A Signal Detection Analysis of Memory for Nonoccurrence in Pigeons

John T. Wixted

In a relatively unexplored memory procedure, pigeons indicated whether or not an event recently occurred. On positive trials, a sample stimulus was presented, followed by a delay, followed by a choice between the *yes* and *no* alternatives (with *yes* being the correct choice). On negative trials, the yes and no choice alternatives were presented without a preceding sample, and on these trials *no* was the correct choice. Recent research on this topic seems to suggest that performance on negative trials is governed by a memory-free default response strategy, whereas performance on positive trials is governed by memory for the sample stimulus. However, a signal detection analysis leads to a different conclusion. In both cases, according to this account, performance was determined by the strength of a memory trace in relation to a decision criterion.

For various reasons, investigations of pigeon short-term memory have often involved sample stimuli consisting of presentations of food versus no food. Typically, the presentation of one of these samples is followed, after some delay, by a choice between two comparison stimuli (e.g., red and green). A response to one comparison is reinforced following samples of food, and a response to the other comparison is reinforced following samples of no food. The precise manner in which these sample stimuli are presented varies from study to study. In some cases, trials are initiated by a warning stimulus (e.g., a white light) followed either by the presentation of food or by nothing at all (i.e., no food). In other cases, no warning stimulus is used, and the samples consist of the brief presentation of food or a brief blackout (i.e., no food). In still other cases, no exteroceptive stimulus change is correlated with no-food samples. In spite of these procedural differences, the consistent finding is that performance following samples of food declines as the retention interval increases, whereas performance following samples of no food does not (Colwill, 1984; Colwill & Dickinson, 1980; Grant, 1991; Wilson & Boakes, 1985).

The obvious question raised by these results is why performance is unaffected by the size of the retention interval on no-food trials. Although food versus no-food samples clearly differ in motivational value, recent research suggests that a more important difference is that one involves the presence of an event (i.e., the occurrence of food), whereas the other involves the absence of an event (i.e., the nonoccurrence of food). Indeed, the same asymmetrical decay functions were observed by Grant (1991) when samples consisted of the presence versus absence of a variety of stimuli (including colors, shapes, and food). In each case, performance following the presence of an event declined as the retention interval increased, but performance following the absence of an event did not.

Colwill (1984), Wilson and Boakes (1985), and Grant (1991) all argued that the absence of a retention interval effect on no-sample trials suggests that memory plays no role on these trials. Instead, pigeons seem to adopt a default response strategy of choosing the comparison stimulus associated with the absence of a sample. The default strategy is overridden on trials involving a sample so long as the memory trace has not completely faded. This explanation accounts for the flat retention function on no-sample trials because, whether the retention interval is short or long, no memory trace is ever present to override the default response. The same account explains why performance on sample trials is often significantly below chance at longer retention intervals: When the memory trace fades completely, subjects revert to their default strategy and reliably choose the wrong comparison stimulus.

This kind of explanation, although intuitively appealing and consistent with the extant data, may not fully capture the most essential elements of the situation. The presence versus absence memory paradigm is the prototypical memory analog of an auditory signal detection task, and an analysis along these lines raises the possibility that performance on no-sample trials involves memory after all. The question of whether or not performance on no-sample trials reflects memory for nonoccurrence, as opposed to a memory-free default response strategy, exactly parallels a once-prominent theoretical question pertaining to auditory signal detection, namely, does performance on no-signal trials involve the perceptual system or not (Swets, Tanner, & Birdsall, 1961)? The strategies used to answer that question can also be applied to the presence versus absence memory paradigm.

In a typical signal detection procedure, human subjects are asked to decide (yes or no) whether or not a quiet tone was just presented. If a tone was actually presented, the subject's response would be classified either as a hit (yes) or as a miss (no). If no tone was actually presented, the response would be classified either as a false alarm (yes) or as a correct rejection (no). Figure 1 illustrates the standard theoretical analysis of performance on this task. The analysis assumes that the subjective intensity of background noise varies from trial to trial according to a Gaussian distribution. The subjective intensity of a tone superimposed on noise is also nor-
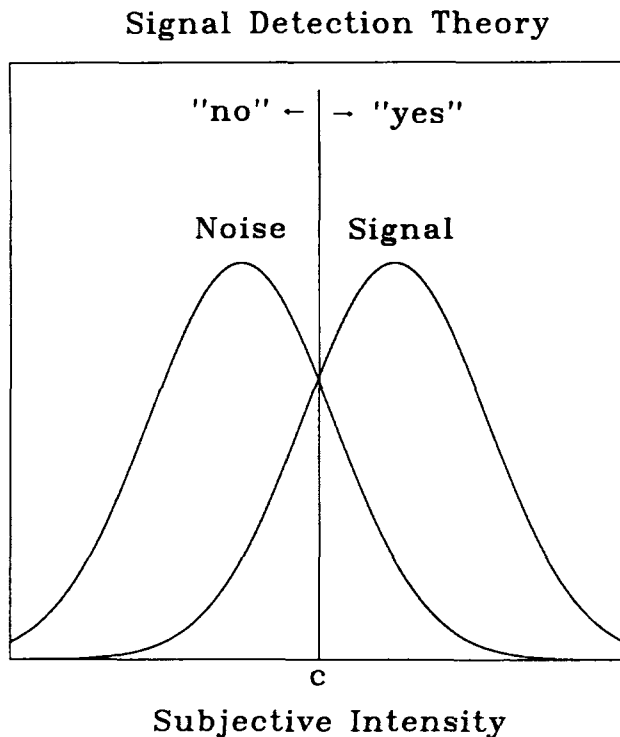
Signal Detection Theory

Figure 1. A graphical illustration of signal detection theory. (The signal and noise distributions correspond to sample and no-sample trials, respectively. c = the decision criterion.)

mally distributed and is, on average, stronger than background noise alone. The vertical line (c) represents the subject's decision criterion. On trials in which subjective intensity exceeds c, the subject responds positively; otherwise, the response is negative.

The placement of the criterion (c) and the distance between the two distributions (d') jointly determine the pattern of hits, misses, correct rejections, and false alarms. Although d' is determined by trial-specific factors (e.g., how loud the tone is), the position of c is determined by more extended experience with the relative payoff probabilities. If the reward for a hit is the same as that for a correct rejection, then the optimal placement of c lies directly between the means of the signal and noise distributions. If the reward for a hit exceeds that for a correct rejection, then c would be shifted to the left and responding would be biased toward the yes alternative. Note that according to this analysis, performance on no-signal trials is determined by the same variable that determines performance on signal trials, namely, subjective intensity.

The theoretical scheme outlined in Figure 1 is particularly relevant to the presence versus absence memory task discussed above. The x-axis in this case represents trace intensity (rather than auditory signal intensity), but the analysis is otherwise identical. On no-sample trials, variations in trace intensity result entirely from random activity (i.e., noise) in the memory system. That is, although no sample was actually presented, the act of retrospection nevertheless produces some sense of trace intensity associated with the sample. On

the majority of trials, the amount of trace intensity resulting from noise falls below the criterion, and the subject makes the correct response (no). Occasionally, however, noise in the system results in trace intensity high enough to produce a false memory (yes). Note that, in agreement with the data just reviewed, performance on no-sample trials should be unaffected by transient variations in the size of the retention interval because that manipulation affects neither the mean of the noise distribution nor the position of c.

