Evidence for Detection of One Duration Sample and Default Responding to Other Duration Samples by Pigeons May Result From an Artifact of Retention-Test Ambiguity

Thomas R. Zentall, Emily D. Klein, and Rebecca A. Singer
University of Kentucky

S. C. Gaitan and J. T. Wixted (2000) noted that when pigeons are trained on a conditional discrimination to associate 1 duration sample with 1 comparison and 2 other duration samples with a 2nd comparison, they detect only the single duration, and on trials involving either of the 2 other duration samples, they respond to the other comparison by default. In 2 experiments, the authors show instead that pigeons tend to treat the retention intervals (such as those used by Gaitan and Wixted) as intertrial intervals, and thus, they tend to treat all trials with a delay as 0-s sample trials. The authors tested this hypothesis by showing that divergent retention functions do not appear when the retention interval is discriminably different from the intertrial interval.

Pigeons can use the duration of a conditional or sample stimulus (e.g., 2 s vs. 10 s) as the basis for choosing the appropriate comparison stimulus (Spetch & Wilke, 1983). By inserting a delay between the offset of the sample and the onset of the comparison, researchers have been able to study the ability of animals to remember intervals of time. Surprisingly, with increasing delay, matching accuracy tends to remain high on trials involving the shorter duration sample and declines rapidly, often to well below chance level, with increasing delay on trials involving the longer duration sample (Spetch, 1987). This finding of divergent retention functions has become known as the choose-short effect because at longer delays it can be described as a bias to choose the comparison associated with a short sample, even on long-duration sample trials.

To account for this finding, Spetch (1987) proposed that memory for time decreases in subjective duration with time since the duration was experienced. Thus, on long-duration sample trials, as the subjective duration of the sample decreases, matching accuracy eventually falls below chance because the subjective duration of the long-duration sample appears more similar to the reference memory for the short-duration sample in training (see Spetch & Wilke, 1983).

Recently, Gaitan and Wixted (2000) used duration-sample matching to examine pigeons’ ability to efficiently represent samples of different durations. They trained pigeons to associate 0- and 2-s samples with one comparison stimulus and 10-s samples with the other comparison stimulus (Experiment 2). When they tested the pigeons with delays inserted between the offset of the sample and the onset of the comparison stimuli, they found, as expected, evidence of a choose-short effect. As the retention interval increased, they found that matching accuracy on 0- and 2-s sample trials remained high, and accuracy on 10-s sample trials declined with increasing delay to below chance level.

In a follow-up experiment, however, when Gaitan and Wixted (2000) associated samples of 0 s and 10 s with one comparison stimulus and samples of 2 s with the other comparison stimulus, surprising results were found. According to the subjective-shortening hypothesis, with increasing delay, matching accuracy on 0-s sample trials should remain high, which it did. Similarly, this hypothesis also predicts that matching accuracy on 2-s sample trials should decline rapidly with increasing delay. This should occur because the delays cause the subjective duration of the 2-s sample to appear more similar to the reference memory for the 0-s sample, and errors should result. This result was found as well. What was surprising was that matching accuracy on 10-s delay trials remained high with increasing delay. This result is contrary to the subjective-shortening hypothesis, which predicts that as the delay increases, the subjective duration of the 10-s sample should begin to appear more similar to the reference memory for the 2-s sample, resulting in an incorrect choice of the comparison associated with the 2-s sample. Eventually, as the delay increases still further, the subjective duration of the 10-s sample should appear more similar to the reference memory for the 0-s sample, and matching accuracy might be expected to rise again. However, Gaitan and Wixted (Experiment 3) found that as the delay increased, there was little evidence on 10-s sample trials for the predicted U-shaped function.

