

Numerosity differences and effects of stimulus density on pigeons' discrimination performance

JACKY EMMERTON

Purdue University, West Lafayette, Indiana

Two experiments are described in which pigeons were trained in a simultaneous conditioning procedure to discriminate small arrays of dots that differed in numerosity. The birds successfully learned to choose the array of each pair that contained fewer dots when these choices were reinforced and choices of the array with more dots led to timeout. For the majority of numerosity values tested, discrimination performance for a fixed S+ value was better when the numerical difference between S+ and S- values was larger rather than smaller. This effect was seen in the first experiment when the numerical difference value was shifted between training trials and novel test trials. In the second experiment, too, performance level depended on the size of the numerosity difference when the birds were concurrently trained with two difference values that varied across trials within sessions. However, discrimination accuracy was influenced secondarily by variations in the density, or interdot spacing, of the stimulus arrays. In order to explain the latter finding, it is suggested that a tendency to "scan" a low-density array incompletely might alter the probability of accepting it as the smaller numerosity (S+) stimulus. This would increase error rates with S- arrays in which the dots are more widely spaced.

Although there has been an interest in investigating animals' numerical abilities since at least the beginning of the century, research on this topic has gathered pace in recent years (see Rilling, 1993). Summaries and critical reviews of much of this work have been provided by Davis and Memmott (1982), Davis and Pérusse (1988), and several contributors to Boysen and Capaldi's (1993) book. Animals' "numerical competence," as it has been termed by Davis and Pérusse, includes a range of abilities. These vary from the capacity to make relative judgments between "more" versus "less," or "many" versus "few," through to the ability to count, or to recognize the absolute, or cardinal, number of items or events. Even in cases in which the stringent criteria for demonstrating counting (see Davis & Memmott, 1982) have not been met, claims that animals can make judgments of relative numerosity, or numerosity, imply that they discriminate among different stimuli on the basis of the number of items, rather than on the basis of other parameters that may easily covary with number. The problem of potential confounds with number was one of which Koehler and his co-workers were well

aware, and for which they made every effort to control when they laid the foundation for obtaining good empirical data in this field (e.g., Koehler, 1943; Koehler, Müller, & Wachholtz, 1935; Koehler & Wachholtz, 1936). In some respects, however, modern technology permits a greater degree of control than was possible in Koehler's pioneering research.

There have been two major approaches to providing evidence for numerical abilities in animals. In one line of research, their discrimination of the number of successive events has been studied. Suitable controls have shown that animals can evaluate the number of events separately from the duration of the series of events. Events have consisted either of external stimuli (e.g., light flashes, Alsop & Honig, 1991; Roberts, Macuda, & Brodbeck, 1995; or tones, Fernandes & Church, 1982) or of the animals' own responses (e.g., keypecks, Fetterman, 1993; or barpresses, Mechner, 1958). In the other line of research, animals have been trained to discriminate displays consisting of varying numbers of simultaneously presented visual items (e.g., Davis, 1984; Honig, 1993; Pepperberg, 1994; Thomas, Fowlkes, & Vickery, 1980). This is the approach on which the present investigation focuses.

Experiments to date with pigeons, a highly visual species and the one used in the present study, have shown that the performance of these animals on numerosity discrimination tasks is remarkably resilient in the face of changes in other attributes of multi-item stimulus displays. Honig and his co-workers (see Honig, 1992) have tested these birds' abilities to discriminate the proportion of items in regular matrices in which the total number of elements in a matrix was controlled. Pigeons' response rates were dependent on the relative, rather than absolute, numbers of "positive" compared with "negative" items in a given

I am grateful to Lori Pecknold for her assistance in running parts of these experiments and to Peter Urciuoli for his comments on an earlier version of the manuscript. The Purdue Research Foundation contributed a Summer Faculty Grant in support of this research. This work was also supported by a grant to J. D. Delius from the Deutsche Forschungsgemeinschaft while J. E. was a visiting professor at the University of Konstanz, Germany, where this manuscript was completed. Some of these data were reported at the 1992 annual meeting of the Psychonomic Society, St. Louis. Correspondence should be addressed to J. Emmerton, Department of Psychological Sciences, Purdue University, West Lafayette, IN 47907-1364 (e-mail: jemmert@psych.purdue.edu).

—Accepted by previous editor, Robert A. Rescorla

matrix. The discrimination gradients were largely independent of the total number of elements in a matrix, as well as of the type of elements in the stimulus displays (e.g., red and blue dots; Os and Xs; small pictures of birds, flowers, or unicorns; see Honig & Stewart, 1989). Similarly, in work by Emmerton, Lohmann, and Niemann (1997), pigeons have been trained to discriminate "many" (6 or 7) from "few" (1 or 2) items in different arrays, and then were tested both with new stimulus exemplars of those numbers as well as the intervening quantities (3, 4, and 5). A series of test conditions showed that numerosity discrimination functions were unaltered when a variety of attributes, including the shape, size, and contrast of constituent elements, were modified. Similarly, discriminability was unaffected when other parameters such as the overall brightness of each array, the summed contour in outline elements, or the total area of the constituent elements were equated across stimuli.

These experiments indicate that the rate of responding (Honig, 1992) or the allocation of conditional choices (Emmerton et al., 1997) varies as a function of stimulus numerosity. Further evidence that numerosity controls responding comes from Honig and Stewart's (1993) report of peak shift effects. They trained pigeons to discriminate arrays containing equal proportions of positive and negative elements (S+ stimuli) from arrays containing more of the negative elements (S- stimuli). When tested with arrays covering the full range of proportions of the mixed elements, the pigeons responded more to certain arrays containing more of the positive elements and fewer of the negative elements than they did to the original S+ mixture containing equal numbers of positive and negative items. This peak shift effect was again independent of the types of elements and occurred both with arrays consisting of red and blue dots and with arrays made up of small horizontal and vertical bars.

The present investigation was designed to extend such work on numerosity discrimination by presenting pigeons with pairs of small arrays consisting of varying numbers of dots. The birds had to differentiate these arrays in a simultaneous discrimination procedure. The first experiment was designed to test the effect on discrimination performance of a sudden shift in numerosity difference within the pairwise displays. The animals were initially trained to a particular criterion of intermediate discrimination accuracy with certain pairs of numerosity values constituting the S+ and S- stimuli. In later test phases, the numerosity difference was either increased or decreased by altering the value of the S-.

There are three possible outcomes of introducing this shift. The first is that discrimination performance with novel stimuli drops to chance level, implying that under the conditions of this experiment, birds do not learn information about numerosity or do not generalize it to new exemplars. The second is that pigeons acquire only limited binary categories of "less" versus "more," irrespective of the exact numerical difference within a stimulus pair. In

this case, test performance should be above chance, but should not differ from performance with familiar training stimuli. The third, and more interesting, possibility is that numerosity differences control differential responding in a more precise way. In this case, discrimination performance ought to be better with test stimuli that constitute a greater numerical difference than that in training pairs. Conversely, performance should decline if the numerosity difference in test stimuli is smaller than in training pairs. The third outcome was the predicted one since this would be in line with previous findings concerning pigeons' numerical abilities.

During this investigation it emerged that another parameter, the density of the arrays or their interdot spacing, could influence discrimination accuracy. The second experiment was in part designed to confirm the dependence of performance level on the size of the numerical difference. But the main purpose of this experiment was to examine more closely the effects of array density on discrimination accuracy. This is a factor that, to date, has not been systematically manipulated in experiments with animals. Some studies on humans, however, have shown that array density can influence perception of numerosity. The effects that were obtained in an experiment to be described here have potential implications for the way in which such multielement arrays are perceived and processed by birds. It will be argued that pigeons might "scan" such visual arrays. In doing so, they could process the constituent elements in a serial manner. This hypothesis implies that a common mechanism might underlie the numerical discrimination of both serial events and simultaneous arrays.