Quite a different prediction applies on trials involving the presentation of a sample stimulus. Figure 2 presents hypothetical trace intensity distributions for three retention interval conditions (short, medium, and long). When the retention interval is short, the signal distribution falls far to the right because, in spite of some variability, trace intensity will generally be quite high (and will fall well above the decision criterion). Thus, the subject would respond correctly (yes) on the majority of sample trials. When the retention interval increases (and trace intensity weakens), the signal distribution gradually shifts toward the noise distribution. As it does, more and more of the distribution falls below c. Thus, unlike the no-sample trials, performance on sample trials should worsen as the retention interval increases because the subject will be increasingly likely to give the wrong answer (no). In the limiting case, the signal and noise distributions will overlap such that performance on sample trials will actually fall significantly below chance. As indicated earlier, below-chance performance has often been observed at long retention intervals on sample trials (e.g., Grant, 1991).

The question addressed by the present series of experiments is whether performance on no-sample trials is based on a default response strategy, as most have assumed, or on memory for nonoccurrence, as suggested by signal detection theory. The first experiment replicates the standard finding (i.e., no effect of retention interval on no-sample trials), and the next two experiments test differential predictions of the two theories. More specifically, Experiment 2 evaluates the effect of retention interval on no-sample trials when each retention interval is in effect for 15 sessions (rather than manipulated within session). Under those conditions, signal detection theory predicts that performance on both sample and no-sample trials should decay at the same rate. Experiment 3 uses the more traditional strategy of examining the form of the receiver operating characteristic (ROC) curve. An ROC curve, which depicts hit rate versus false alarm rate, is generated by manipulating the subject's inclination to respond yes or no without changing the strength of the memory trace. Default response theories typically predict a linear ROC curve, whereas signal detection theory predicts a curvilinear function (Murdock, 1965; Swets et al., 1961).

Experiment 1

The first experiment was designed to replicate earlier findings concerning the asymmetrical decay functions on sample versus no-sample trials. As indicated above, the differing decay functions have been taken as direct evidence for the idea that performance on sample and no-sample trials is gov-
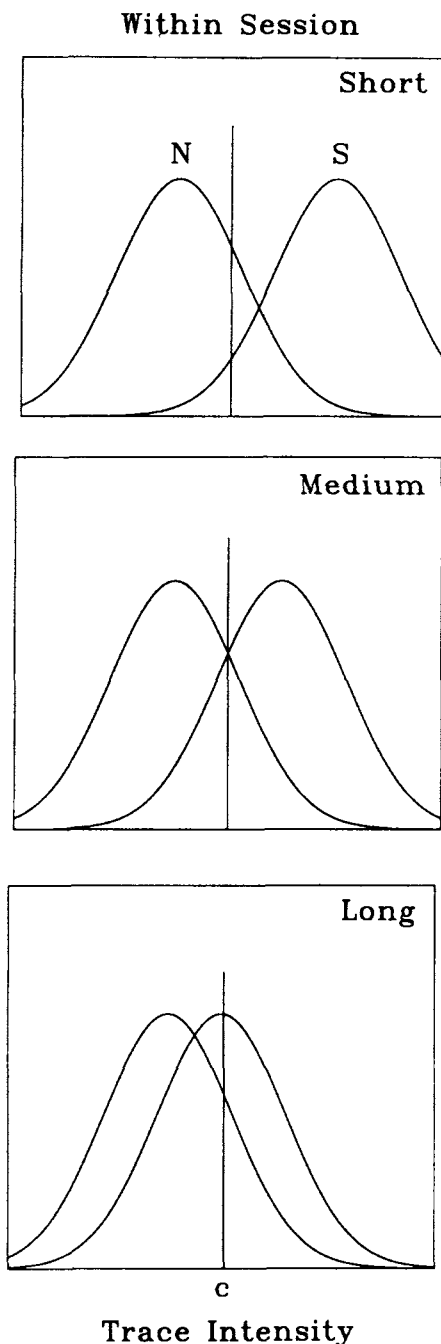
Within Session

Short

Medium

Long

c

Trace Intensity

Figure 2. Hypothetical signal (S) and noise (N) distributions (corresponding to sample and no-sample trials, respectively) for three retention intervals (short, medium, and long) manipulated within session. (The location of the decision criterion, *c*, is fixed.)

erned by different variables. Samples in the present experiment consisted of the presentation of a color stimulus on some trials or no stimulus at all on others (i.e., no exteroceptive stimulus change). Following a delay that varied within session, two comparison stimuli were presented (red vs. green). A response to one comparison was reinforced on trials initiated by a color sample, and a response to the other

comparison was reinforced on trials in which no sample appeared.

Method

Subjects. The subjects were 4 White Carneaux pigeons (*Columba livia*) with extensive experience on the delayed matching-to-sample (DMTS) task. They were each maintained at 80% of their free-feeding weight throughout the experiment.

Apparatus. Experimental chambers were constructed of Plexiglas (35 cm long × 32 cm wide × 34 cm high), except for one wall that was made of metal. Three Plexiglas response windows (6.5 cm high × 4.4 cm wide) were mounted side by side in the metallic wall of the chamber 20 cm above the chamber floor. Visual stimuli were presented on a VGA graphics monitor situated against the response windows on the opposite side. The stimuli used throughout these experiments consisted of a green square (approximately 3.5 cm × 3.5 cm) and an arrangement of four red squares (which together occupied an area of approximately 3.5 cm × 3.5 cm). Thus, the stimuli differed in both color and pattern. A microswitch located at the base of each response window recorded pecks. Food pellets were delivered by a 45-mg Gerbrands pellet dispenser through an opening into a pellet tray centered on the metal wall 4 cm above the chamber floor. The chamber was illuminated before and after sessions by a standard 28-V houselight, and extraneous noise was masked by an exhaust fan. The entire apparatus was enclosed in a sound-attenuating chamber, and the experiment was controlled (and data were recorded) by an IBM-compatible computer.

Procedure. All 4 pigeons had extensive experience on a standard DMTS procedure involving two sample stimuli. From the outset of this experiment, trials were arranged in the following way. Sample trials were initiated by the presentation of a color sample in the center window. The first response after 5 s terminated the sample and initiated the retention interval. Following the retention interval, the red and green comparison stimuli appeared in the side windows (with position randomly determined before each trial). A response to the stimulus that matched the sample (red for 2 birds and green for the other 2) was reinforced with two 45-mg food pellets. A response to the other stimulus terminated the trial. Following a 15-s intertrial interval (ITI), the next trial was presented.

No-sample trials were identical to the sample trials except that no sample stimulus was presented. More specifically, following the ITI, a 5-s timer counted down and was followed by the onset of the retention interval. On completion of the retention interval, the red and green choice stimuli were presented on the two side windows. For the 2 pigeons that received the red stimulus on sample trials, a response to green was reinforced with two 45-mg food pellets on no-sample trials. For the 2 pigeons that received the green stimulus on sample trials, a response to red was reinforced with two 45-mg food pellets on no-sample trials.

Half of the trials in every session were initiated by a sample stimulus and half by no sample stimulus. The trials were presented in random order with the restriction that no more than four of one type occur in a row. Initially, a retention interval of 0.5 s was used on every trial, and a correction procedure was in effect (i.e., incorrect responses resulted in a repetition of the previous trial). Once performance reached 85% correct, longer retention intervals were gradually introduced (up to 12 s). At this point the experiment proper began. A retention interval of 0.5, 2, 6, or 12 s was randomly selected prior to each trial (and during this phase the correction procedure was not in effect). The birds were run under these conditions until stable performance was achieved (25 sessions).

