trial). A paired-samples t-

369 test confirmed that this increase was statistically significant, $t(6) = 4.94$, $p = .003$, $d_z =$

370 1.87. When the first no-sample test (without peck) was re-run, preference for

371 "short"/"small" returned to a lower value (75%), increasing again for the second no-

372 sample test (with peck), to 84%. Once more, the difference in preference between tests

373 was statistically significant, $t(6) = 2.48$, $p = .048$, $d_z = 0.94$.

374   Even though the relative pattern between tests was maintained, preferences in the

375 second run tended to be lower than in the first run. With the repetition of the tests, the

376 pigeons could have begun to learn that no-sample trials never ended with food, which

377 could possibly lead to responding approaching the chance level of 50%. The fact that the

378 difference between runs was the most pronounced in the test that was run last is

379 consistent with this possibility. However, a repeated-measures ANOVA with test type

380 (two levels) and run (two levels) as factors revealed only a main effect of test, $F(1, 6) =$

381 71.25, $p < .001$, $\eta_G^2 = 0.306$, and no main effect of run, $F(1, 6) = 2.43$, $p = .170$, $\eta_G^2 =$

382 0.137. The interaction was also not statistically significant, $F(1, 6) = 1.24$, $p = .307$, η_G^2

383   $= 0.033$. That is, the tendency for the second run to have less extreme preferences was

384   not significant and did not affect the finding that performance differed between the two

385   no-sample tests.

386   The results of the delay test are presented in Figure 2: on the top panel percent

387   correct following each sample as a function of delay, and on the bottom panel the same

388   data is plotted as preference for the "short"/"small" comparison. The introduction of a

389   delay led to a decrease in correct choices following both samples, with accuracy

390   following 2-s samples stabilizing at a higher level than following 6-s samples (Figure 2,

391   top panel). A repeated-measures ANOVA with sample duration (two levels) and delay

392   (five levels) as factors revealed a significant main effect of sample, $F(1, 6) = 5.99$, $p =$

393   $.05$, $\eta_G^2 = 0.157$, and of delay, $F(4, 24) = 21.45$, $p < .001$, $\eta_G^2 = 0.498$. The interaction

394   between factors was not statistically significant, $F(4, 24) = 1.19$, $p = .339$, $\eta_G^2 = 0.085$.

395   That is, while the effect of the delay was similar for both samples (accuracy initially

396   decreasing and then stabilizing), that initial decrease was more pronounced for 6-s

397   sample trials.

398   When plotting the data as preference for the "short"/"small" comparison (Figure

399   2, bottom panel), it becomes clear that, for delays lasting 5 s or longer, choices were

400   indistinguishable between samples. That is, independently of the sample presented,

401   following 5, 10, and 20-s delays, animals showed a preference for the comparison

402   associated with the shorter sample. Even though this preference was not extreme

403   (following the three longest delays, 59.5% of choices were made to the "short"/"small"

404   comparison), it was consistent enough to be significantly above 50%: 95% Confidence

405   Interval for 2-s samples = [52.4%-65.7%], 95% Confidence Interval for 6-s samples =

406   [53.4%-66.6%].

407   Figure 3 plots the proportion of choices to the "short"/"small" comparison as a

408   function of pecks preceding choice during the no-sample tests. The data for each pigeon

409   is shown individually, with the group average on the lower panel of the right column.

410   The unfilled data points refer to training trials: overall (including the peck to initiate a

411   trial), during a 2-s sample there was an average of 6.3 pecks, whereas during a 6-s

412   sample the pigeons pecked an average of 16.2 times. Therefore, the empty point on the

413   left always corresponds to 2-s sample trials, and the empty point on the right always

414   corresponds to 6-s sample trials. The choice preferences for these two data points reflect

415   the contingencies trained: the percentage of choices to the "short"/"small" comparison

416   was high following 2-s samples (average: 90.0%), and low following 6-s samples

417   (average: 9.3%).

418   In an attempt to assess whether the number of pecks during the sample was

419   guiding choice, 2-s and 6-s sample trials were separated between trials that ended with

420   correct and incorrect choices. During 2-s sample presentations, there was an average of

421   5.3 pecks on trials that ended with a correct choice, and an average of 5.0 pecks on trials

422   that ended with an incorrect choice. During 6-s sample presentations, there was an

423   average of 15.2 pecks on trials that ended with a correct choice, and an average of 15.1

424   pecks on trials that ended with an incorrect choice. These differences were not

425   statistically significant (2-s samples: $t(6) = 1.46$, $p = .195$, $d_z = 0.55$; 6-s samples: $t(6) =$

426   0.69, $p = .518$, $d_z = 0.26$).