EXPERIMENT 1

Method

Subjects

Sixteen pigeons (*Columba livia*) of mixed breed served as subjects. All but 3 were experimentally naive at the start of the experiment. The 3 experienced birds had been used in a different laboratory as subjects in a differential outcomes experiment. The stimulus conditions and procedure were unlike those used in the present study, but to control for prior operant experience, these birds were allocated to different training groups. The animals were housed in individual cages in a colony room with a 14:10-h light:dark cycle. During the experiment they were maintained at 80% of their ad-lib body weights, but water and grit were freely available in their home cages.

Apparatus

The conditioning chamber (34 × 32 × 34 cm) was of in-house design. On the response panel were two horizontally arranged keys (3 cm in diameter) made of matt acrylic plastic. Their centers were 21 cm above the grid floor of the chamber and were 7.5 cm apart. Stimuli were displayed on these keys by opening solenoid-operated shutters located behind each key. The stimuli were back-projected onto the keys from a Zeiss-Ikon Unimat 2500 slide projector with a 70–120 mm zoom lens. Light intensity was reduced by diffusely spraying an internal collector lens with black paint. In the midline of the operant panel and 7 cm below the center line of the keys, a 5-mm² red light-emitting diode (LED) protruded through the panel. The LED could be illuminated and was equipped in such a way that pecks to it were registered as to a response key.

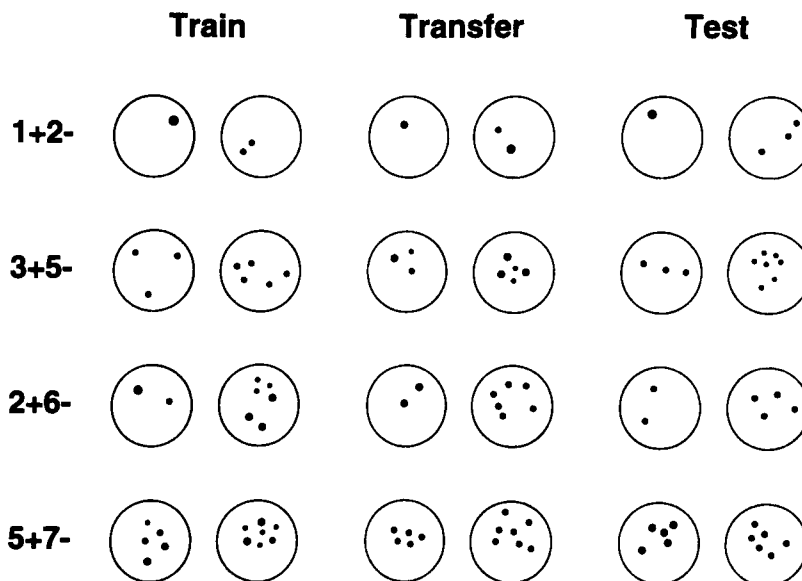


Figure 1. Examples of paired stimulus arrays used in the different phases of Experiment 1. Numbers beside each row indicate stimulus values for the four training conditions. In the transfer phase, new arrays with the same numerical values were introduced. In the test phase, novel arrays with altered values for the negative stimulus were included. (See Table 1 for stimulus values.) The positive value is on the left and the negative value on the right in each stimulus pair here. Actual stimuli consisted of white dots on a dark background. For illustration purposes, arrays are shown in reversed contrast.

The conditioning chamber was illuminated by a 1.7-W house-light centered 7 cm above the level of the response keys. A second light, directly beside the houselight, was switched on during food reinforcement periods. A small shelf just below these lights shielded the keys from scattered light. Access to a solenoid-operated food hopper was via an opening 6.5 cm below the LED. A C-64 Commodore computer, programmed in PSYCHOBASIC and connected to an interface (Xia, Wynne, von Münchow-Pohl, & Delius, 1991), was used to control events in the operant chamber, to drive the projector, and to register responses.

Stimulus configurations were prepared by hand as arrays of Letra-set black dots mounted on sheets of drawing paper. The dot configurations were designed to be random and variable. The size of dots also varied. Within a single array, dot sizes could be homogeneous or mixed. Within stimulus pairs, dot sizes could be the same across arrays, uniform within an array but different within a pair, or heterogeneous for both arrays of a pair. These arrays were then photographed and the photographic negatives were used as stimuli. Fixed pairs of stimuli were prepared by mounting negatives over each of two 1-cm-diameter holes in specially constructed 5×5 cm slide frames. When projected on the keys, the stimuli consisted of varying patterns of white dots on a dark background. The sizes of these dots ranged in diameter between 2 and 3.5 mm. Examples of paired stimulus arrays from different phases of this experiment are shown in Figure 1. Across all the experimental phases, there was a total of 60 stimulus pairs per numerosity condition, as will be described below. Thus, in the course of the first experiment, each bird experienced 480 different stimulus arrays, arranged in fixed pairs on 240 slides.

Procedure

Response training. All the pigeons first underwent a modified autoshaping procedure (Brown & Jenkins, 1968). Following a 20-sec intertrial interval (ITI), the keys were illuminated for 6 sec with a black cross-hatched pattern on a white ground. A peck to either key during this stimulus period produced immediate food reinforcement for 3 sec. If a bird failed to respond, reinforcement was

presented concurrently with the last 3 sec of the stimulus period. If an animal showed a strong preference for either the right or the left key, the preferred key was disabled in the next session. There were 40 trials per session. These sessions were continued for up to 15 days until each bird pecked consistently when the keys were lit.

They were then trained with an instrumental procedure to peck the red LED when it was illuminated. A single peck switched off the LED and a randomly chosen left or right pecking key was lit up with the cross-hatch pattern. A single peck to the lit key darkened it again and delivered food reward for 3 sec, followed by a 2-sec ITI. There were 40 trials per session and the birds were trained for five sessions with this procedure.

Discrimination training. During simultaneous discrimination training, each trial began with the illumination of the red LED. A single peck to it served as a "ready" response and activated the key shutters, revealing a pair of dot arrays. A single peck to the stimulus designated as the S+ (defined below) led to 3-sec access to food reinforcement, followed by a 2-sec ITI during which the next slide was set up in the projector. If instead the bird pecked the incorrect S- stimulus, a 3-sec timeout followed, during which the houselight was switched off. Then the 2-sec ITI ensued, with the houselight on. The same slide was presented on the next trial, which constituted a correction procedure. Correction trials could be repeated until the bird ultimately chose the correct stimulus. Choices on each trial were recorded but data from correction trials were not used to assess discrimination performance. Choice latencies between a bird's peck to the LED and its choice of a stimulus on the pecking keys were also recorded. Median latencies for correct choices varied between about 0.6 and 2.3 sec, but subsequent analyses failed to reveal any differential patterns linking choice latencies to stimulus values. They will therefore not be discussed further.

There were four different training conditions. The 16 subjects were assigned initially to one of four groups, matching them as far as possible for prior experimental experience and any key preference demonstrated during autoshaping. During the course of the experiment, all the birds were trained and tested under all four condi-

Table 1
Numerosity Combinations Used in the
Three Phases and Four Conditions of Experiment 1

Training Phase: S+/S- Stimuli	Transfer Phase: Novel Exemplars	Test Phase: Numerosity Shift
1+/2-	1/2	1/3
3+/5-	3/5	3/7
2+/6-	2/6	2/4
5+/7-	5/7	5/6

+Pecking was reinforced; -Pecking led to timeout. Only the values of nonreinforced stimuli are shown for transfer and test phases. These phases also included training stimuli (see text).

tions in a counterbalanced design. The numerosity combinations used in the four sets of training, transfer, and test conditions are given in Table 1, and examples of stimulus arrays are shown in Figure 1. The numerosities chosen for the stimulus arrays fell within a range that pigeons had previously been capable of discriminating (Knorn, 1987; see also Emmerton & Delius, 1993). The numerosity combinations were also designed to include both smaller (e.g., 1 vs. 2) and larger (e.g., 5 vs. 7) numbers of stimulus elements.