Gaitan and Wixted (2000) interpreted the results of these experiments and, in particular, the absence of errors with increasing delay on 10-s sample trials as evidence that the pigeons were treating the task as a detection task. That is, the pigeons had learned to detect whether one of the three samples had been presented. If it had been presented, they would respond to the appropriate comparison; otherwise, they would respond to the other comparison by default. Treating a task as a detection task has often been referred to as the development of a single-code/default...
coding strategy (Grant, 1991; Sherburne & Zentall, 1993). The development of such a coding strategy is an appealing hypothesis because it represents an efficient process. The pigeon need only remember the sample when one of the three samples is presented, that is, the one sample that appears in a one-to-one relation with its comparison. When 0- and 2-s samples are associated with one comparison and 10-s samples are associated with the other comparison, the pigeons need only remember the sample on trials involving the 10-s sample to achieve high levels of matching accuracy. Similarly, when 0- and 10-s samples are associated with one comparison and 2-s samples are associated with the other comparison, the pigeons need only remember the sample on trials involving the 2-s sample to achieve high levels of matching accuracy. In both cases, with increasing delay, only on trials involving the single to-be-remembered sample would one expect to see a decline in matching accuracy because this would be the only case in which the pigeons should experience forgetting.

The purpose of the present experiments was to replicate the finding by Gaitan and Wixted (2000, Experiment 3) and then to test an alternative account of their results, namely, that the divergent retention functions found with differential duration samples may result from an artifact of the procedure. Specifically, it is possible that the similarity of the intertrial interval (ITI) to the novel delays results in all samples being treated as 0-s samples (no-sample) on long-delay trials.

Experiment 1

Method

Subjects

The subjects were 4 White Carneaux (Columba livia) pigeons that were retired breeders (5–8 years of age) purchased from the Palmetto Pigeon Farm (Sumter, SC). The pigeons were individually housed in wire cages and were maintained at 80% to 85% of their free-feeding weights for the duration of the experiment. The pigeons had free access to water and grit in their home cages. The pigeon colony room was maintained on a 12:12-hr light–dark cycle. The pigeons were cared for in accordance with University of Kentucky animal care guidelines. All pigeons had previous experience in unrelated studies involving simple simultaneous discriminations and two-sample, two-alternative, matching-to-sample discriminations.

Apparatus

The experiment was conducted in an LVE/BRS (Laurel, MD) sound-attenuating operant chamber. The inside measurements of the chamber were 32.0 cm high, 32.0 cm across the response panel, and 28.0 cm from the response panel to the back wall. Three square response keys (2.5 × 2.5 cm) were aligned horizontally on the response panel and were separated by 0.8 cm. The bottom edge of the response keys was 25.5 cm from the wire mesh floor. A 12-stimulus in-line projector (Industrial Electronics Engineering, Van Nuys, CA) with 28 V, 0.1 A lamps (GE 1820) was mounted behind each response key. The left and right response keys could be illuminated with red and green hues (Kodak Wratten filters nos. 26 and 60, respectively). Only the left and right response keys were used.

The sample stimulus consisted of a shielded houselight (GE 1820) located above the center response key, 3.5 cm from the ceiling of the chamber. The houselight lamp was surrounded by a blue filter (Kodak Wratten filter no. 38A). A second, unfiltered houselight (GE 1820) was mounted in the center of the chamber ceiling. A rear-mounted feeder provided mixed grain reinforcement (Purina Pro Grains with the feeder illuminated) through a 5.1 × 5.5 cm aperture centered horizontally on the response panel and vertically midway between the response keys and the floor of the chamber. Reinforcement consisted of 2-s access to mixed grain. The experiment was controlled and data collected by a computer, with a Med Associates (St. Albans, VT) interface, located in an adjacent room.

Procedure

Training. Because the pigeons were not experimentally naive, we began training immediately with duration-sample matching. At the start of each trial, the blue houselight, which served as the sample, was lit for 0, 2, or 10 s. Offset of the sample was followed by a 1-s dark retention interval and subsequent presentation of the red and green comparison stimuli on the side response keys. For half of the pigeons, choice of the red comparison was reinforced following a 2-s sample, and choice of the green comparison was reinforced following 0- and 10-s samples. For the remaining pigeons the reinforced choices were reversed. A 15-s ITI, during which the houselight was not lit, followed all comparison choices. Each training session consisted of 64 trials, with thirty-two 2-s sample trials, sixteen 10-s sample trials, and sixteen 0-s sample trials per session. Training sessions were conducted once a day, 6 days a week.

During training, a correction procedure was in effect. If a pigeon made an incorrect choice, after a 15-s ITI, the trial was repeated to a maximum of five times. On a repeated trial, a correct response was not reinforced, but the feeder light was turned on for 2 s.