Results and Discussion

The data for each bird were averaged over the last five sessions. Figure 3 shows mean proportion correct for sample and no-sample trials as a function of retention interval duration (the error bars represent standard errors). The figure clearly suggests that, in agreement with previous studies, performance on no-sample trials remained flat, and performance on sample trials exhibited a steady decline as the retention interval increased. An analysis of variance performed on these data revealed a significant effect of retention interval, $F(3, 9) = 7.39$, as well as a significant interaction between sample condition and retention interval, $F(3, 9) = 9.62$. An alpha level of .05 was used for these and all subsequent analyses.

The results summarized in Figure 3 are representative of the forgetting functions exhibited by the individual subjects. Table 1 presents the slopes (and standard errors of the parameter estimates) of straight lines fitted to the individual sample and no-sample forgetting functions. In all cases, the sample functions have negative slopes (although the value for P29 does not differ significantly from zero). By contrast, the slopes of the no-sample functions are close to zero in all four cases (and in no case do the values differ significantly from zero).

These data suggest that, even with extended exposure to within-session variations in the size of the retention interval, the findings reported by Grant (1991) are obtained. Thus, the results are consistent with the idea that the pigeons in this experiment adopted a default response strategy of choosing the comparison stimulus associated with the absence of a sample unless memory for the sample was present (in which
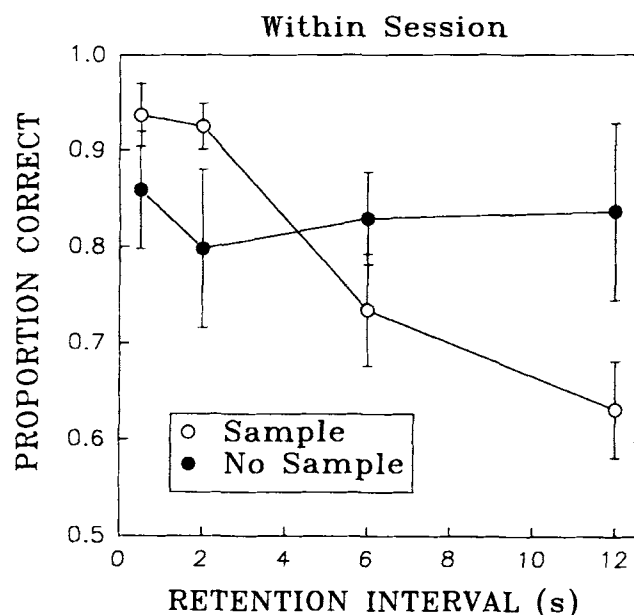
Table 1

Forgetting Function Slopes on Sample and No-Sample Trials for Individual Subjects (Experiment 1)

Subject	Sample		No sample	
	Slope	SE	Slope	SE
P29	−1.20	0.63	0.15	0.27
P47	−2.48	0.45	0.31	0.71
P8	−3.76	0.48	−0.49	0.89
P38	−3.80	0.67	0.11	0.04
M	−2.83	0.34	0.01	0.25

Note. SEs are standard errors associated with the parameter estimates.

case the other comparison stimulus was chosen). An important aspect of this theoretical interpretation is that it explicitly assumes that memory plays no role on no-sample trials.

As indicated earlier, these data are also consistent with a signal detection analysis (Figure 2), which assumes that a retrospective memory decision occurs on every trial. According to this model, the location of the noise distribution is always fixed, whereas the location of the signal distribution varies with the size of the retention interval on a given sample trial. The longer the retention interval, the smaller the distance between the signal and noise distributions. The position of c is also malleable, but it does not vary on a trial-by-trial basis. Instead, its position is determined by the subject's experience with the payoff probabilities averaged over trials.

Because performance on no-sample trials is determined by the position of the noise distribution and the placement of c, transient variations in the retention interval should have no effect. By contrast, performance on sample trials is determined by the position of the signal distribution relative to c and should therefore be dramatically affected by any change in the size of the retention interval. On both sample and no-sample trials, however, the task faced by the pigeon is the same and it involves a query of memory: Did the sample appear or not? On some occasions, according to this account, the pigeon falsely remembers the presentation of a sample.

Experiment 2

All of the experiments using the presence versus absence memory paradigm have varied the retention interval within a session. Under those conditions, the default response hypothesis and signal detection theory make the same (correct) predictions. When the retention interval is varied across conditions, however, theoretical predictions diverge. If pigeons simply rely on a default response strategy of choosing the comparison associated with the absence of a sample on no-sample trials, then the pattern of results should not differ whether the retention interval is varied within or between sessions. A signal detection analysis, by contrast, suggests that the between-session manipulation of retention interval will affect performance on both sample and no-sample trials.

Figure 4 again presents hypothetical trace intensity distributions for three retention interval conditions under the
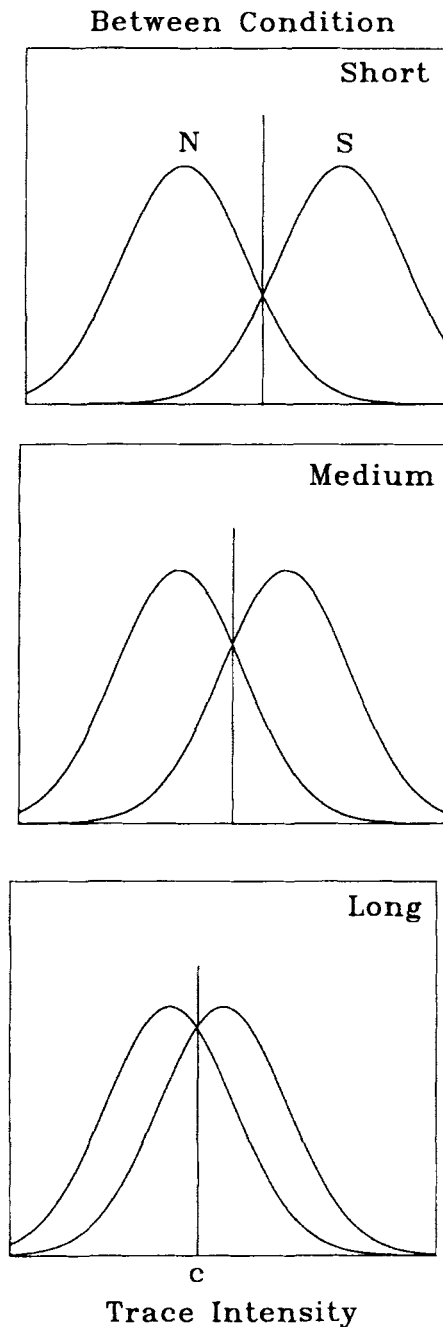

Figure 3. Proportion of correct responses as a function of retention interval on the sample and no-sample trials of Experiment 1 averaged over subjects. (The error bars represent the standard errors associated with each mean value.)

Between Condition

Short

N   S

Medium

Long

c

Trace Intensity

Figure 4. Hypothetical signal (S) and noise (N) distributions (corresponding to sample and no-sample trials, respectively) for three retention intervals (short, medium, and long) manipulated between conditions. (The location of the decision criterion, c, changes with each condition.)

assumption that the retention interval is varied between conditions. The only difference between Figure 4 and Figure 2 is the placement of the decision criterion. Neglecting the effect of inherent response biases, one might assume that the criterion will generally be placed at the point that maximizes the payoff probabilities averaged over trials (i.e., directly between the two distributions). Thus, unlike the within-

session case, each retention interval is associated with a different value of c.