427   Focusing on the comparison between 2-s sample trials and no-sample tests, the

428   no-sample tests are represented by the filled data points in Figure 3, the test where no

429   pecks preceded choice plotted under "0" on the x axis, and the test where a single peck

430   was required to initiate a trial plotted under "1" on the x axis. The data from the two runs

431   were combined for each test. Even though, as far as number of pecks is concerned, the

432   two no-sample tests are closer to each other than to a 2-s sample trial, choice preferences

433   on 1-peck no-sample trials were generally closer to preferences on 2-s trials than on 0-

434   peck no-sample trials. In fact, with the exception of pigeon PG18 (Figure 3, top panel of

435   right column), there appears to be a discontinuity in preference between the two no-

436   sample tests. Paired-samples t-tests confirmed that, while preference on 1-peck no-

437   sample trials did not differ from preference on short-sample trials, $t(6) = 0.36$, $p = .734$,

438     $d_z = 0.13$, preference between the two no-sample trials did differ, $t(6) = 8.44$, $p < .001$, d_z

439     $= 3.19$.

440

441   Discussion

442   To analyze choice biases that may occur in discrimination tasks where the stimuli

443 fall along a quantitative continuum, pigeons were trained on a symbolic matching-to-

444 sample task with samples that differed in duration, where pecking during sample

445 presentation was required. Subsequently, a delay test and two types of no-sample tests

446 (with and without the requirement of a peck to initiate no-sample trials) were run.

447   One of the aims of the present work was to test whether the commonly-found

448 preference for the comparison associated with the short sample in no-sample testing

449 (Church 1980; Pinto et al. 2017; Pinto and Machado, 2015, 20167; Spetch and Wilkie

450 1983) would be maintained with the pecking requirement. Another of our goals was to

451 compare performance on delay and no-sample testing in this variant of the task, to test an

452 assumption from previous work (Kraemer et al. 1985; Sherburne et al. 1998; Spetch and

453 Wilkie 1983): that the effect of a delay between sample and comparisons is similar to

454 having no sample presented to begin with.

455   A preference for the "short"/"small" comparison in no-sample testing occurred

456 even with the peck requirement, thus not supporting the hypothesis that this preference

457 could be due to the animals failing to attend to some short-sample trials (and learning

458 that when there was no sample – from the animal standpoint – the "short" comparison

459 was correct). In fact, our results were similar to experiments with no response

460 requirement during sample presentation (and also to studies where the sample was not

461 presented on a pecking key, so responding on the sample stimulus was not possible).

462 Table 1 lists no-sample test results from temporal-discrimination tasks. Among those

463   studies, the extent of the preference for the "short" comparison is fairly consistent, even

464   between different species and experimental setups. The results of our no-sample tests

465   without initial peck (Figure 1, data points on the left) – 79% "short" (range: 68 – 85) in

466   the first run and 75% "short" (range: 58 – 93) in the second run – fall within a range

467   similar to these studies.

468   In a different type of task where responding was required – Fetterman and

469   MacEwen (1989) trained four pigeons in a symbolic matching-to-sample task where the

470   samples consisted of 10- and 40-peck (fixed-ratio) requirements –, in no-sample testing

471   there was a strong preference for the comparison associated with the "small", 10-peck

472   sample: 89% of choices were made to this comparison (estimated range: 78 – 100). It is

473   worth noting that the response requirement did not guarantee that the animals were

474   making a numeric discrimination: as a large response requirement takes longer to

475   complete than a small response requirement, responding could have been based on

476   duration, and the preference for the "small" comparison would be another instance of the

477   preference for "short". In any case, a preference for the comparison associated with the

478   sample closest to zero (be it number or duration) is consistently found in no-sample tests,

479   with the fixed-ratio setup (Fetterman and MacEwen, 1989) appearing to yield stronger

480   preferences than other studies. However, given that it is a single study, with four animals,

481   generalizations regarding the effects of different procedures are only tentative at this

482   moment.

483   A choice bias was also found in the delay test, consistent with the choose-

484   short/small effect: following delays of 5 s or longer, the pigeons preferred the

485 "short"/"small" comparison, irrespective of the sample presented (Figure 2). The

486 preference for "short"/ "small" was consistent but not very pronounced, stabilizing at

487 around 60% for the longer delays. As aforementioned, it is commonly assumed that when

488 forgetting occurs during a delay, the animal responds as if no sample had been presented

489 at all. If that is the case, performance on no-sample tests and following the longer delays

490 (where forgetting is most likely) should converge. Although not as clear as in previous

491 studies where results from delay and no-sample tests went in different directions (Church

492 1980; Pinto and Machado 2015, 2017), even though we found a "short"/"small" bias in

493 both tests, the extent of the preferences appears to be distinct. While in the delay test,

494 following the longer delays, 59.5% of choices were to the "short"/"small" comparison, in

495 all no-sample tests preference for "short"/"small" was stronger (79% and 75% without

496 initial peck, 94% and 84% with initial peck), all of these values falling outside the 95%

497 confidence intervals for choices of "short"/"small" following the longer delays. Thus, the

498 present results lend further credence to the possibility that the effect of a delay is not to

499 put an animal in a "no-sample state".