The pigeons were trained to a criterion of 80% correct or better for each of the three sample types for five consecutive sessions. When criterion was attained, the correction procedure was discontinued, and training continued to a criterion of 75% correct for two consecutive sessions.

Testing. Testing sessions were similar to training sessions with the exception of the duration of the retention interval. During testing, the duration of the retention interval was 1, 2, 4, or 12 s, sixteen trials of each, with eight 2-s sample trials, four 0-s sample trials, and four 10-s sample trials at each retention interval, randomly presented in each testing session. There were 10 testing sessions conducted, one each testing day, with 2 days of training between testing sessions.

Results

The pigeons reached criterion in an average of 33.8 training sessions. For purposes of data analysis, we pooled testing data over the 10 testing sessions. During testing, the pigeons’ matching accuracy on 0- and 10-s sample trials remained consistently high with increasing retention interval, whereas matching accuracy on 2-s sample trials declined rapidly with increasing retention interval to a level well below chance. The test-session data from Experiment 1 are presented in Figure 1.

We performed a two-way, repeated measures analysis of variance (ANOVA) on the testing data, with sample type (0, 2, & 10 s) and retention interval (1, 2, 4, & 12 s) as factors. The results of the analysis indicated that there was a significant effect of sample type, F(2, 6) = 119.00, p < .01, and a significant effect of retention interval, F(3, 9) = 30.11, p < .01. It is important to note that there was also a significant Sample-Type × Retention-Interval interaction, F(6, 18) = 30.18, p < .01. As can be seen in Figure 1, the interaction can be attributed to divergent retention functions for samples of 2 s versus those of 0 s and 10 s. To further isolate the source of the interaction, we performed a planned comparison involving samples associated with each of the two comparisons and delay as factors. Again, the Sample-Type × Delay interaction was significant, F(3, 18) = 52.25, p < .01. Thus, as can be seen in Figure 1, the slope of the retention function on 2-s delay trials was substantially steeper than on 0- and 10-s delay trials.
hybrid one-to-one/many-to-one matching with 2-s samples (squares) asso-
longer the delay, the more it becomes similar to the ITI and thus,
comparison stimuli after a period of darkness. Furthermore, the
retention test, all trials with a delay involve the appearance of the
stimuli appear after a period of darkness (the ITI). During the
comparison associated with the 0-s sample when the comparison
important ambiguity. In training, pigeons learn to choose the
retention intervals. But this sequence of events may result in an
ples. This generally means that the ITIs are dark, as are the
duration samples, illumination of the houselight defines the sam-
trial). Specifically, in most research involving discrimination of
samples steeply declined. These data are consis-
tent with the hypothesis that the pigeons develop an efficient
strategy in which they detect only the 2-s sample, and on 0- and
10-s sample trials, they choose the alternative comparison stimulus
by default.
Our results are also consistent with the results of Gaitan and
Wixted's (2000) Experiment 3. When pigeons were trained on a
hybrid one-to-one/many-to-one matching task involving 0- and
10-s samples associated with one comparison stimulus and 2-s
samples associated with the other, the pigeons produced divergent
retention functions; matching accuracy on trials involving 0- and
10-s samples remained consistently high, whereas accuracy on
trials involving 2-s samples steeply declined. These data are con-
sistent with the hypothesis that the pigeons develop an efficient
strategy in which they detect only the 2-s sample, and on 0- and
10-s sample trials, they choose the alternative comparison stimulus
by default.
Conversely, it is also possible that unrelated to the duration of
the sample, there is something special about the sequence of events
on 0-s sample trials that makes all trials involving a delay longer
than 1 s appear to be similar to a no-sample trial (a no-sample
trial). Specifically, in most research involving discrimination of
duration samples, illumination of the houselight defines the sam-
ples. This generally means that the ITIs are dark, as are the
retention intervals. But this sequence of events may result in an
important ambiguity. In training, pigeons learn to choose the
comparison associated with the 0-sample when the comparison
stimuli appear after a period of darkness (the ITI). During the
retention test, all trials with a delay involve the appearance of the
comparison stimuli after a period of darkness. Furthermore, the
longer the delay, the more it becomes similar to the ITI and thus,
the more similar the conditions are to the 0-sample trials experi-
enced in training. According to this ambiguity hypothesis, on all
retention-test trials, pigeons should show a bias to choose the
comparison that was correct on 0-sample trials in training, as was
found by Gaitan and Wixted (2000) and in our Experiment 1.
In Gaitan and Wixted's (2000) Experiment 3 and in our Ex-
periment 1, 0- and 10-s samples were associated with one comparison
stimulus, and 2-s samples were associated with the other; the
ambiguity hypothesis predicts that performance on retention tests
should have declined with increasing delay on 2-s sample trials but
remained high on 0- and 10-s sample trials, not because of any
failure to detect 10-s samples but because long retention intervals
make all test trials appear similar to 0-sample training trials. By
the same token, in Gaitan and Wixted's Experiment 2, when 0- and
2-s samples were associated with the same comparison stimulus, it
would have been performance on 10-s sample retention tests that
would have declined because increasing retention intervals made
them appear more similar to 0-sample training trials.