Note that the anticipated change in c now suggests that performance on no-sample trials will not remain constant as the retention interval increases. As c moves leftward, more of the noise distribution falls to the right of c such that more false alarms (and correspondingly fewer correct rejections) should occur. Thus, performance on both sample and no-sample trials should decline as the retention interval increases. The existence of a retention interval effect on no-sample trials would at least raise the possibility that the memory system is involved on these trials.

Method

Subjects and apparatus. The subjects and apparatus were the same as those used in the previous experiment.

Procedure. The procedure was identical to that used in Experiment 1 except that each retention interval was in effect for an extended period of time to allow performance to reach stability. The pigeons were exposed to an ascending series of retention intervals (0.5, 2, 6, 12, and 24 s), with each condition in effect for 15 sessions. An ascending sequence was used because the random selection of retention intervals tended to produce disruptions in performance when a long retention interval condition immediately followed a short one (the most notable problem being the complete cessation of responding to the sample stimulus). Also, the use of an ascending series seemed unlikely to disrupt any default response strategy that may have developed at the shorter retention intervals. At the end of the sequence, all birds were returned to the .50-s condition for 15 sessions.

Results and Discussion

The data were again averaged over the last five sessions of each condition. Figure 5 shows the mean proportion correct as a function of retention interval for both the sample and no-sample trials. In contrast to the results from Experiment 1, performance declined on both kinds of trial. An analysis of variance performed on these data revealed a significant effect for retention interval, $F(4, 12) = 9.32$, but the effect of trial type and the interaction between trial type and retention interval did not approach significance. Although performance at the 24-s retention interval declined to 65.0% and 72.6% correct on sample and no-sample trials, respectively, accuracy increased to 92.4% and 85.6% correct over the last five sessions of the final condition (with the retention interval again set to 0.5 s).

The results shown in Figure 5 are, for the most part, representative of the performance of individual subjects. As shown in Table 2, the forgetting functions generated by these subjects exhibited negative slopes in every case. All of these slopes differed significantly from zero except for Subject 38 in the no-sample condition. The no-sample retention function for this bird was nearly flat. Thus, 1 bird exhibited the pattern predicted by the default response hypothesis (i.e., the absence of a retention interval effect on no-sample trials), and 3 exhibited the pattern predicted by signal detection theory.

The results of Experiment 2 demonstrate that performance on no-sample trials can be affected by the size of the retention
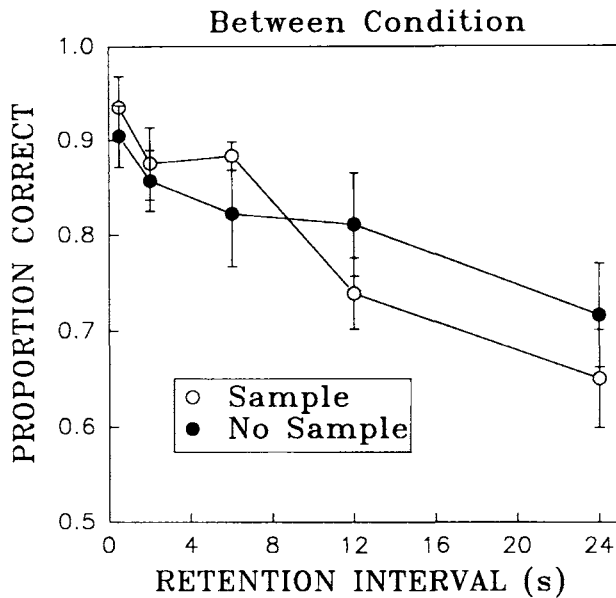
Between Condition

Figure 5. Proportion of correct responses as a function of retention interval on the sample and no-sample trials of Experiment 2 averaged over subjects. (The error bars represent the standard errors associated with each mean value.)

interval under the right conditions. Indeed, had an extremely long retention interval been used, it seems almost certain that performance would have fallen to chance levels. That is, if trace intensity on signal trials is never stronger than that on noise trials, then the task essentially reduces to a discrete-trials choice procedure with equal probabilities of reinforcement on the two alternatives. One might expect that, under those conditions, responding would be more or less equally divided between the yes and no alternatives (i.e., chance performance would be observed).

These results are in accordance with the predictions of the signal detection analysis illustrated in Figure 4. Because each retention interval was in effect for an extended period of time, the decision threshold (c) was, according to this account, adjusted in such a way as to maintain more favorable payoff probabilities (averaged across trials). As a result, performance on both signal and noise trials should diminish as the retention interval increases.

Table 2
Forgetting Function Slopes on Sample and No-Sample Trials for Individual Subjects (Experiment 2)

Subject	Sample		No sample	
	Slope	SE	Slope	SE
P29	−0.66	0.16	−0.57	0.14
P47	−1.73	0.25	−0.60	0.09
P8	−1.61	0.36	−1.49	0.28
P38	−0.77	0.11	−0.01	0.22
M	−1.23	0.12	−0.76	0.08

Note. SEs = standard errors associated with the parameter estimates.

The default response hypothesis, which readily explained the results from Experiment 1, cannot explain the data from this experiment unless it is assumed that pigeons altered their default response strategy when the retention interval was increased across conditions. That is, when a short retention interval was in effect, the pigeons may have adopted a default strategy of choosing, almost exclusively, the alternative associated with the absence of a sample. As the retention interval increased across conditions, they may have chosen to adopt a less exclusive default strategy (for whatever reason) and divided their responses more evenly between the two alternatives. If that were true, then perhaps one could still argue that performance on no-sample trials did not involve memory in spite of the fact that the forgetting functions on sample and no-sample trials declined at essentially the same rate. The viability of this account obviously depends on an adequate explanation for why a new default strategy was implemented with each retention interval.

At least one fairly straightforward modification of the default response hypothesis can accommodate the results of Experiment 2. Swets et al. (1961) evaluated the plausibility of a "high threshold" theory of auditory detection that is, in many respects, similar to the default response hypothesis under consideration here. According to high threshold theory, a yes response occurs on signal trials if the strength of the signal exceeds a fixed sensory threshold. On no-signal trials, however, yes responses (i.e., false alarms) are merely guesses that do not arise because of any sensory process. Instead, the subject is assumed to adopt a default response strategy of saying yes or no a certain proportion of the time in the absence of an above-threshold sensation. As the payoff for a correct yes response increases relative to that for a correct no response, the subject will increasingly respond yes by default on both no-signal trials and below-threshold signal trials. These new yes responses are not the result of a sensory process exceeding a decision criterion on a greater number of trials (as in signal detection theory), but result from what amounts to a default response strategy of increasingly choosing the yes alternative to improve the payoff.