In training sessions, 40 pairs of stimuli were presented in quasi-random order (Gellermann, 1933). Within each pair, the stimulus consisting of the smaller number of dots was defined as the S+. The percentage of correct choices on noncorrection trials was calculated at the end of each bird's session. Training continued until a bird's score equaled or just exceeded 70% correct choices. The next phase of the experiment began as soon as a bird had met this individual learning criterion.

Transfer phase with novel exemplars. The next phase lasted for three sessions. Ten of the training slides were replaced by novel exemplars consisting of stimulus pairs with the same numerosity values as for the remaining 30 familiar training slides. The order of slide presentation was reshuffled according to Gellermann (1933) sequences. Whereas correct choices of the S+ numerosity values were rewarded on training trials, no reinforcement contingencies followed choices on the novel transfer trials. A single peck to either key simply terminated the trial, which was immediately followed by an ITI. The percent correct choices, defined as responses to the smaller numerosity, were calculated separately at the end of each session for familiar training and for novel transfer trials.

Test phase with numerosity shift. Upon completion of the transfer phase, the birds were tested in a further three sessions. In this phase, their discrimination performance was tested on trials in which the numerosity difference within a stimulus pair was either increased (in two testing conditions) or decreased (in the other two conditions; see Table 1). For these sessions, the 10 transfer slides used in the previous phase were replaced by 10 new test slides. The order of slide presentation was again changed according to Gellermann sequences. Correct choices of the smaller numerosity were still reinforced on familiar training trials. Choice of either stimulus had no reinforcement consequences on test trials, but merely terminated the trial and initiated an ITI. Again, the percent correct choices were calculated separately for training and for test trials.

After completing a numerosity shift phase, a bird underwent training with one of the other numerosity combinations, followed by the corresponding transfer and test phases, until all the birds had been exposed to all four conditions of the experimental design. Within each of these four conditions, one quasi-random sequence of slide presentation was used for all the training sessions, a different sequence for all three transfer sessions, and yet another sequence for the remaining three test sessions. Different sets of these Gellermann sequences were used across the four numerosity conditions so that the birds experienced 12 orders of slide presentation during the whole experiment.

Results

An overall summary of mean data across all 16 subjects is given in Figure 2. For each of the four numerical conditions, this figure presents the results from the training, transfer, and test subsets of trials in the transfer and test phases. It also shows the mean level of discrimination on final acquisition sessions in which the training criterion was achieved.

The data that were mainly of interest were those pertaining to nonreinforced trials with novel stimuli. These were included in the six consecutive sessions of the transfer and test phases for the four training numerosity conditions. For each condition, the data from nonreinforced trials were analyzed to determine whether there were any differences in discrimination performance across these six sessions. These analyses indicated whether the changes in numerosity difference between the transfer and test stimuli had any reliable effect on discrimination accuracy. Improvements in accuracy were found for increases in numerosity difference, as seen in the left-hand graphs of Figure 2. When numerical differences were diminished, however, the data analyses failed to support the predicted decrease in performance (see right-hand graphs of Figure 2).

For all these sessions, data from ongoing training trials were scored, too. These data served as baseline results against which transfer and then test scores were compared. Since these training trials were essentially an extension of the initial acquisition phase, the training scores alone were also analyzed to determine whether there was any significant improvement in the discrimination of training numerosity values. For most training conditions, these scores remained stable. The only significant improvement was found in scores for the 2+/6- condition.

Analyses of results will be presented first for initial acquisition and ongoing training trials and then for nonreinforced transfer and test trials.

Acquisition and Further Training

An acquisition criterion of at least 70% correct choices had been set to decide when each subject would enter the transfer phase. The time taken for the birds to reach this criterion varied across numerosity conditions. A Friedman two-way analysis of variance (ANOVA) by ranks (Siegel & Castellan, 1988), used to analyze the number of sessions for each bird to achieve criterion, revealed a significant difference in speed of acquisition across the four training conditions [$F_r(3, N = 16) = 10.0, p < .05$]. Birds took a mean of 1.9 sessions to reach criterion in the 2+/6- training condition, 4.1 sessions with the 3+/5- stimuli, 4.5 sessions for the 5+/7- discrimination, and 7.1 sessions with the 1+/2- stimuli. There was a similar result from an analysis of the number of sessions to reach criterion for only the first training condition per bird. A Kruskal-Wallis ANOVA of these initial acquisition data also showed a significant difference across numerosity conditions [$H(3, 16) = 10.8, p < .05$], with the smallest number of sessions (3.8) being needed to reach criterion

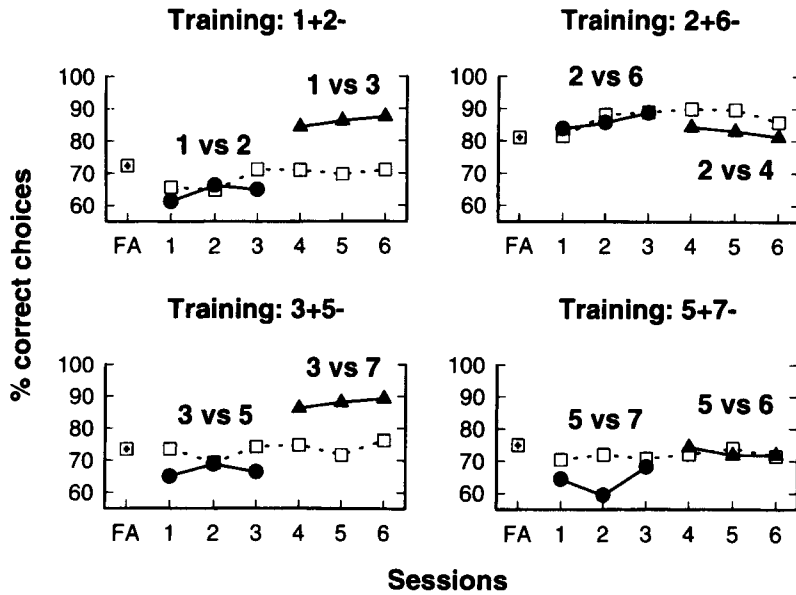


Figure 2. Mean scores on trials in transfer sessions (1–3) and test sessions (4–6) following each of the four training phases in Experiment 1. S+ and S– numerosity values on training trials are shown above each graph. Scores for reinforced training trials are indicated by unfilled symbols. Numerosity values in transfer and test phases are given above the curves with filled symbols that display results for nonreinforced trials. Choices on the latter trials were scored as correct when birds chose the smaller numerosity value corresponding to the S+ of training trials in each condition. The data point above FA on the abscissa is the mean score on the final acquisition session.

for the discrimination of 2+/6– and the largest number of sessions (14.0) for the discrimination of 1+/2–.

Scores on training trials across the six transfer and test sessions were examined with repeated measures one-way ANOVAs to see whether performance on these trials remained stable in spite of ongoing training. For conditions 1+/2–, 3+/5–, and 5+/7–, there were no reliable changes in performance [$F(5,75) = 1.7, 2.0, \text{ and } 0.4$, respectively, $p > .05$]. For the training condition 2+/6–, there was a significant change [$F(5,75) = 4.6, p < .01$], but Newman–Keuls post hoc tests showed that this was mainly attributable to the lower score on just the first session. Training scores were also well above the acquisition criterion, ranging between 81% and 90%, whereas they remained between 65% and 76% in the other three numerosity conditions.