Thus, there are two hypotheses. In the case of the single-dur-
tion detection hypothesis, when pigeons encounter a hybrid
one-to-one/many-to-one matching task involving duration sam-
ples, they develop an efficient detection strategy by detecting
the one-to-one sample and choosing the comparison associated
with the many-to-one samples by default.
Alternatively, in the case of the ITI/delay similarity or ambigu-
ity hypothesis, on all delay trials involving a retention interval
longer than that experienced during training, pigeons should in-
creasingly treat the delay as an ITI and thus, with increasing delay,
should choose the comparison associated with 0-s samples during
training. This hypothesis accounts for the results of our experiment
as well as those of Gaitan and Wixted (2000).

Experiment 2
A better way to distinguish between the two hypotheses is to ask
how clearly distinguishing between the ITIs and the delays might
affect the retention functions. That is, if the ITIs were brightly lit
instead of dark, and the delays remained dark, would the retention
functions be affected? According to the single-duration detection
hypothesis, there is no obvious reason why lighting the ITI should
affect the slopes of the retention functions. On the other hand,
according to the ITI/delay similarity hypothesis, if the two inter-
vals were disambiguated, it is possible that the pigeons would
detect all three sample durations independently, and divergent
retention functions would no longer be found.
Thus, the purpose of Experiment 2 was to assess the effect on
the retention functions of using ITIs that were distinctive from the
delays. For pigeons in Group ITI-lit, the second houselight, which
was mounted in the center of the chamber ceiling, was lit during all
ITIs during training and retention testing. For pigeons in Group
ITI-lit, the retention intervals remained dark, as they were in
Experiment 1. For Group ITI-dark, the ITIs were dark, as were the
delay intervals. In all other respects, the design of Experiment 2
was similar to the design of Gaitan and Wixted's (2000) Experi-
ment 2. That is, samples of 0- and 2-s were associated with choice
of one comparison, and samples of 10-s were associated with
choice of the other comparison.
Method

Subjects and Apparatus

The 8 pigeons used in Experiment 2 were similar to those used in Experiment 1. The apparatus was the same as that used in Experiment 1.

Procedure

The procedure used in Experiment 2 was similar to that used in Experiment 1 with the following exceptions. Experiment 2 involved two groups of pigeons. The first group, Group ITI-dark, was treated the same as the pigeons in Experiment 1 except that for half of the pigeons, choice of the red comparison was reinforced following 10-s samples, and choice of the green comparison was reinforced following 0- and 2-s samples. For the remaining pigeons, the contingencies were reversed. To maintain an equal number of trials on which red and green comparisons were potentially reinforced, in Experiment 2, each training session consisted of thirty-two 10-s sample trials, sixteen 0-s sample trials, and sixteen 2-s sample trials. Testing sessions were the same as in Experiment 1 except that there were eight 10-s sample trials, four 0-s sample trials, and four 2-s sample trials at each retention interval, randomly presented in each testing session. The second group of pigeons in Experiment 2, Group ITI-lit, was treated the same as pigeons in Group ITI-dark except that the second houselight (unfiltered), which was mounted in the center of the ceiling, was illuminated during all ITIs. This unfiltered (white) houselight was substantially brighter than the blue houselight used to signal the sample duration, and its location also differentiated it from the blue houselight.