The high-threshold model as it pertains to the presence versus absence memory task is depicted in Figure 6. The distribution on the right represents trace intensity on sample trials and the distribution on the left represents trace intensity on no-sample trials. The threshold, the location of which cannot change, is represented by T. From the subject's point of view, a subthreshold memory trace is undetectable (i.e., the memory trace is effectively absent). The probability that trace intensity exceeds the fixed threshold on sample trials is represented by p, and the probability that it falls below that threshold is therefore equal to $1-p$. If trace intensity exceeds threshold, the pigeon is assumed to respond yes; otherwise, the default response strategy is implemented. Note that the default strategy is always implemented on no-sample trials because trace intensity on these trials never exceeds threshold. Thus, the high threshold model shown in Figure 6 is formally equivalent to the default response hypothesis advanced to explain performance on no-sample trials (Colwill, 1984; Grant, 1991; Wilson & Boakes, 1985).
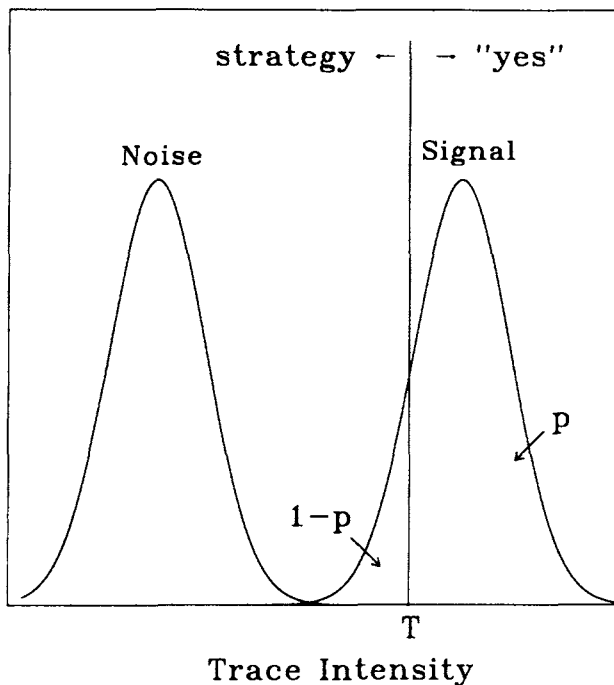
High Threshold Theory

Figure 6. A graphical illustration of high-threshold theory. (The signal and noise distributions correspond to sample and no-sample trials, respectively. T = the memory threshold; p = the probability that trace intensity will exceed the threshold on signal trials.)

The default response strategy that is implemented on subthreshold trials can be symbolically represented as some probability, δ, of choosing the yes alternative in the absence of a memory trace (with $1-\delta$ representing the probability of choosing the no alternative under those conditions). To maintain a favorable payoff, that strategy might be adjusted according to the average probability of reinforcement for choosing the yes and no alternatives in the absence of a memory trace. That value depends on the location of the signal distribution in Figure 6, a point that can be most easily seen by considering the two extreme cases. When the entire signal distribution falls above threshold (e.g., when the retention interval is always very short) then the probability of reinforcement for choosing the yes alternative in the absence of a memory trace is zero (because the memory trace is absent on no-sample trials only). Under those conditions, the optimal default response strategy is to choose the no alternative exclusively. When the entire signal distribution falls below threshold (e.g., when the retention interval is always very long), then the probability of reinforcement for choosing the yes alternative in the absence of a memory trace is .50 (because the memory trace is absent on both sample and no-sample trials). In that case, the default response strategy should be adjusted to include more yes responses. Indeed, one reasonable assumption is that the default response strategy tracks the relative probability of reinforcement for a yes response in the absence of a memory trace, a value that is

given by

$$p(\text{rft} \mid \text{yes}) = \frac{(1-p)N_s}{(1-p)N_s + N_n}, \quad (1)$$

where N_s and N_n represent the number of sample and no-sample trials, respectively, and p represents the probability that trace intensity will exceed threshold (cf. Figure 6). The numerator represents the number of sample trials in which the memory trace falls below threshold. For all of these trials, a default yes response will produce a reinforcer. The denominator represents the total number of trials in which the memory trace is absent (i.e., the subset of sample trials with a faded trace plus all of the no-sample trials). Note that if $N_s = N_n$, the right side of Equation 1 reduces to $(1-p)/(2-p)$, which is the probability of reinforcement for a yes response on subthreshold trials. If the default probability of choosing the yes alternative, δ, is determined by the probability of reinforcement for choosing the yes alternative in the absence of a memory trace, then

$$\delta = (1-p)/(2-p). \quad (2)$$

This equation immediately implies that as p decreases (e.g., with the longer retention intervals of Experiment 2), the probability of choosing the yes alternative by default increases.

What does Equation 2 predict about the probability of a correct response on sample and no-sample trials (hits and correct rejections, respectively)? The probability of a hit (H) is given by

$$H = p + \delta(1-p), \quad (3)$$

which is the probability of a correct yes response due to memory trace exceeding threshold (p) plus the probability of a correct default yes response when the strength of the trace falls below threshold $[\delta(1-p)]$. The probability of a correct rejection (CR) is equal to $1-\delta$, which is simply the probability of a correct no response occurring by default on no-sample trials. Substituting the right side of Equation 2 for δ in Equation 3 and simplifying yields

$$H = 1/(2-p) \quad (4)$$

and

$$CR = 1/(2-p). \quad (5)$$

Thus, hits and correct rejections should decline at the same rate as p decreases owing to a lengthening retention interval, a prediction that corresponds to the principal finding of Experiment 2. Note that this prediction arises even though the model assumes that memory is not involved on no-sample trials.

The revised default response analysis is also consistent with the results of Experiment 1 in which a flat CR function was observed. The reason is that the probability of reinforcement for a yes response in the absence of a memory trace does not change on a trial-by-trial basis. Instead, the default response strategy (like the placement of the signal detection

criterion) is determined by experience with reinforcement probabilities averaged across trials. Thus, the default strategy will not change from trial to trial, and a flat no-sample retention function should be (and is) observed.

Thus far, two very different accounts can explain the pattern of results obtained in the first two experiments. One assumes that performance on no-sample trials occasionally involves false memories (signal detection), whereas the other assumes that performance on these trials is independent of memory processes (the default response hypothesis). One way to distinguish between these two theories is to examine the form of an empirical ROC curve.

As detailed by Swets et al. (1961), the high threshold model predicts a linear function when the probability of a hit is plotted as a function of the probability of a false alarm (i.e., a linear ROC curve). That this is true of the model just outlined is easily proved. According to Equation 3, H is equal to $p + \delta(1 - p)$. The probability of a false alarm, FA, is simply equal to δ (the probability of a default yes response on no-sample trials). Thus, substituting FA for δ in Equation 3 yields $H = p + (1 - p)FA$, a simple linear function that intersects the H axis at p and extends to the upper right corner of the plot (note that this analysis assumes that p is constant). Signal detection theory, by contrast, predicts a characteristically bowed ROC curve that extends upward from the origin to the upper right corner. Although an algebraic expression for this function does not exist, a close approximation is $H = 1/\{e^{-\alpha d'}[(1/FA) - 1] + 1\}$, where α is equal to 1.7 (Ogilvie & Creelman, 1968).

Swets et al. (1961) used the form of the ROC curve to rule out a high threshold theory of auditory signal detection, and Murdock (1965) used a similar approach to reject a high threshold account of human short-term memory. An ROC curve has never been examined using a presence versus absence memory procedure with animals, although an apparently curved ROC function was obtained by Harnett, McCarthy, and Davison (1984), using a standard DMTS procedure. In Experiment 3, ROC curves were obtained for four pigeons by manipulating the probability of reinforcement for hits and correct rejections across conditions.