Transfer and Test Scores

Next, similar ANOVAs were performed on the discrimination data from nonreinforced trials across sessions in the transfer and test phases. For the training conditions 1+/2– and 3+/5–, transfer trials included novel exemplars of the same numerosity pairings, whereas on corresponding test trials the numerosity difference within stimulus pairs increased. Performance on nonreinforced trials was significantly better in test than in transfer sessions for both of these conditions [$F(5,75) = 14.1, p < .01$ for data corresponding to the 1+/2– training condition; $F(5,75) = 15.7, p < .01$ for scores corresponding to the

3+/5– training condition]. Significant results ($p = .01$) in Newman–Keuls tests confirmed that discrimination with the novel slides and a larger numerosity difference (1 vs. 3 or 3 vs. 7) on each of the test sessions was better than that with novel slides (1 vs. 2 or 3 vs. 5) in each of the transfer sessions (see also Figure 2).

For the other training conditions 2+/6– and 5+/7–, the numerosity difference in novel test trials decreased compared with that used in novel transfer trials. Thus, poorer discrimination had been predicted on test trials. For the 2+/6– condition, although scores on test sessions for the numerosity combination 2 versus 4 were marginally lower than those on transfer sessions (2 vs. 6), an ANOVA revealed no significant difference across sessions in discrimination performance on these nonreinforced trials [$F(5,75) = 0.9, p > .05$]. For the condition 5+/7–, discrimination was, against expectation, slightly better in test sessions with numerosity values of 5 versus 6, compared with discrimination on the preceding transfer sessions with combinations of 5 versus 7 [$F(5,75) = 2.5, p < .05$]. Newman–Keuls tests revealed, however, that the only significant difference ($p = .05$) was between performance on the second transfer session and that on the first test session (Sessions 2 and 4 in Figure 2).

Comparisons Between Reinforced and Nonreinforced Trials

Further ANOVAs compared within-session discrimination scores for nonreinforced (novel stimulus) and re-

inforced (training stimulus) trials in the transfer and in the test phase. In the transfer phase, numerosity values in training and transfer trials were identical, so no differences in scores were expected between these trials. In fact, the only condition in which scores differed significantly was for 5 versus 7, in which performance on transfer trials was significantly poorer than on training trials [$F(1,15) = 11.0, p < .01$].

In the test phase, there was a change in numerosity difference between training and test trials. An increase in performance was expected on test trials, on which the stimulus numerosity difference exceeded that on training trials. Significantly better scores on test trials were indeed found [$F(1,15) = 64.3, p < .01$ for 1 vs. 3 compared with 1+/2- training trials, and $F(1,15) = 106.0, p < .01$ for 3 vs. 7 compared with 3+/5-]. Conversely, a decline in discrimination accuracy was predicted when the numerosity difference on test trials was less than that on training trials. For one condition (2 vs. 4 compared with 2+/6-), performance was poorer [$F(1,15) = 4.6, p = .05$]. However, there were no reliable differences in performance for the other condition [$F(1,15) = 0.01, p > .05$ for 5 vs. 6 compared with 5+/7-].

In all these analyses, there were no significant differences in overall percent correct choices across the three sessions of each phase, and interactions between trial type and sessions were not significant.

Discussion

When the numerosity difference within stimulus pairs was increased (from 1 vs. 2 to 1 vs. 3, and from 3 vs. 5 to 3 vs. 7), there was, as expected, a corresponding increase in the level of discrimination performance. This was against a background of stable scores on training trials throughout both the transfer and the test phases. The results were less clear-cut for the remaining conditions, in which numerosity differences decreased and in which a drop in performance level between transfer and test results had been expected.

In one of these conditions, the training and transfer numerosity values were 2 versus 6 and the test values were 2 versus 4. The analyses of results yielded only weak support for a decrease in performance as the numerosity difference decreased: This was in the comparison of scores for training and test trials in the three sessions of the test phase. However, set against this finding was a lack of statistical difference between transfer and test trials. One difficulty in this experimental condition was that acquisition scores rapidly exceeded the learning criterion, and then there was some further improvement in discrimination on the ongoing training trials. Compared with all the other numerosity conditions, scores on all types of trials were high throughout the transfer and test phases. The apparent ease with which all the numerosity combinations were discriminated in this condition may have masked any tendency for the birds to show reliable changes in performance.

The data for the remaining experimental condition, which began with training on 5+/7-, yielded an unex-

pected effect. In spite of a decrease in the numerosity difference from 5 versus 7 on training and transfer trials to 5 versus 6 on test trials, test and training results were similar in the test phase, but the test data were actually better than transfer results when only the nonreinforced trials across all sessions were considered. The latter finding is probably mainly due to scores on transfer trials being worse than, rather than similar to, scores on training trials in the initial transfer phase, even though the numerosity values on all trials were identical.

In an attempt to understand why the transfer and test results following training in the 2+/6- and the 5+/7- conditions were not clear-cut, the number of sessions the birds had needed to reach the initial learning criterion of 70% correct choices was considered. It had seemed possible that the poor results in the 5+/7- condition were due to particular difficulties in acquiring this discrimination. However, the results did not support this idea.

Not surprisingly, acquisition proved to be quickest in the 2+/6- condition, when acquisition times were considered after all the birds had been trained under all four conditions. This tallies with the ongoing ease of discrimination in this condition. The next easiest discrimination to acquire was with stimuli of the 3+/5- type. However, it took the birds only slightly longer to learn the 5+/7- discrimination. The birds took the longest to reach criterion with the 1+/2- discrimination. Similar results were seen when the mean number of sessions to criterion in the very first training condition for each group of birds was considered. Although the discrimination of 5+/7- did not seem to have been particularly difficult for the birds, the discrimination of 1+/2-, assumed to be an easy one, had unexpectedly taken them much longer to acquire. Why this should be so was not apparent by the end of this first experiment. Discussion of this point will be deferred until later.

By and large, choice accuracy was influenced by the numerosity differences within stimulus pairs. However, examination of some trial-by-trial data suggested that another factor might have influenced discriminability. Stimuli for presentation on different trials had been prepared so that their configurations varied, leaving only their numerosities as the intended characteristic that certain stimuli had in common. In some instances, a variation in the configuration of component dots also involved altering the spacing of these elements. For the 1+/2- condition, obviously the density of only the S- could vary. A post hoc assessment of responses with just a couple of stimulus pairs revealed that more errors were made when the dots in the S- stimulus were widely spaced than when they were closely spaced. For most of the stimulus pairs, particularly those in which both S+ and S- consisted of multiple elements, a clear-cut classification of the stimuli in terms of their densities was not possible. Instead, a further experiment was conducted, with the chief aim of examining the effects of array densities on discrimination performance, as well as assessing once more how choice accuracy levels relate to numerosity differences.

EXPERIMENT 2

The effect on relative numerosity judgments of varying interitem spacing has been examined in several studies with human subjects. Probably the best known effects are those that were reported by Piaget (1952) when he studied the numerical abilities of young children at a "pre-observational" stage of development. He found that if objects were presented spread out in a row, young children judged them to be more numerous than if the same number of objects were lined up close together. Most studies with adults have used 2-D arrays of dots as stimuli. However, stimulus presentation in these studies has differed in two respects from the displays used in the present experiments. With humans, the arrays usually consist of a larger number of items (varying between about 15 and 200 in different experiments). The stimuli are also presented only briefly (for times varying between about 150 and 750 msec). The use of large numbers of dots and brief display times is necessary with human subjects to prevent their enumeration or one-by-one counting of the dots.