Results

The pigeons reached criterion in an average of 40.0 training sessions. Pigeons in Group ITI-dark acquired duration-sample matching somewhat faster (32.8 sessions) than pigeons in Group ITI-lit (47.2 sessions), but this difference was not statistically reliable, $F(1, 6) = 2.02, p > .05$.

Retention Testing for Group ITI-Dark

Again, for purposes of data analysis, we pooled the data from testing sessions over the 10 testing sessions. During testing, the matching accuracy of pigeons in Group ITI-dark on 0- and 2-s sample trials remained consistently high with increasing retention interval, whereas matching accuracy on 10-s sample trials declined rapidly with increasing retention interval, to a level well below chance. The test-session data from Group ITI-dark are presented in the left panel of Figure 2.

We performed a two-way, repeated measures ANOVA on the testing data from Group ITI-dark, with sample type (0, 2, & 10 s) and retention interval (1, 2, 4, & 12 s) as factors. There was a significant effect of sample type, $F(2, 6) = 82.50, p < .01$, and a significant effect of retention interval, $F(3, 9) = 40.24, p < .01$. It is noteworthy that there was a significant Sample-Type × Retention-Interval interaction, $F(6, 18) = 57.35, p < .01$. As can be seen in the left panel of Figure 2, the interaction can be attributed to divergent retention functions for samples of 10 s versus those for 0 and 2 s. To further isolate the source of the interaction, we performed a planned comparison involving samples associated with each of the two comparisons and delay as factors. Again, the Sample-Type × Delay interaction was significant, $F(3, 18) = 69.37, p < .01$. As can be seen in the left panel of Figure 2, the slope of the retention function on 10-s sample delay trials was substantially steeper than on 0- and 2-s sample delay trials.

Retention Testing for Group ITI-Lit

During testing, matching accuracy of pigeons in Group ITI-lit appeared quite different from that of pigeons in Group ITI-dark. For pigeons in Group ITI-lit, divergent retention functions were not found (see right panel of Figure 2). We performed a two-way, repeated measures ANOVA on the testing data from Group ITI-lit, with sample type (0, 2, & 10 s) and retention interval (1, 2, 4, & 12 s) as factors. Although there was a significant effect of retention interval, $F(3, 9) = 24.22, p < .01$, there was not a significant effect of sample type ($F < 1$), nor, most important, was there a significant Sample-Type × Retention-Interval interaction ($F < 1$). As can be seen in the right panel of Figure 2, the slopes of the

Figure 2. Experiment 2: Retention functions (M with SEMs) following hybrid one-to-one/many-to-one matching with 10-s samples (triangles) associated with one comparison and 0- and 2-s samples (circles and squares, respectively) associated with the other comparison. Data from the group that had the dark intertrial intervals (Group ITI-dark) are presented on the left. Data from the group that had the lit intertrial intervals (Group ITI-lit) are presented on the right.
retention functions for the three sample types were virtually indistinguishable. For comparison purposes with the data from Group ITI-dark, a planned comparison involving trials with samples associated with the same comparison (0- and 2-s vs. 10-s) and delay indicated that the Sample-Type × Delay interaction was not statistically significant (F < 1).

Discussion

The results of Experiment 2 support the ITI/delay similarity hypothesis. Retention functions for Group ITI-dark mirrored those found by Gaitan and Wixted (2000, Experiment 2). Matching accuracy was consistently high with increasing retention interval on 0- and 2-s sample trials but steeply declined on 10-s sample trials, to a level below chance. Retention functions for Group ITI-lit were quite different, however. As predicted by the ITI/delay similarity hypothesis, for all three sample durations, retention functions declined gradually and at the same rate, to a level somewhat better than chance.