Experiment 3

Several methods are available to generate empirical ROC curves. One method is to vary the ratio of signal to noise trials across conditions (e.g., 90/10, 50/50, or 10/90), and another is to differentially manipulate the probability of reward for hits and correct rejections across conditions (e.g., 0.8/0.2, 0.6/0.6, or 0.2/0.8). Both strategies generate a range of hit and false alarm probabilities, presumably without affecting memory characteristics. From a signal detection point of view, for example, these strategies are assumed to induce changes in the placement of c without affecting the distance between the signal and noise distributions (d'). According to the high threshold model, these strategies alter the default probability of a yes response (δ) without affecting the proportion of trials in which trace strength exceeds the threshold (p). In the present experiment, ROC curves were generated

for each of four retention intervals by varying the probability of reinforcement for hits and correct rejections.

Method

Subjects and apparatus. The subjects and apparatus were the same as those used in the previous experiment.

Procedure. The procedure was identical to that used in Experiment 1 (i.e., the retention interval was randomly varied within session) except that the reinforcement probabilities for hits and correct rejections were varied across conditions, and the retention intervals within each condition were 0.5, 2, 4, and 12 s. Each reinforcement condition was in effect for 15 sessions. Following baseline training during which the probability of reinforcement was 0.6 for both hits and correct rejections (0.6/0.6), the values were changed to 0.20/1.0 for two birds and 1.0/0.2 for the other 2 birds. After 15 sessions under these conditions, these values were reversed for 15 sessions. In the final phase of the experiment, the reinforcement probabilities were reset to their baseline values (0.6/0.6).

Results and Discussion

ROC function analysis. Figure 7 presents the ROC functions (i.e., hit vs. false alarm probabilities) for each retention interval averaged across subjects. Each point represents the average performance for the 4 birds during the last five sessions of each condition (the values for the pre- and post-0.6/0.6 conditions were averaged together). For each plot, the middle of the three points reflects the hit and false alarm probabilities obtained when the probability of reinforcement for a correct yes response equaled that for a correct no response (0.6/0.6). The point higher and to the right in each graph reflects the hit and false alarm probabilities obtained when the probability of reinforcement for a correct yes response exceeded that for a correct no response (1.0/0.2). The point lower and to the left in each graph reflects the hit and false alarm probabilities obtained when responding was biased in the opposite direction (0.2/1.0).

As indicated earlier, one reasonable version of the default response theory requires a linear ROC function of the form $H = p + (1 - p)FA$, where p represents the proportion of sample trials in which trace intensity falls above a fixed threshold (Figure 6). This one-parameter line extends from the upper right corner to a point, p, on the left vertical axis. Signal detection theory by contrast, requires a curved ROC function of the approximate form $H = 1/\{e^{-\alpha d'}[(1/FA) - 1] + 1\}$. This one-parameter function follows a curvilinear trajectory from the lower left corner to the upper right corner. The larger the value of d', the more bowed the function becomes.

The solid curves shown in Figure 7 represent the best-fitting ROC functions based on signal detection theory, whereas the dashed lines represent the best-fitting linear functions based on the default response account. Table 3, which shows the percentage of variance accounted for by these two functions for each ROC plot, corroborates the visual impression: In all four cases, the performance of the curvilinear function far exceeds that of the linear function. Indeed, the linear fits are so inaccurate that the version of high
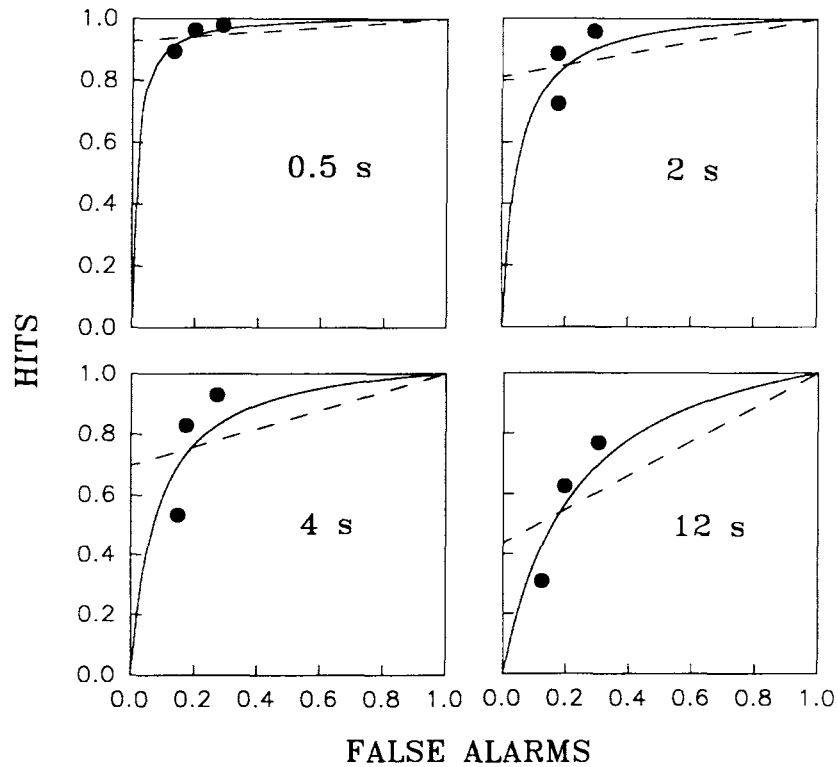
Figure 7. Empirical receiver operating characteristic (ROC) curves for the four different retention intervals used in Experiment 3. (Each graph depicts the hit rate vs. the false alarm rate for three reinforcement outcome conditions. The solid curves represent the best-fitting ROC functions based on signal detection theory, whereas the dashed lines represent the best-fitting linear functions based on high threshold theory.)

threshold theory under consideration here can be safely rejected. A similar ROC analysis performed on the data from individual subjects produced essentially the same results. Although the data were more variable, of the 16 individual fits (4 per subject), the curvilinear function accounted for a higher percentage of the variance than the linear function 12 times, and the linear function slightly outperformed the curvilinear function only once (variable data prevented convergence in the remaining three cases).

Although the fits based on signal detection theory are far better than those based on the default response account, some systematic deviation is apparent. In particular, the range of variation in false alarm probabilities was generally less than would be expected given the fairly wide range of hits induced

Table 3
Percentage of Receiver Operating Characteristic Data Variance Accounted for by Default Response (Linear) and Signal Detection (Curvilinear) Models

Retention interval	Linear	Curvilinear
0.5	20.8	79.4
2	13.8	38.1
4	14.6	46.5
12	35.9	80.0

by manipulating the reinforcement outcomes (this is especially true for the 2- and 4-s retention intervals). Similar deviations are often apparent in the data from auditory signal detection studies involving human subjects. As noted by Swets et al. (1961), such deviations might arise because the assumption that the signal and noise distribution are of exactly equal variance may be too strong.

Maximum likelihood analysis. To this point, the predictions of signal detection theory have been evaluated only in relative terms (i.e., relative to high threshold theory). An absolute test of signal detection theory asks whether or not the observed deviations from its predictions are greater than would be expected on the basis of chance. The least squares procedure used above is not the appropriate way to address this question. As described by Ogilvie and Creelman (1968), maximum likelihood estimation provides a much better alternative. In this approach, the parameters of the signal detection model are adjusted to maximize the theoretical probability of obtaining the observed distribution of responses, and the model's goodness of fit is evaluated by the chi-square statistic. A significant chi-square suggests that the deviations are greater than would be expected by chance (in which case the model can be rejected).