500   In this study, a bias for "short"/"small" was found even with the inter-trial

501 interval (ITI) and the delay signalled differently (the ITI was houselight-illuminated

502 while the delay was spent in darkness), which goes against the predictions of the

503 "confusion hypothesis" (Sherburne et al. 1998) of the choose-short effect. This

504 hypothesis states that the source for choice biases in delay testing is confusion between

505 delay and ITI, when they are signalled in a similar fashion. In those cases, a delay could

506 be interpreted as an ITI before a new trial, leading to disengagement of the current trial –

507   and when a comparison had to be chosen, the animals would behave as if no sample had

508   been presented (which would lead to a bias towards "short").

509   The pecking requirement during sample presentation could have led the pigeons

510   to base their responses not on the duration of the sample, but on the time spent pecking

511   on the sample or on the number of pecks given. If animals were basing their choices on

512   the number of pecks during sample presentation, we could perhaps see a relation between

513   number of pecks and incorrect choices: for instance, the 2-s sample trials where the

514   incorrect comparison was chosen could be the trials where the animal pecked a higher

515   number of times (thus approaching the number of pecks typically given during a 6-s

516   sample). Conversely, an incorrect choice following a 6-s sample could happen in trials

517   where the animal pecked a lower number of times. In our data, the number of pecks to

518   each sample did not allow predicting which choice would follow: there was no

519   significant difference between the number of pecks given before correct and incorrect

520   choices (for 2-s samples, 5.3 vs 5.0 pecks, respectively; for 6-s samples, 15.2 vs 15.1

521   pecks). However, this analysis does not conclusively set aside the possibility that the

522   animals were tracking the number of pecks: the restrictions imposed on pecking (at least

523   one peck every 1.5 s) may have limited the range of pecks that could be given during

524   each sample presentation.

525   Disentangling time and number may indeed prove to be challenging. While

526   posing as different strategies to interpret the world, some studies propose that when in a

527   situation where both duration and number co-vary, more often than not, individuals

528   process both in a rather automatic fashion (e.g., Meck and Church 1983 (rats); Meck et

al. 1985 (rats); Roberts and Mitchell 1994 (pigeons)). Droit-Volet and colleagues (2003) conducted an experiment with children (5- and 8-years old) and adults in which time and number co-varied. In a typical bisection task, participants were either instructed to attend to the stimuli duration while disregarding the number of stimuli, or to process the number of stimuli while ignoring stimuli duration. Results showed that time and number were not processed separately i.e., individuals were not able to exclusively pay attention to only one of the stimuli features. This was particularly true for the 5-year-olds but interference between duration and number was also found in the older children and adults, even if not as pronounced.

Additionally, the nature of the interference between time and number is not clear. In the aforementioned study (Droit-Volet et al. 2003), results pointed to the fact that the relationship between time and number is asymmetrical – it appears as though number interferes more with time than time with number. This asymmetry has propelled discussion that culminated in the emergence of two explanatory theories – theory of magnitude (ATOM) (Walsh 2003), and metaphor theory. ATOM proposes that time, space, and number are all processed by a unified system, relying on common neural structures (e.g., Cantlon et al. 2009; Feigenson 2007; Meck and Church 1983). Given that time, space and number are represented together, there is no reason not to believe that these dimensions interfere with one another in a symmetrical way. The metaphor theory, however, claims that the relationship between time, space, and number is asymmetrical. Merritt and collaborators (2010) found that interference between space and time was asymmetrical in humans but not in monkeys or pigeons (DeCorte et al. 2017). Literature

551 on (a)symmetry between number and time in non-humans is scarce and future research

552 should dwell on that. What we do know is that, aside from the direction of interference,

553 time and number go hand in hand (Roberts and Mitchell 1994), much like in the present

554 experiment.

555   One implication of the animals responding on the basis of how long they pecked

556 or of how many pecks were given in a trial can be seen in the no-sample tests: the no-

557 sample test with a peck to initiate the trial would functionally be a "very short/small"

558 sample: in those trials the animals pecked for a very short duration (a single peck).

559 Following the same reasoning, a no-sample trial without an initiating peck could be seen

560 as the de facto no-sample trial: no stimulus was presented and no pecks occurred. This

561 difference could lead to dissimilar choice behavior. Alternatively, if the animals were

562 exclusively timing the duration of the sample keylight, performance should be similar

563 between the two tests, given that no sample was presented in both no-sample tests.

564   We found that performance in the two no-sample tests was different: when

565 animals had to peck a key to initiate a no-sample trial, preference for the "short"/"small"

566 comparison was more pronounced than when no such initiating peck was required

567 (Figure 1). These results may suggest that the absence of a sample may not belong to the

568 same continuum as non-zero samples, but a clearer picture can be seen when no-sample

569 trials are plotted together with 2-s and 6-s sample trials, to obtain a generalization

570 gradient.