An alternative interpretation of the difference in results between Group ITI-dark and Group ITI-lit is that when the ITI was dark, the 0-s sample was actually a nonevent (i.e., in training, the ITI proceeded seamlessly into the 1-s delay), whereas lighting the ITI allowed the 0-s sample to become an event (i.e., the offset of the ITI followed by the 1-s delay). According to this hypothesis, one would predict that parallel retention functions should be found whenever all of the sample durations used in an experiment are greater than 0-s because they should all be detectable events. However, most differential-duration-sample matching experiments involve sample durations that are all greater than 0 s. Typically, the shortest sample that is used in differential-duration-sample matching experiments is 2-s long and, with the exception of the results of an experiment by Sherburne, Zentall, and Kaiser (1998) in which the ITI was differentiated from the delay interval, all of these studies have reported divergent retention functions.

General Discussion

The results of the present experiments have implications not only for duration-sample experiments involving the hybrid one-to-one/many-to-one design but for all experiments involving duration samples. Even when the shorter of two samples is greater than 0 s, the similarity between the ITI and the delay may produce a condition that is more similar to a 2-s sample in training than to a 10-s sample in training, and a choose-short bias may result. In support of this hypothesis, Sherburne et al. (1998) found that when the ITIs and delays were clearly distinguished, (by lighting the ITIs and keeping the delays dark), parallel retention functions were found.

Our results may also have implications for present/absent-sample matching in which differential durations are not involved (e.g., Sherburne & Zentall, 1993). For example, in the case of present/absent samples, the absent sample is typically more similar to the retention interval than the present sample. If this is the case, then on retention tests, the novel delay may be judged by the pigeon to be a variant of the absent sample, and an absent-sample bias may result.

Similarly, Zentall, Kaiser, Clement, Weaver, and Campbell (2000) have suggested that response similarity also may produce a bias. Specifically, pigeons tend to peck at the present sample and refrain from pecking at the absent sample, and they typically do not peck during the dark retention interval. Thus, if the pigeons use the presence versus the absence of sample pecking as a basis for choosing the comparison, then inserting a delay between the sample and the comparisons can result in an absent-sample bias.

With regard to studies that have examined the development of single-sample-detection/default strategies, the above concerns can be greatly reduced by avoiding the use of absent samples. Clement and Zentall (2000) trained pigeons on a hybrid one-to-one/many-to-one task in which one stimulus hue was associated with one comparison, and four different stimulus hues were associated with the other comparison. When retention tests were introduced, the pigeons’ matching accuracy remained high on the many-to-one trials, but accuracy declined rapidly with increasing delay on one-to-one trials. These results suggest that the pigeons were detecting the one sample uniquely associated one-to-one with a comparison stimulus, and they were choosing the alternative comparison by default.

The present research raises the larger issue of how to interpret results of experiments in which transfer of training is involved. Whenever an animal is exposed to novel test conditions, it is quite possible for the experimenter to overinterpret the animal’s response to test conditions (see Zentall, 1997). This is especially true when training involves matching to sample and when delays are inserted between the offset of the sample and the onset of the comparison stimuli. Not only is it possible that stimulus similarity (between the retention interval and the ITI or between the retention interval and one of the samples) or response similarity (between the behavior during the retention interval and in the presence of one of the samples) can bias comparison choice, but also that the novelty of the retention interval can result in a decrement in matching accuracy that may not be attributable to loss of memory even when no bias is present. Instead, the decrement in matching accuracy may merely reflect a response to novelty that becomes greater as the trial becomes more novel (i.e., as the delay increases).

Given the limits of retention testing, following training without a delay, researchers should develop procedures to deal with this problem. One approach has been suggested by Dorrance, Kaiser, and Zentall (2000), who trained pigeons on delayed matching with duration samples in which delays of various durations were experienced from the start. Under delay conditions that typically result in a strong choose-short effect, Dorrance et al. found parallel and relatively flat retention functions. Only when the pigeons were trained with relatively long retention intervals did the slopes of the retention functions begin to decline, and, even then, there was little evidence of a choose-short effect. Thus, the present results, as well as those of several other experiments (Dorrance et al., 2000; Sherburne et al., 1998; Zentall et al., 2000), suggest that in many cases, divergent retention functions may be produced by procedural artifacts rather than by detection strategies or by the compression of memory for time.
References


Received September 18, 2003
Revision received December 8, 2003
Accepted December 10, 2003