Nevertheless, some recent studies have shown fairly consistent effects of dot spacing on relative numerosity judgments (decisions about whether a test display contained more or fewer items than a reference display). Mostly, it has been found that when dots are fairly evenly spread out within a display area, their numerosity is overestimated compared with equally numerous displays in which the dots are randomly distributed. Random configurations, in turn, are judged more numerous than clustered dot displays (see, for instance, Krueger, 1972; Vos, van Oeffelen, Tibosch, & Allik, 1988; see also Allik & Tuulmets, 1991a, 1991b). The overall size or edge-to-edge extent of the display does not seem to be critical (Allik, Tuulmets, & Vos, 1991), but rather the sum of the areas within the display that are perceived as being "filled" with dots. Whereas some authors have simply made empirical observations (e.g., Krueger, 1972), Allik and Tuulmets (1991a) developed an algorithm to account for these effects. Their "occupancy model" postulates that each dot has an impact on the surrounding area according to a fixed spread function of visual effect. If the dots' "occupancy territories" do not overlap, their summed effects will be greater than if there is substantial overlap of the spread functions among neighboring dots. Hence dots that are fairly well spread out will be perceived as being more numerous than ones that are very closely spaced.

Although interdot spacing has been reported to have differential effects on these types of "more" or "less" numerosity judgments with humans, not all of the effects are consistent. An older study by Mokre (1928) used a moderate number of dots (a reference number of 15) and varied their diameters and their regular center-to-center spacing. Dot spacing seemed to be more important than dot size in influencing numerosity judgments. Nevertheless, the effects of spacing were not consistent across subjects. Only 2 of his 7 subjects showed effects similar to those reported in more recent studies, with a tendency

to overestimate relative numerosity with increasing interdot spacing. For 3 subjects, there was a tendency in the opposite direction—to overestimate numerosity when dots were more closely spaced and underestimate the relative number as spacing increased. For the remaining 2 subjects, varying the density of the displays appeared to have very little effect on numerosity judgments.

So far, no systematic attempts have been made with animals to examine the effects on numerosity discrimination of varying the display density. As mentioned previously, informal analysis of some partial data in the present study suggested that this variable might at least modulate pigeons' ability to differentiate arrays of different numbers of dots. The next experiment was therefore designed to explore this issue. To provide data that were comparable with those of the first experiment, the numerical range of dots in the stimulus displays remained the same. However, since discrimination in one condition had seemed exceptionally easy, that numerical condition was changed to assess the birds' discrimination of 2 dots from 3 or 4 (rather than 4 or 6, as in Experiment 1). The combination of 5 versus 6 or 7 dots was retained. In part, this was to avoid going beyond the range of numerosities that pigeons had previously been able to discriminate with similar displays (Knorn, 1987; see also Emmerton & Delius, 1993). This combination was also used to see whether more consistent data could be obtained with these numbers of items. A further modification was introduced in the next experiment: In order to assess birds' choice behavior when the numerosity difference was varied, performance levels were compared within sessions when two levels of numerosity difference were intermixed quasi-randomly across trials.

Method

Subjects

One of the birds became sick after the completion of the first experiment. Two of the other birds in the first experiment had consistently been the slowest to complete training in all of the numerosity conditions. For the second study, the 12 birds that had been quickest in acquisition were chosen as subjects in order to reduce individual variability in learning rates. They were randomly assigned to four different groups.

Apparatus

The apparatus was the same as in the first experiment. New sets of slide stimuli were prepared. Although the stimulus drawings were prepared by hand so that the positioning of the dots was determined subjectively, the stimuli were designed so that the dot density within each multiple-dot array was manipulated. The dots were arranged so that when they were projected on the response keys, they would be bunched near together (the N or high-density condition) or they would be spaced far apart from each other (the F or low-density condition). Interdot spacing was not constant but was slightly irregular within each slide. Obviously, density could not be manipulated when stimuli consisted of only 1 dot. In that case, the dot's size and position on the key was varied. In multiple-item stimuli, dot size could also vary. As in the first experiment, dot sizes could be the same or different, either within a single array or across pairs of arrays. Fixed pairs of arrays were used so that each slide was unique.

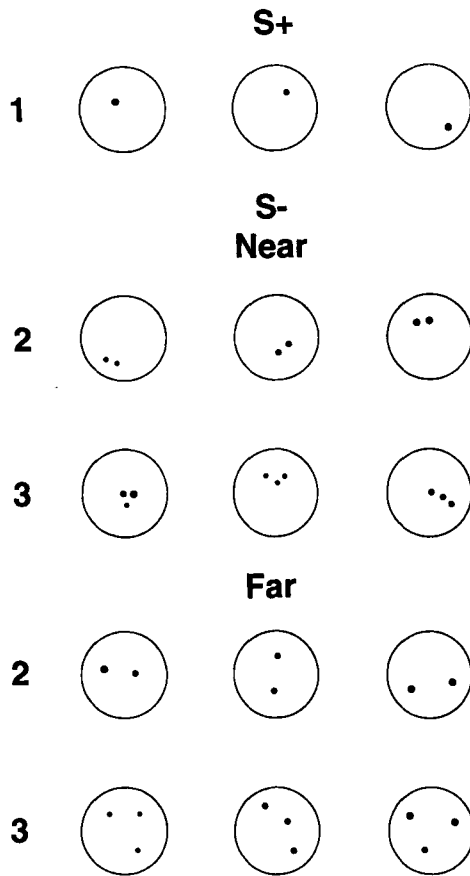


Figure 3. Samples of the S+ and S- stimuli used in the numerosity pairings 1+2- and 1+3- in Experiment 2. S- stimulus dots were either situated near together or were spaced far apart. For illustration purposes, the contrast of the arrays has been reversed.

Stimuli were prepared with the numerosity combinations 1+2-, 1+3-, 2+3-, 2+4-, 3+5-, 3+7-, 5+6-, and 5+7-, where + denotes the S+ value, and -, the S- value. When the S+ consisted of only 1 dot, there were two density arrangements for each numerosity combination: S- could be of the type N (near spacing) or the type F (far spacing). Examples of these stimuli are shown in Figure 3. In all the other stimuli, both the S+ and the S- consisted of multiple dots. Each of these stimulus components could consist of high- or low-density dot arrays. Thus S+/S- pairings could be of four density combinations: N+/N-, F+/F-, N+/F-, and F+/N-. Examples of multiple-dot arrays in different density arrangements are shown in Figure 4.

Procedure

The simultaneous discrimination training procedure was the same as in the previous experiment. Correct choices on all trials were followed by food reward. Incorrect responses led to a timeout and initiated a correction procedure. At the end of each session, separate percentages of correct choices on noncorrection trials were calculated for the two intrasession numerosity combinations. Sessions consisted of 32 noncorrection trials and a variable number of correction trials.

The numerosity value of the S+ was the same on all trials within a session, but this number was combined with two different S- values (e.g., 1+2- and 1+3-; see Figure 5). These two numerosity combinations were presented in quasi-random order, with 16 slides in each combination. In each set of 16 slides with the combinations

1+2- and 1+3-, half were of the low-density type (F-) and half of the high-density type (N-). For all the other training conditions, there were again two numerosity combinations per session, with 16 slides in each combination. Of these 16 slides, there were four sub-groupings of slides with the following different density combinations: N+/N-, F+/F-, N+/F-, and F+/N-. Again the order of slide presentation was quasi-random. The four groups of birds were trained for 10 sessions in each of four numerosity conditions that can be designated by the S+ values: 1+, 2+, 3+, or 5+ (each of which was associated with two different S- values, as mentioned above). The order of training was counterbalanced across groups.

For each numerosity condition, two slide sequences were used. The first sequence was used for the first five sessions, and then the slide order was changed according to the second sequence and stayed the same for the remaining five sessions. Across all four training conditions, each bird thus experienced 8 different quasi-random orders of stimulus presentation.

Results

For each bird and session, the percentages of correct choices made on subsets of noncorrection trials were calculated. Data subsets were grouped according to the numerosity and dot density combinations used on individual trials (e.g., 3+7- F+/N- constituted a subset).