571   Generalization gradients along the temporal dimension tend to not be

572 symmetrical, approaching the shape of a step function (e.g., Spetch and Cheng 1998;

573 Vieira de Castro and Machado 2012). To illustrate, picture a task where a short stimulus

574 is an S- for a given response and a long stimulus is an S+ for the same response. In a

575 generalization test (where new, untrained, durations are also presented), proportion of

576 occurrence or rate of the response is low following the S-, high following the S+, and

577 tends to remain high for stimuli longer than the S+ (as the stimulus presented moves

578 along the temporal continuum in the direction opposite S-). Hence the asymmetrical,

579 step-like generalization function. The implication for the present study is that, if the data

580 conform to a generalization gradient, proportion of "short"/"small" choices following

581 samples shorter than 2s should remain at a level similar to 2-s samples. That prediction is

582 confirmed for no-sample trials with an initiating peck (the putative "very-short sample"):

583 average choice proportions between that no-sample test (89.1%) and 2-s samples (90%)

584 are very similar (Figure 3). However, the pattern is broken for the no-sample test without

585 a peck to initiate the trial: there is a decrease in proportion of "short"/"small" choices

586 (77.3%).

587   Another result that does not conform to what is expected by stimulus

588 generalization is that the distance between stimuli does not appear to predict preference:

589 the gap between the 2-s sample (Figure 3, leftmost empty point) and the no-sample tests

590 (Figure 3, filled points) is greater than the distance between the two no-sample tests, but

591 it is between the no-sample tests where the largest difference in preference occurs. This

592 pattern, similar to a study that had no peck requirements (Pinto and Machado 2015),

593 could be suggestive that the sample-stimulus continuum does not include a zero, that is,

594   that the no-sample test without any peck does not fall along the same continuum as the

595   other trials where pecking was involved.

596   However, given that the birds are required to peck to initiate a trial, some

597   alternative possibilities are to be considered. Firstly, the no-sample trials with no initial

598   peck, by lacking a "start" response, are considerably different from all other trials, which

599   could result in generalization decrement, and thus explain the lower preference for the

600   "short"/"small" comparison found in this test. Secondly, given the absence of a "start"

601   stimulus, the pigeons may respond on no-sample tests with no initial peck not as a new

602   trial, but as if it is still part of the previous trial. In that case, the ITI would function as an

603   additional 30-s delay before the animal has a chance to choose again. Even though

604   preferences for "short"/"small" are significantly stronger following no-sample tests with

605   no initial peck (75% and 79% of choices) than following long delays (approximately

606   60% of choices), the pigeons could be alternating between responding as if the

607   comparisons referred to the previous trial (leading to a weaker preference for

608   "short"/"small") or a new trial (leading to a stronger preference for "short"/"small") –

609   for an illustration on how animals may alternate between different stimuli to base their

610   choices, please refer to Pinto et al. 2017.

611   In closing, for stimulus dimensions that fall along a quantitative continuum,

612   choice biases are not uncommon, and these biases may be valuable in revealing how

613   stimuli are perceived or what learning strategies animals may employ. In the present

614   study we tested the possibility – which was not confirmed – that the typical preference

615   for "short" in no-sample tests was the product of accidental learning. The mechanisms

616   behind this preference still remain to be fully identified: as our results (and Table 1)

617   show, preferences following no-sample trials are reliably above chance levels, but are

618   perhaps too low for what would be expected by stimulus generalization. Our results also

619   underline that open questions remain on the role and effect delays may have. As such,

620   further exploration of choice biases may prove fruitful.

621   Reference List

622   Bakeman R (2005) Recommended effect size statistics for repeated measures

623   designs. Behav Res Methods, 37:379–384. https://doi.org/10.3758/BF03192707

624   Blough DS (1959) Delayed matching in the pigeon. J Exp Anal Behav 2(2):51–

625   160. https://doi.org/10.1901/jeab.1959.2-151

626   Cantlon JF, Platt ML, Brannon EM (2009) Beyond the number domain. Trends

627   Cogn Sci 13(2):83–91. https://doi.org/10.1016/j.tics.2008.11.007

628   Church RM (1980) Short-term memory for time intervals. Learn Motiv

629   11(2):208–219. https://doi.org/10.1016/0023-9690(80)90013-2

630   Cohen J (1988) Statistical power analysis for the behavioral sciences. Lawrence

631   Earlbaum Associates, Hillsdale, NJ

632   D'Amato MR (1973) Delayed matching and short-term memory in monkeys. In:

633   Bower GH (ed) Psychology of learning and motivation: Advances in research and theory,

634   vol. 40. Academic Press, New York, pp 227–269

635   De Corte BJ, Navarro VM, Wasserman EA (2017) Non-cortical magnitude

636   coding of space and time by pigeons. Curr Biol 27(23):R1264-R1265.