The data from each retention interval were subjected to a maximum likelihood analysis using the full signal detection
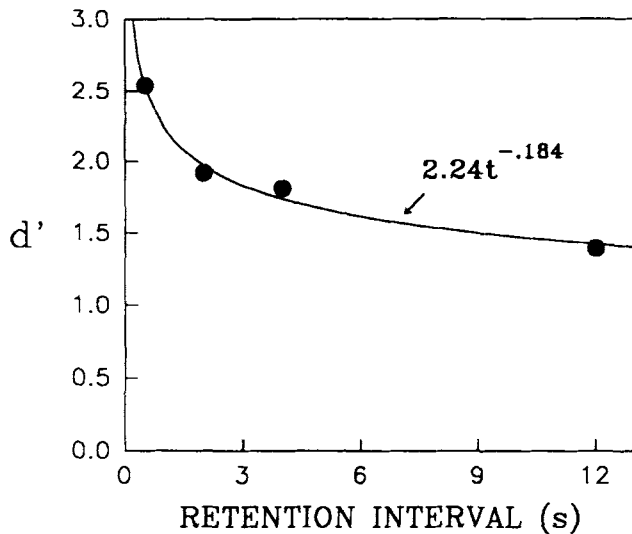
Figure 8. Forgetting function (*d'* vs. retention interval) from Experiment 3. (The *d'* estimates were obtained from a maximum likelihood analysis of the receiver operating characteristic curves shown in Figure 7.)

model (i.e., allowing for the possibility of unequal signal and noise distribution variances). The full model was used because inspection of the ROC curves shown in Figure 7 suggests that the deviations from the simpler model, though not extreme, are more systematic than would be expected on the basis of chance. Each fit of the full model involved estimating five parameters: *r* (the standard deviation of the noise distribution relative to the signal distribution), *d'* (the distance between the means of the signal and noise distributions), and three decision criteria, c_1, c_2, and c_3 (one for each condition). Because each fit involved five parameters and six data points (three hit rates and three false alarm rates), the chi-square goodness-of-fit statistic for a given retention interval was associated with one degree of freedom. The resulting chi-square values for retention intervals one through four were 0.02, 3.99, 0.93, and 2.21, respectively such that the overall test, $\chi^2(4, N = 3460) = 5.72$, was not significant.[1]

The maximum likelihood estimates of *r*, c_1, c_2, and c_3 did not vary systematically as a function of retention interval. The mean value of *r* across the four retention intervals was 2.99, which implies that the variance of the signal distribution was approximately one third that of the noise distribution. This would account for the fact that hit rates changed more rapidly than false alarm rates across conditions (Figure 7). Also, as would be expected, $c_1 < c_2 < c_3$ for conditions 1.0/0.2, 0.6/0.6, and 0.2/1.0, respectively. In other words, the decision criterion moved up as the payoff for a correct no response increased relative to that for a correct yes response. The mean maximum likelihood estimates of c_1, c_2, and c_3 (averaged across retention intervals) were 0.59, 0.86, and 1.08, respectively. The only parameter that did change systematically as a function of retention interval was *d'* (which is to be expected). The maximum likelihood values for this parameter were 2.54, 1.92, 1.81, and 1.39 for Retention Intervals 1–4, respectively.

Dependent measure implications. The preceding analysis suggests that the signal detection model offers a viable account of the present results and highlights an important issue regarding one's choice of a dependent measure. For example, as argued previously on many occasions (e.g., Davison & Tustin, 1978), the percentage of correct responses (which is the preferred dependent measure in the animal memory literature) can be misleading because it fails to take into account the effect of response bias. That is, an experimental manipulation that changes the placement of the decision criterion will affect this measure even though the strength of the memory trace is unaltered. For this reason, the use of *d'* (or a similar measure) is advisable because it is theoretically immune to changes in response bias.

On the other hand, even *d'* is potentially misleading because it is usually calculated using tables that assume equal variances for the signal and noise distributions. An experimental manipulation that affects the variance of the signal distribution without affecting the mean strength of the memory trace will produce different table *d'* values (falsely implying a change in the strength of memory). From a signal detection point of view, an ROC curve analysis is the safest way to determine which property of memory performance (bias, signal variance, or signal strength) is affected.

Although Experiment 3 was performed to test the predictions of signal detection and high threshold theory, the considerations presented above suggest that the data may be useful for another purpose as well. Specifically, Wixted and Ebbesen (1991) investigated the mathematical form of forgetting in human and animal subjects and found that, in both cases, the power function outperformed a number of reasonable alternatives (including the exponential and the hyperbola). However, one reasonable objection to the fits of the pigeon data reported in that experiment concerns the choice of the dependent measure (viz., percentage correct). What form would the forgetting function assume if the dependent measure took into account the possibility of response bias and unequal signal and noise distribution variances?

The results of Experiment 3 provide an answer to that question. Figure 8 shows the forgetting function obtained in this experiment using the *d'* values from the maximum likelihood analysis described above. The curved line represents the least squares fit of the power function (of the form at^{-b}), which accounted for 98.7% of the variance. More important, the deviations from the best-fitting power function are nonsystematic. By contrast, the hyperbola and exponential accounted for 86.5% and 81.2% of the variance, respectively, and the deviations were highly systematic.

Although the possibility is remote, these results could be an artifact of combining the data from the 4 pigeons prior to fitting the three mathematical functions. Thus, for example, the aggregate curve shown in Figure 8 might best be described by the power function even though the individual curves might best be described by the hyperbola or exponential. To evaluate this possibility, a separate maximum

[1] The details of the maximum likelihood estimation procedure for ROC curves can be found in Ogilvie and Creelman (1968).

likelihood analysis was performed for each bird (allowing a d' forgetting function to be plotted for each one). In most cases, a chi-square goodness-of-fit test for the full signal detection model was not possible because cells for the individual birds contained expected values of less than 5. Nevertheless, using the maximum likelihood estimates of d', the power function clearly provided the best fit for 2 of the 4 birds, the hyperbola slightly outperformed the power function in one case, and the exponential provided the best fit in the remaining case. The mean percentage of individual data variance accounted for by the power function, hyperbola, and exponential was 92.6, 88.7, and 85.7, respectively. Thus, although some variability is evident, the power function fit to the aggregate curve shown in Figure 8 reflects its relatively strong performance on the level of individual pigeons. Moreover, and not necessarily, these results corroborate previously reported findings based on an atheoretical dependent measure (percentage correct).

Differential outcomes. An issue that could complicate the interpretation of these results concerns the generally positive effect of differential outcomes (e.g., Santi & Roberts, 1985; Urcuioli, 1991). That is, previous research has demonstrated that the use of differing reinforcement probabilities associated with the two choice alternatives leads to enhanced performance relative to the nondifferential case. In Experiment 3, two conditions involved differential reinforcement outcomes (1.0 vs. 0.2 and 0.2 vs. 1.0) and one involved nondifferential outcomes (0.6 vs. 0.6). The ROC curve analysis assumed a single p in the linear case (which represents the proportion of sample trials in which trace intensity exceeds threshold) and a single d' in the curvilinear case (which represents the separation of the signal and noise distributions) for all three reinforcement conditions. If either d' or p were higher in the differential conditions (as previous research would seem to suggest) then such an analysis would be invalid.