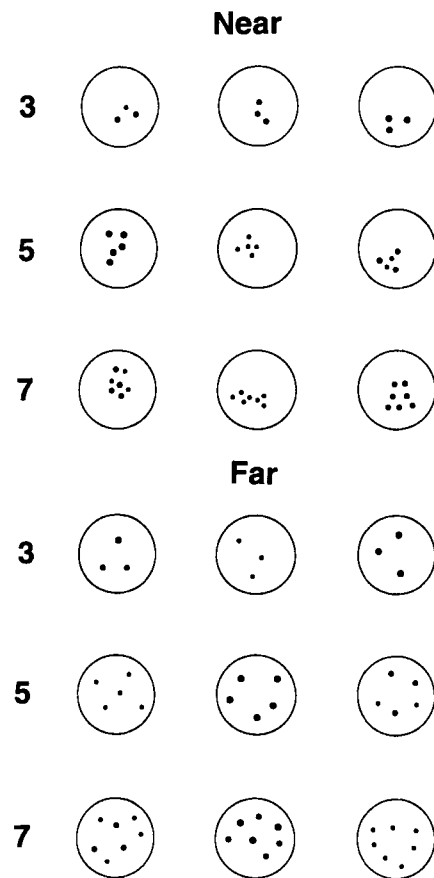


Figure 4. Samples of multiple-dot arrays used as either S+ or S- in Experiment 2. In both S+ and S- stimuli, dots were placed either near together (N) or far apart (F). This yielded four possible density combinations: N+N-, F+F-, N+F-, and F+N- (see text for further details). For illustration purposes, the stimulus contrast has been reversed.

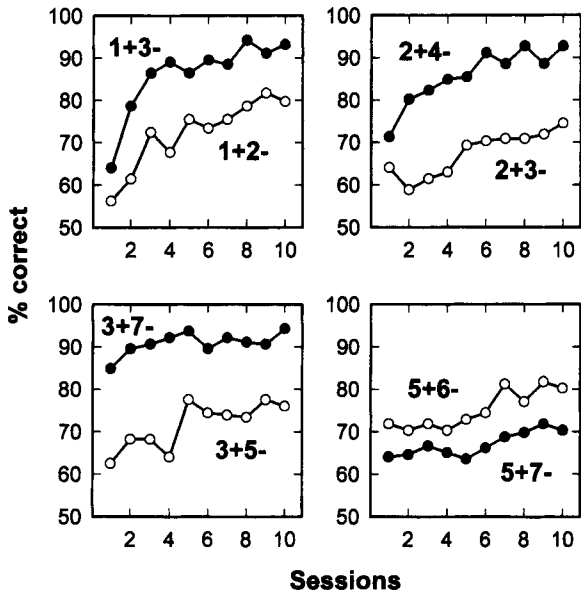


Figure 5. Mean discrimination scores obtained in each of the four numerosity conditions in Experiment 2. In each condition, the numerosity of the S+ was the same on all trials but the S- numerosity differed across trials within the same session. Each graph shows the average discrimination performance for these two types of trials. Numerosity values of S+ and S- stimuli are indicated next to the mean data curves.

For each S+ numerosity condition, three-way ANOVAs were then performed, with numerosity combinations, dot density combinations, and sessions as repeated measures. Each of these main factors had a reliable effect on discrimination performance. In some cases, there were also significant interactions, chiefly between display densities and some of the numerosity conditions. The outcome of these analyses will be described in more detail below.

Effect of Numerosity Difference on Performance

The mean discrimination scores across 10 sessions for the various numerosity differences are summarized in Figure 5. For three out of the four sets of numerosity combinations, the greater the numerosity difference on trials within a session, the better the birds' discrimination accuracy. For all numerosity combinations, performance level also improved across sessions.

For the training combinations 1+2- compared with 1+3-, 2+3- versus 2+4-, and 3+5- versus 3+7-, choice accuracy depended significantly on the numerosity difference, with better discrimination corresponding to the greater numerosity discrepancy in each condition [$F(1,11) = 89.7, 138.8, \text{ and } 94.7$, respectively; $p < .01$ in each case]. For the numerosity pairs 5+6- and 5+7-, there was a significant difference in discrimination level between these two stimulus conditions [$F(1,11) = 10.5$, $p < .01$]. However, mean discrimination accuracy in this case was higher with the smaller numerosity difference

(5+6-) than with the larger difference (5+7-). Discrimination accuracy improved significantly across sessions for all four sets of training combinations [$F(9,99) = 17.7, 7.2, 4.8, \text{ and } 3.0$ for Conditions 1+, 2+, 3+, and 5+, respectively; $p < .01$ in each case]. There were no significant ($p > .05$) interactions between numerosity conditions and sessions.

Effect of Array Density on Performance

The other main factor was the relative spacing or density of the stimulus dots. Overall, in each of the four sets of numerosity combinations, this factor had a significant effect on performance. For three of these conditions, there was also a significant interaction between the numerosity difference that had to be discriminated and the density of the stimulus arrays. An overall summary of these interaction effects is provided in Figure 6. Since the density of the S- stimulus in each pair seems to have been particularly important in producing these results, performance is encoded by different types of symbols in the figure. Filled symbols have been used for conditions in which S- arrays were of high density (dots were near together). Unfilled symbols represent conditions in which S- arrays were of low density (dots were far apart).

In the numerosity combinations 1+2- and 1+3-, obviously the density could only vary in the S- stimuli. The overall effect of S- density on discrimination accuracy was significant [$F(1,11) = 23.1$, $p < .01$], with better discrimination being achieved when the S- elements were near together. In addition, there was a significant interaction between numerosity and density conditions [$F(1,11) = 10.5$, $p < .01$].

In all the other numerosity conditions, dot spacing varied in both S+ and S- stimuli. In all cases, there was a reliable effect of density variations on overall discrimination performance. For intrasession numerosity combinations 2+3- and 2+4-, 3+5- and 3+7-, and 5+6- and 5+7-, the respective F ratios were $F(3,33) = 15.4, 7.2, \text{ and } 8.8$; $p < .01$ in each case. For training with the 2+3- and 2+4- stimulus pairs, there was a significant interaction between numerosity pairing and stimulus density [$F(3,33) = 9.0$, $p < .01$]. This interaction term did not reach significance for the stimulus pairs 3+5- and 3+7-. The interaction of numerosity difference and stimulus densities was again significant for the training pairs 5+6- and 5+7- [$F(3,33) = 15.6$, $p < .01$]. The latter training condition, with S+ = 5, was the only one in which a significant interaction was also found between density conditions and training sessions [$F(27,297) = 2.5$, $p < .01$]. None of the three-way interactions of numerosity \times density \times sessions were significant ($p > .05$).