637   https://doi.org/10.1016/j.cub.2017.10.029

638   Droit-Volet S, Clément A, Fayol M (2003) Time and number discrimination in a

639   bisection task with a sequence of stimuli: A developmental approach. J Exp Child

640   Psychol 84(1):63-76. https://doi.org/10.1016/S0022-0965(02)00180-7

641   Feigenson LR (2007) The equality of quantity. Trends Cogn Sci 11(5):185–187.

642   https://doi.org/10.1016/j.tics.2007.01.006

643   Fetterman JG, MacEwen D (1989) Short-term memory for responses: the

644   "choose-small" effect. J Exp Anal Behav 52(3):311–324.

645   https://doi.org/10.1901/jeab.1989.52-311

646   Fetterman JG, MacEwen D (2003) Acquisition and retention in compound

647   matching with hue and peck number elements. Learn Motiv 34:354–371.

648   https://doi.org/10.1016/S0023-9690(03)00034-1

649   Grant DS (2001) Memory for empty time intervals in pigeons. Anim Learn Behav

650   29:293–301. http://dx.doi.org/10.3758/BF03192896

651   Grant DS, Spetch ML (1991) Pigeons' memory for event duration: Differences

652   between choice and successive matching tasks. Learn Motiv 22(1-2):180–199.

653   https://doi.org/10.1016/0023-9690(91)90022-Z

654   Grant DS, Spetch ML (1993) Analogical and nonanalogical coding of samples

655   differing in duration in a choice-matching task in pigeons. J Exp Psychol Anim Behav

656   Processes 19(1):15–25. https://doi.org/10.1037/0097-7403.19.1.15

657   Grant DS, Talarico DC (2004) Processing of empty and filled time intervals in

658   pigeons. Learn Behav 32:477–490. http://dx.doi.org/10.3758/BF03196043

659   Hunter WS (1913) The delayed reaction in animals and children. Anim Behav

660   Monogr 6(2):1–86. https://doi.org/10.1037/13779-000

661   Kelly R, Spetch ML (2000) Choice biases in delayed matching-to-sample

662   duration with pigeons: Manipulations of ITI and delay illumination. Q J Exp Psychol

663   53B(4):309–323. http://dx.doi.org/10.1080/713932737

664   Kraemer PJ, Mazmanian DS, Roberts WA (1985) The choose-short effect in

665   pigeon memory for stimulus duration: Subjective shortening versus coding models. Anim

666   Learn Behav 13(4):349–354. http://dx.doi.org/10.3758/BF03208009

667   Lakens D (2013) Calculating and reporting effect sizes to facilitate cumulative

668   science: a practical primer for t-tests and ANOVAs. Frontiers Psychol 4:1–12.

669   https://doi.org/10.3389/fpsyg.2013.00863

670   Meck WH, Church RM (1983) A mode control model of counting and timing

671   processes. J Exp Psychol: Anim Behav Process 9(3):320–334.

672   https://doi.org/10.1037/0097-7403.9.3.320

673   Meck WH, Church RM, Gibbon J (1985) Temporal integration in duration and

674   number discrimination. J Exp Psychol: Anim Behav Process 11(4):591–597.

675   https://doi.org/10.1037/0097-7403.11.4.591

676   Merritt DJ, Casasanto D, Brannon EM (2010) Do monkeys think in metaphors?

677   Representations of space and time in monkeys and humans. Cogn 117(2):191-202.

678   https://doi.org/10.1016/j.cognition.2010.08.011

679   Olejnik S, Algina J (2003) Generalized eta and omega squared statistics:

680   Measures of effect size for some common research designs. Psychological Methods, 8:

681   434–447. https://doi.org/10.1037/1082-989X.8.4.434

682   Pinto C, Fortes I, Machado A (2017) Joint stimulus control in a temporal

683   discrimination task. Anim Cogn 20(6):1129–1136. http://dx.doi.org/10.1007/s10071-

684   017-1130-6

685   Pinto C, Machado A (2011) Short-term memory for temporal intervals:

686   Contrasting explanations of the choose-short effect in pigeons. Learn Motiv 42(1):13–25.

687   https://doi.org/10.1016/j.lmot.2010.05.001

688   Pinto C, Machado A (2015) Coding in pigeons: multiple-coding versus single-

689   code/default strategies. J Exp Anal Behav 103(3):472–483.

690   https://doi.org/10.1002/jeab.153

691   Pinto C, Machado A (2017) Unraveling sources of stimulus control in a temporal

692   discrimination task. Learn Behav 45(1):20–28. http://dx.doi.org/10.3758/s13420-016-

693   0233-2

694   Roberts WA, Mitchell S (1994) Can a pigeon simultaneously process temporal

695   and numerical information? J Exp Psychol: Anim Behav Process 20(1):66–78.