The fact that the signal detection model produced an adequate fit to the data already suggests that this was not a problem. Two additional analyses confirm that a differential outcome effect was not observed in this experiment. First, an analysis of variance comparing the percentage of correct responses in the nondifferential condition (0.6/0.6) versus the two differential conditions averaged together yielded a significant effect of retention interval, $F(3, 9) = 30.28$, as expected, but neither the effect of outcome condition nor the interaction between outcome condition and retention interval approached significance. Indeed, at all four retention intervals, performance in the nondifferential condition equaled or slightly exceeded performance in the differential outcome condition. Second, d' values for the two differential conditions averaged together closely matched those for the nondifferential condition. The values from the differential conditions were 2.73, 2.03, 1.90, and 0.98 for Retention Intervals 1–4, respectively, whereas the corresponding values from the nondifferential condition were 2.73, 2.20, 1.91, and 1.21. Note that these d' values were necessarily determined in the usual way (i.e., using a d' table that assumes equal variances).

Given the robust nature of the differential outcome effect elsewhere, its absence here may seem somewhat surprising.

However, the apparent discrepancy may have a simple explanation. The sample–no-sample procedure used throughout these experiments seems likely to encourage a retrospective strategy. Indeed, on no-sample trials, a retrospective strategy is essentially forced because no exteroceptive event signals an impending memory test. The standard explanation for the differential outcome effect assumes a prospective strategy. Specifically, outcome expectancies generated by the differential outcomes are assumed to guide responding to the correct alternative (e.g., Peterson, Wheeler, & Armstrong, 1978; Santi & Roberts, 1985). If pigeons in this experiment relied on a retrospective strategy, then any outcome expectancies that might have occurred on sample trials would not be expected to facilitate performance.

General Discussion

The question addressed by these experiments is whether or not pigeons can be said to exhibit memory for the prior non-occurrence of an event. That is, on trials without a sample, is responding during the choice phase governed by a memory-based retrospective strategy or by a memory-free default response strategy? Previous research using the presence versus absence memory procedure has been interpreted as supporting the latter idea (e.g., Colwill, 1984; Colwill & Dickinson, 1980; Grant, 1991; Wilson & Boakes, 1985). The typical finding in this literature is that sample trials produce a negatively sloped retention function (as expected), but no-sample trials produce a flat retention function. The absence of a retention interval effect on no-sample trials has been taken as direct evidence that memory is not involved on these trials. Instead, responses to the no key are assumed to occur by default on trials without a sample.

The present results suggest that, in spite of appearances to the contrary, performance on no-sample trials may actually involve memory for nonoccurrence. Interpreted in terms of signal detection theory, the flat retention function observed on no-sample trials is to be expected when the retention interval is manipulated within session. Under those conditions, the noise distribution and decision criterion remain fixed despite variations in the size of the retention interval (Figure 2). When the retention interval is manipulated across conditions, by contrast, the criterion should shift to the left as the retention interval increases (Figure 4). Only in the latter case would no-sample performance be expected to decline as a function of the retention interval. In both cases, however, performance on no-sample trials is assumed to be governed by the same variable that governs performance on sample trials, namely, the strength of a memory signal relative to a decision criterion.

Although the results of Experiments 1 and 2 were consistent with the predictions of signal detection theory, it is possible to revise the default response model to make the same predictions if one assumes that pigeons alter their default strategy when the retention interval is manipulated across conditions. Thus, for example, if that strategy (which involves a probabilistic choice between the yes and no alternatives in the absence of a memory trace) is adjusted to maintain a favorable outcome, then performance would be

expected to decline at the same rate on sample and no-sample trials under the conditions used in Experiment 2. Models based on this idea generally predict a linear relationship between the probability of a hit and the probability of a false alarm (i.e., a linear ROC curve), whereas signal detection theory always predicts a curvilinear relationship between these two measures (Murdock, 1965; Swets et al., 1961). The results of Experiment 3 suggested that this seemingly reasonable version of the default response model is also untenable. Although the fits were by no means perfect, the ROC curves were far more accurately described by the one-parameter curvilinear function based on signal detection theory. Using the same strategy, Murdock (1965) reached the same conclusion with regard to human short-term memory.

The present results do not rule out all possible default response models. A variety of more complicated models can be proposed that will fit an apparently curvilinear ROC function even though responding is assumed to occur by default on no-sample (or no-signal) trials (Lockhart & Murdock, 1970). For example, Luce's (1960) low-threshold model holds that default yes responses can occur when trace intensity falls below threshold (as with the high-threshold model) and that default no responses can occur when trace intensity falls above threshold. This model predicts a bitonic function with two linear segments that can often fit an ROC curve as well as a curvilinear function derived from signal detection theory. Nevertheless, at the very least, the present results suggest that the absence of a negatively sloped retention function on no-sample trials does not necessarily indicate that memory is not involved. Indeed, on balance, the evidence is more easily reconciled with the idea that performance on these trials reflects a retrospective decision about whether or not an event recently occurred.

References

Colwill, R. M. (1984). Disruption of short-term memory for reinforcement by ambient illumination. Quarterly Journal of Experimental Psychology, 36B, 235–258.

Colwill, R. M., & Dickinson, A. (1980). Short-term retention of "surprising" events by pigeons. Quarterly Journal of Experimental Psychology, 32, 539–556.

Davison, M. C., & Tustin, R. D. (1978). The relation between the generalized matching law and signal-detection theory. Journal of the Experimental Analysis of Behavior, 29, 331–336.

Grant, D. S. (1991). Symmetrical and asymmetrical coding of food and no-food samples in delayed matching in pigeons. Journal of Experimental Psychology: Animal Behavior Processes, 17, 186–193.

Harnett, P., McCarthy, D., & Davison, M. (1984). Delayed signal detection, differential reinforcement, and short-term memory in the pigeon. Journal of the Experimental Analysis of Behavior, 42, 87–111.

Lockhart, R. S., & Murdock, B. B. (1970). Memory and the theory of signal detection. Psychological Review, 74, 100–109.

Luce, R. D. (1960). Detection thresholds: A problem reconsidered. Science, 132, 1495.

Murdock, B. B. (1965). Signal-detection theory and short-term memory. Journal of Experimental Psychology, 70, 443–447.

Ogilvie, J. C., & Creelman, C. D. (1968). Maximum-likelihood estimation of receiver operating characteristic curve parameters. Journal of Mathematical Psychology, 5, 377–391.

Peterson, G. B., Wheeler, R. L., & Armstrong, G. D. (1978). Expectancies as mediators in the differential reward conditional discrimination of pigeons. Animal Learning & Behavior, 6, 279–285.

Santi, A., & Roberts, W. A. (1985). Reinforcement expectancy and trial spacing effects in delayed matching-to-sample by pigeons. Animal Learning & Behavior, 13, 274–284.

Swets, J. A., Tanner, W. P., & Birdsall, T. G. (1961). Decision processes in perception. Psychological Review, 68, 301–340.

Urcuioli, P. J. (1991). Retardation and facilitation of matching acquisition by differential outcomes. Animal Learning & Behavior, 19, 29–36.

Wilson, B., & Boakes, R. A. (1985). A comparison of the short-term memory performance of pigeons and jackdaws. Animal Learning & Behavior, 13, 285–290.

Wixted, J. T., & Ebbesen, E. (1991). On the form of forgetting. Psychological Science, 2, 409–415.