Rank Order of Density Effects on Discrimination With Multidot Numerosity Pairs

When both discriminanda consisted of multiple elements, the densities of the S+ and S- stimuli also affected discrimination accuracy in the separate numerosity pairings, as can be seen in Figure 6. On the whole, a consistent

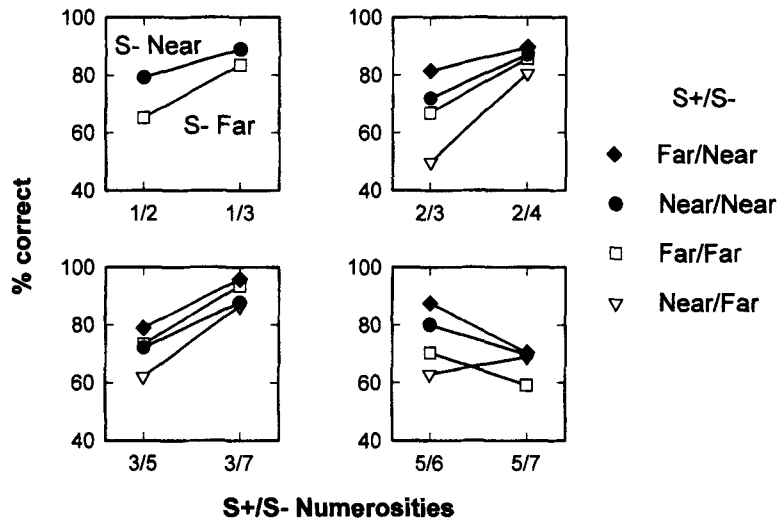


Figure 6. Interactions between numerosity and dot-density combinations in the four training conditions of Experiment 2. Filled symbols show scores for density conditions in which the S- array was of the near type (with closely spaced dots). Open symbols show scores for combinations in which S- arrays were of the far type (with widely spaced dots). When S+ consisted of just 1 dot (top-left graph), only the density of the S- stimulus could vary. For all the other numerosity conditions, there were four density combinations. These are coded by separate symbols as shown to the right of the figure.

pattern emerged in the results. For each of the separate numerosity pairings (e.g., 2+/3- alone), the best discrimination scores were found for the F+/N- density condition, when the S+ was of the far type (with dots spaced far apart) and the S- was of the near type (with dots close together). With the exception of the numerosity pairing 5+/7-, discrimination accuracy was worst under the N+/F- density condition.

Further two-way repeated measure ANOVAs were performed post hoc in order to compare the effects of density conditions by applying subsequent Newman-Keuls tests. In the majority of cases, Newman-Keuls tests showed that the differences in performance level under the two conditions F+/N- and N+/F- were significant ($p < .05$ for 2+/4-; $p < .01$ for 2+/3-, 3+/5-, 3+/7-, and 5+/6-). The only inconsistency in this general pattern of results was that, for the numerosity pairing 5+/7-, discrimination accuracy was worst under the F+/F- condition, when both S+ and S- stimuli were of the far type, with dots spaced apart. For 5+/7-, the difference in discrimination accuracy under these best (F+/N-) and worst (F+/F-) density conditions was also significant ($p < .05$). For most numerosity combinations, intermediate levels of discrimination were obtained when S+ and S- had similar densities (i.e., the combinations N+/N- and F+/F-). However, whether performance was relatively better, worse, or approximately equal when the N+/N- and F+/F- conditions were compared varied across numerosity pairings.

Discussion

Most of the data in this experiment confirmed that, for a given S+ value, discrimination accuracy was greater

when the numerosity difference between S+ and S- was larger than when it was smaller. Although numerosity differences were apparently the primary factor determining the pigeons' choice behavior, their discrimination performance was also influenced by variations in the spacing of items in the stimulus arrays. However, the effects of dot spacing appeared to be very different from those reported in most studies of numerosity discrimination in humans. In part, this may be due to the differences in stimulus presentation conditions for humans and animals that were noted earlier. Be that as it may, in humans, numerosity tends to be overestimated when dots are spread out and underestimated when they are clumped together (compared with estimates made with random configurations). If similar effects had operated in birds, then arrays of the far type should have had their apparent numerosity enhanced in comparison with the same numerosity of the near arrangement. Thus, in the combinations 1+/2- and 1+/3-, discrimination should have been better when S- was of the far type rather than the near type. Instead, the opposite was found. Furthermore, in combinations in which both S+ and S- consisted of multiple dots, and choice of the smaller numerosity was always correct, discrimination should have been optimal with the density combination N+/F-. For most numerosity pairings (with the exception of 5+/7-), this was the combination that instead produced the poorest discrimination accuracy, whereas the spacing combination that led to optimal performance was F+/N-. Thus, at the descriptive level, the data show opposite effects of interitem spacing on numerosity assessment in humans and pigeons.

GENERAL DISCUSSION

Exactly how animals process stimulus arrays to assess their numerosity is unknown at present. However, analyses of the general processes underlying other discrimination tasks suggest that choice behavior depends on a combination of stimulus effects, often represented in signal detection models by a normal distribution, and decision criteria. The model that best describes the behavior of pigeons when they are acquiring a discrimination task or have reached stable performance levels is one based on Markov choice processes that has been developed by Wright (1990, 1992; Wright & Sands, 1981). Although this model was derived from empirical data obtained in matching-to-sample and oddity-from-sample tasks, its general principles should apply to the simultaneous discrimination procedure used in the present experiment.

Wright (1990, 1992) observed that pigeons sometimes pecked at the first comparison stimulus they viewed, but would also switch back and forth, perhaps several times, between the two comparison stimuli before finally pecking a key. To demonstrate this, he recessed the stimuli behind the keys and monitored the birds' behavior with an infrared camera. Such an arrangement was not technically feasible in the present experiment, but may not have been necessary for the present argument if just the main features of Wright's model are considered. In the Markovian model, the probability of accepting or rejecting each stimulus is independent of the number of switches in observing behavior. In fact, Wright's (1990, 1992) empirical data show that, although multiple switches do occur, the majority of pecking responses are made to the first stimulus the bird views, or else to the alternative stimulus that it examines after making a single switch between keys. Thus, as a simplification, these initial tendencies may be considered to account for the broad trend of the discrimination results.

Wright's (1990, 1992) discrimination model assumes distributed stimulus effects as well as a decision criterion (or acceptance/rejection level) at which each stimulus triggers a response. However, in contrast to signal detection models, the acceptance criterion stays relatively stable with respect to the distribution of S+ effects. What changes is the relation between the decision criterion and S- effects. As discrimination improved during acquisition, the probability of rejecting the S- when it was viewed first increased. Conversely, the likelihood of its being falsely accepted decreased (Wright, 1992). Choice behavior was also analyzed in well-trained birds when the physical difference between stimuli (wavelength values of color stimuli) varied (Wright, 1990). If the birds examined the S+ first, they mostly responded to that stimulus rather than switching to look at the other one. This acceptance probability remained stable across different levels of discrimination difficulty. What varied was the behavior toward the S-, if that stimulus was observed first. When the discrimination task was an easy one (i.e., the physical difference between S+ and S- was

large), there was a fairly low proportion of immediate choices of the S-, or false acceptances. This meant that there was a high proportion of correct rejections that resulted in a single switch to and correct acceptance of the S+. Conversely, if the discrimination task was more difficult (i.e., the difference in stimulus values was smaller), then the opposite pattern of behavior was observed: There was a greater occurrence of false acceptances on first viewing the S- and relatively fewer correct rejections with subsequent choices of the S+.

Throughout this study, each bird was trained with the same procedure, not only in the two experiments described here, but also in some intervening control conditions that were not reported. The animals received equivalent amounts of experience with each of the stimulus sets used here. But since performance accuracy varied in a fairly consistent way, there must be something about the stimuli themselves that leads to differences in their acceptance or rejection rates. At first, and in keeping with Wright's (1990, 1992) model, the results appeared to be explicable through differences in the processing of the S- stimuli alone. One possibility is that the birds "scan" across a key (probably by making eye and/or head movements in front of the key—see Goodale, 1983—or else through shifts of attention within their binocular visual field). S- is always the stimulus whose numerosity exceeds the fixed value of the S+. Thus, when S- dots are more widely spaced, there is an increased chance of missing some dots or making a false acceptance response before all the dots have been encountered. The greater the number of dots in the S-, the greater the likelihood that a number in excess of the S+ value will be encountered and the S- will be correctly rejected. Hence, choice accuracy should be greater when the numerosity difference between S+ and S- is larger. For most numerosity conditions, this was the case. But also, for a given numerosity difference, the closer the dots' proximity, the greater the probability that each will be detected, in which case an S- is more likely to be correctly rejected. On the other hand, the farther apart the dots, the more likely it is that one or more might be missed, increasing the false acceptance rate to the S-. This would be sufficient to explain the pattern of interactions between numerosity and dot density for the 1+/2- and 1+/3- training conditions.