696   https://doi.org/10.1037/0097-7403.20.1.66

697   Santi A, Hornyak S, Miki A (2003) Pigeons' memory for empty and filled time

698   intervals signaled by light. Learn Motiv 34(3):282–302. https://doi.org/10.1016/S0023-

699   9690(03)00021-3

700   Santi A, Ross L, Coppa R, Coyle J (1999) Pigeons' memory for empty time

701   intervals marked by visual or auditory stimuli. Anim Learn Behav 27:190–205.

702   http://dx.doi.org/10.3758/BF03199675

703   Sherburne LM, Zentall TR, Kaiser DH (1998) Timing in pigeons: The choose-

704   short effect may result from pigeons' "confusion" between delay and intertrial intervals.

705   Psychon Bull Rev 5(3):516–522. http://dx.doi.org/10.3758/BF03208831

706   Spetch ML (1987) Systematic errors in pigeons' memory for event duration:

707   Interaction between training and test delay. Anim Learn Behav 15(1):1–5.

708   http://dx.doi.org/10.3758/BF03204897

709   Spetch ML, Cheng K (1998) A step function in pigeons' temporal generalization

710   in the peak shift task. Anim Learn Behav 26(1):103–118.

711   https://doi.org/10.3758/BF03199165

712   Spetch ML, Grant DS (1993) Pigeons′ Memory for Event Duration in Choice and

713   Successive Matching-to-Sample Tasks. Learn Motiv 24(2):156–174.

714   https://doi.org/10.1006/lmot.1993.1010

715   Spetch ML, Wilkie DM (1982) A systematic bias in pigeons' memory for food

716   and light durations. Behav Anal Lett 2(5):267–274.

717   Spetch ML, Wilkie DM (1983) Subjective shortening: a model of pigeons'

718   memory for event duration. J Exp Psychol Anim Behav Process 9(1):14–30.

719   https://doi.org/10.1037/0097-7403.9.1.14

720   Vieira de Castro AC, Machado A (2012) The interaction of temporal

721   generalization gradients predicts the context effect. J Exp Anal Behav 97(3): 263–279.

722   http://dx.doi.org/10.1901/jeab.2012.97-263

723   Walsh V (2003) A theory of magnitude: common cortical metrics of time, space

724   and quantity. Trends Cogn Sci 7(11):483-488. https://doi.org/10.1016/j.tics.2003.09.002

725   Wixted JT (1989) Nonhuman short- term memory: A quantitative reanalysis of

726   selected findings. J Exp Anal Behav 52(3):409–426.

727   https://doi.org/10.1901/jeab.1989.52-409

728   Figure Captions

729   **Fig. 1** Mean (with SEM) percent of choices to the comparison associated with the

730   short (2s) sample on no-sample tests. The data points on the left refer to the tests where

731   no key peck was required to initiate no-sample trials, and the data points on the right

732   refer to the tests where one key peck was required to initiate no-sample trials. The filled

733   data points refer to the first run of the tests, and the empty data points refer to the

734   replication of the tests

735   **Fig. 2** Top panel: Mean (with SEM) percent correct responses to 2-s (empty data

736   points) and 6-s (filled data points) samples as a function of delay duration. Bottom panel:

737   Mean percent of choices to the comparison associated with the short (2s) sample

738   following 2-s (empty data points) and 6-s (filled data points) samples as a function of

739   delay duration. The dashed line represents 50% of choices

740   **Fig. 3** Mean percent of choices to the comparison associated with the short (2s)

741   sample plotted against the number of key pecks before choice for each pigeon, during the

742   no-sample test sessions. The filled data points refer to no-sample test trials: the leftmost

743   filled point for when no peck was required, and the rightmost filled point for when a

744   single peck was required. The empty data points refer to training trials with a sample: the

745   leftmost empty point corresponds to 2-s sample trials, and the rightmost empty point

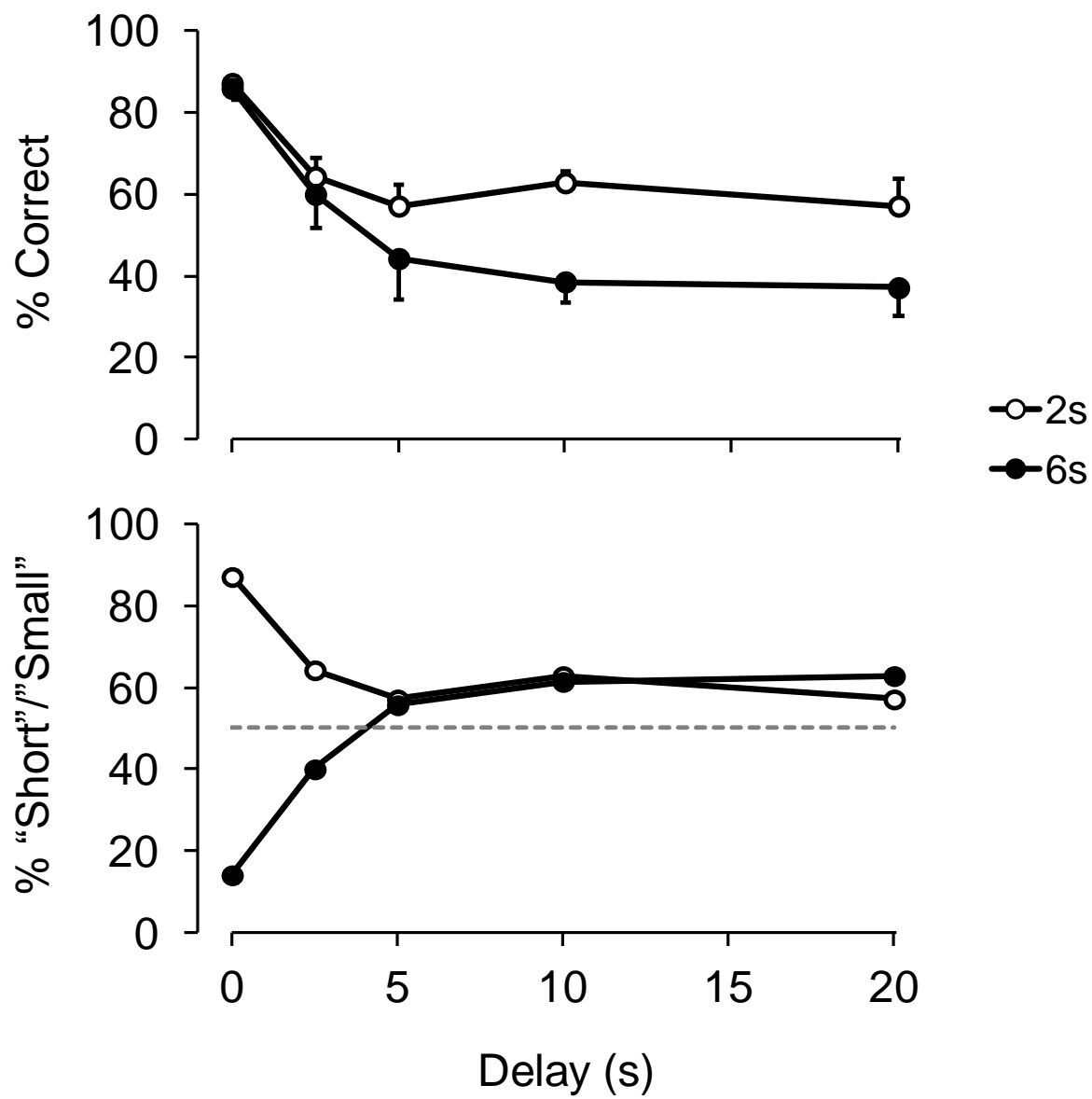746   corresponds to 6-s sample trials

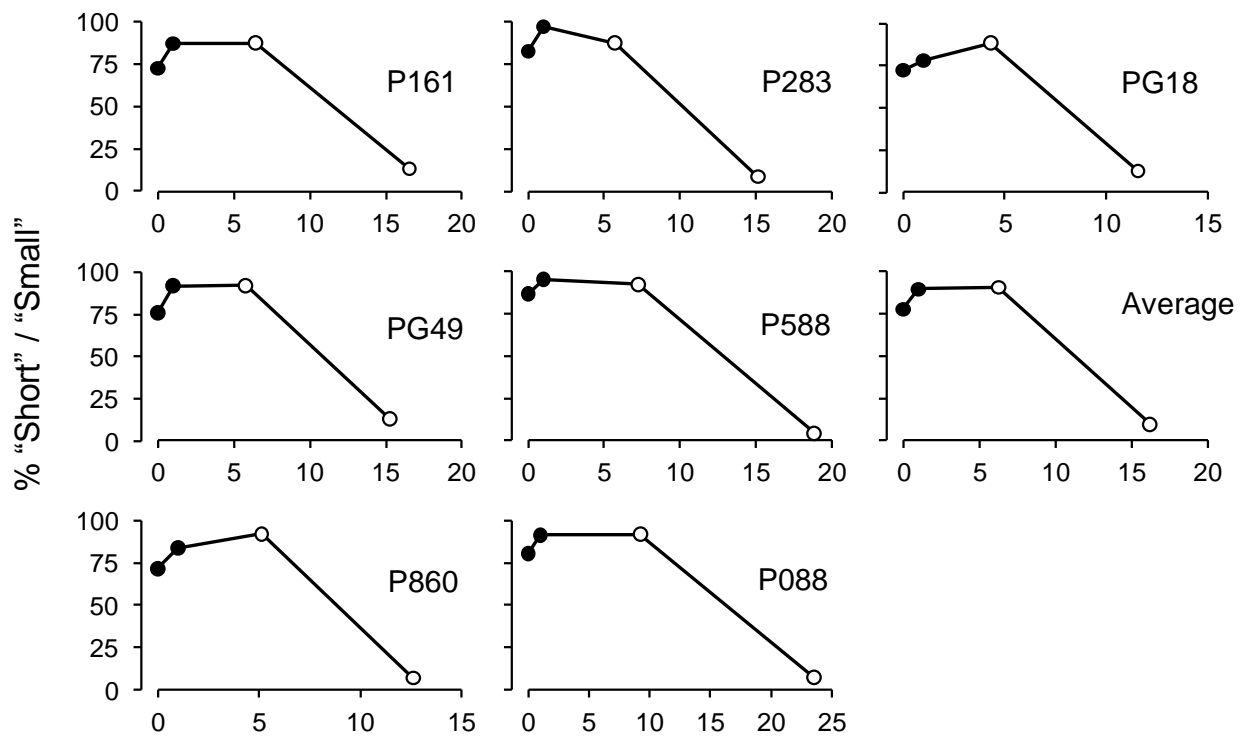747

748
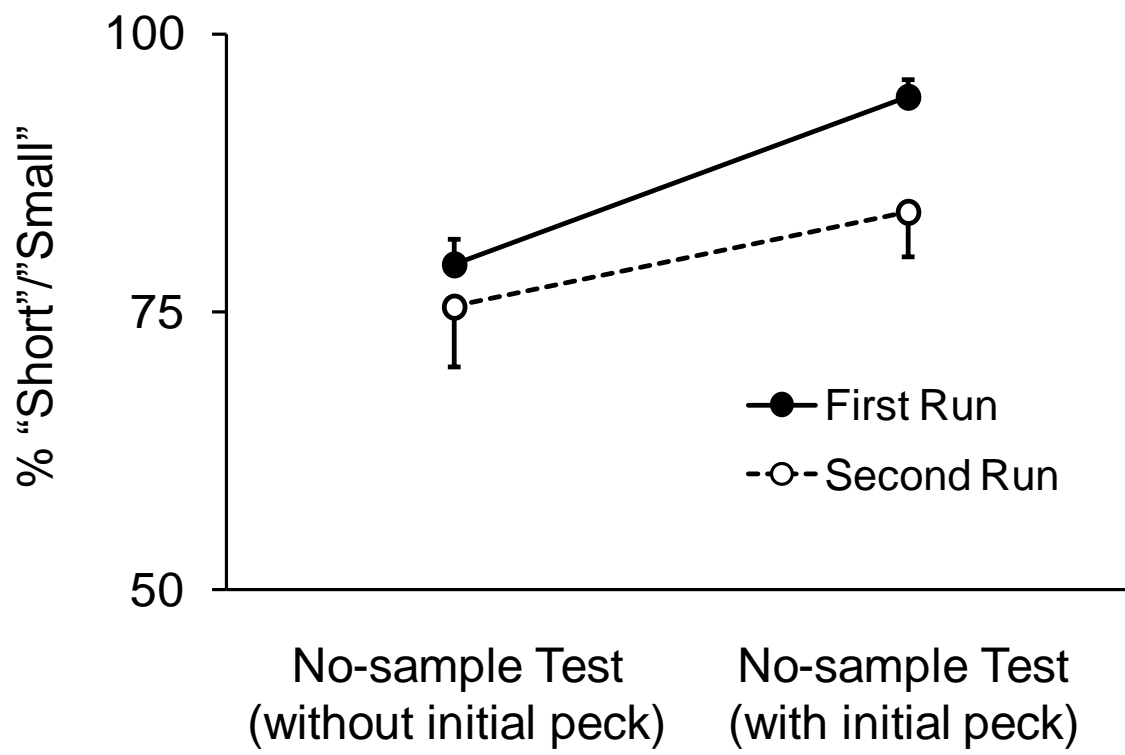
749

750

751　Pinto & Sousa - Figure 1

752

753

754   Pinto & Sousa - Figure 2

755

756

% "Short" / "Small"

Number of pecks during sample (including trial-initiating peck)

757

758

759   Pinto & Sousa - Figure 3

760

761   Table 1. Choices of the comparison correct following the short sample ("short") in no-

762   sample tests

Reference	Species	Sample Stimulus	% "short" (range)
Church (1980, Experiment 3A)	Albino Norway rat (*Rattus norvegicus*)	Houselight	79%
Pinto, Fortes and Machado (2017)	Pigeon (*Columba livia*)	Keylight	73% (45 – 93)
Pinto and Machado (2015)	Pigeon (*Columba livia*)	Keylight	72 % (55 – 80)
		Keylight	68% (50 – 77)
Pinto and Machado (2017)	Pigeon (*Columba livia*)	Keylight	68% (47 – 83)
Spetch and Wilkie (1983, Experiment 4)	Pigeon (*Columba livia*)	Houselight or Keylight	77% (66 – 92) [a]

[a] values estimated from figure

763