This proposition, that low dot density and incomplete scanning of the S- increase error rates, would also account for the two most consistent findings with the remaining numerosity combinations. The first is that discrimination accuracy is generally better when the numerosity difference is greater. The second is that, within a given numerosity pair, performance is best with the density combination F+/N- and (with one exception) worst with the combination N+/F-. However, this proposition alone does not quite account for the results with the remaining density combinations. If differences in choice accuracy were to be attributed solely to differences in the probability of processing errors with the S-, then results for the combinations F+/N- and N+/N- should be

equivalent. Similarly, discrimination levels should be equivalent for the F+/F- and N+/F- combinations. The data provided mixed evidence on this point.

If an additional factor has to be postulated that affects the overall decision process, there are two possibilities. The more parsimonious one is complementary to the effects that spacing might have on S- decisions. Thus, with F+ stimuli, a bird could also accept a stimulus as having the smaller number of dots before it had completely scanned all of them. If this happened, correct acceptance rates to F+ stimuli would be slightly enhanced. The other possibility is that initial approach to one or other key is biased by stimulus factors. In Wright's (1992; Wright & Sands, 1981) experiments with recessed stimuli, the birds had to approach one key first and look through it before they could see a stimulus. Preferences for investigating a particular side first differed among animals, so initial key approach was essentially a random variable. In the present experiment, it was not technically feasible to recess the stimuli, nor was it possible to monitor which key the pigeons looked at first. Most likely both keys were visible from a distance since the birds were free to move around in the conditioning chamber. It is therefore possible that they made a quick assessment of the location where there appeared to be only a few stimulus items (maybe using lateral vision to do so; see Friedman, 1975; Goodale & Graves, 1982). If the local dot density was low (and a bird noticed just a few dots on the key), this could have been used as a cue to approach that key and examine it more closely. Then the pigeon would have had to switch to frontal, binocular vision, on which it relies when viewing things close up and pecking them (see Goodale, 1983). Initial approach could thus be biased toward the key with the far stimulus on it only if local densities between keys differed. This would imply a bias to approach F+ with the F+/N- combination but not with the F+/F- combination. Similarly, there would be a bias to approach S- with the N+/F- combination but no approach bias with the N+/N- combination. As a second stage in the decision process, a close examination of the approached key would lead to a response if the bird accepted the stimulus as the correct one, or to a switch to the other side if it rejected the stimulus as the incorrect one. The present data cannot differentiate between these possibilities, but they may be worth considering in future experiments.

A couple of findings in these two experiments are harder to explain. Mostly, discrimination was easier when the numerical difference between paired arrays was larger rather than smaller. The exception to this was with the numerosity combinations 5+/6- and 5+/7-. The pattern of results for the latter numerosity pairing also deviated from that of other combinations when the effects of density on discriminability were considered. At present, the reason for these discrepant results is not clear.

The other unexpected finding was that it took the birds longest to acquire the 1+/2- discrimination in the first experiment. This effect was seen both in the overall mean acquisition times and in the time to reach the learning criterion when the birds' initial training was with 1+/2-.

This slower learning might be related to the feature negative effect (Jenkins & Sainsbury, 1970) if a single dot is considered as the common feature on both keys and a second dot is considered as a distinctive feature. However, the effect appears to be restricted to this numerosity condition since, with multiple elements on both keys, the birds had no difficulty in learning to choose the key with the lesser amount on it. Jenkins and Sainsbury favored a simultaneous discrimination theory over a search theory to explain the feature negative effect. According to the simultaneous discrimination theory, both common and distinctive features are perceived by an animal as soon as a stimulus is displayed. As discrimination learning proceeds, however, these features are associated with differential reinforcement. Responding to the common feature increases, but then is maintained by partial reinforcement, since this feature appears in both S+ and S-. Responses to the distinctive feature in the S-, though, have a zero probability of being reinforced. The search theory proposes that the gross features, common to displays, are initially processed and only through differential reinforcement do the animals come to search for the distinctive feature. If the distinctive feature appears in the S-, nonreinforcement will not elicit attention to this differential feature. The present hypothesis is more akin to Jenkins and Sainsbury's search theory in proposing that displays have to be scanned in order for a bird to pick up information about all their features. But the further assumption must be that scanning is curtailed when the reinforced stimulus consists of a single dot, rather than several (as in the other numerosity conditions). However, because of the variable stimulus configurations, the distinction between common and distinctive features in the present experiments becomes rather abstract. Across stimuli, dots varied in size and location, so it is only the supernumerariness of a dot that makes it distinctive. Jenkins and Sainsbury, by contrast, generated stimuli in which the distinctive feature was some consistent physical characteristic, such as a gap in a line or a star shape among dots.

In both experiments, the numerosity difference between arrays nevertheless seemed to be the most important factor guiding birds' choices in the simultaneous discrimination task. This adds to the accumulating findings that a variety of species are capable of responding differentially to the relative or absolute number of items that are presented either simultaneously, as here, or sequentially (see Boysen & Capaldi, 1993; Davis & Pérusse, 1988). With the few sets of numerosity values that were used in the present experiments, it is not certain whether the birds learned about the absolute or the relative difference between each S+ and S- value. Some post hoc calculations suggest that the birds relied on relative numerosity differences. For Experiment 1, disparity ratios (the difference divided by the sum of paired numerosities) were calculated for each of the paired values in transfer and in test trials. The overall mean percent correct scores for each set of nonreinforced trials were transformed to a logit scale (Macmillan & Creelman, 1991). The correlation between disparity ratios and logit scores proved to

be significant ($r = .74, N = 8, p < .05$). For Experiment 2, a similar logit transform was applied to the overall mean percentages across the last five sessions, in which performance was tending to level off, for each pair of numerosity values. The disparity ratios for these numerosities were also calculated. Again, the correlation coefficient for the numerical ratios versus logit discrimination scores was significant ($r = .78, N = 8, p < .05$). These correlations suggest that Weber's law might apply to simultaneous numerosity discrimination, as it does to numerical discrimination of sequential events (Fetterman, 1993; Meck & Church, 1983).

Previous research with pigeons has shown that the form of discrimination functions remains fundamentally the same when other stimulus parameters, that are not essentially numerical, are altered. Particularly in an experiment by Emmerton et al. (1997), a wide variety of potentially confounded attributes were manipulated without having any systematic effect on discrimination functions. Although there were slight random variations across stimuli in the spacing of elements, those variations were not correlated with the numerosity values. The present experiments show that, although dot spacing is not the primary factor guiding choice behavior, it can nevertheless influence the accuracy of numerosity discrimination. Although this statement summarizes the empirical findings, the explanation for the effects of array density remains speculative for now. Nevertheless, the working hypothesis that is suggested here rests on a few assumptions that are compatible with other observations about visual behavior and the nature of the discrimination process in birds.

If, indeed, birds "scan" visual arrays and thereby process separately each of the constituent items, this would have further implications for the type of mechanism that underlies numerosity assessment of such stimuli. A scanning strategy implies serial processing of simultaneously presented array elements so that each encounter with an element could be considered as an "event." Meck and Church (1983) proposed a model that involves the accumulation of gated pulses from an oscillator to account for both counting of events and timing of series duration when rats discriminated tone sequences. This model has been modified by Roberts and Mitchell (1994) to explain data from pigeons that were tested with sequences of light flashes. A model for serial counting might also apply to pigeons' processing of multi-item visual arrays.

REFERENCES

ALLIK, J., & TUULMETS, T. (1991a). Occupancy model of perceived numerosity. *Perception & Psychophysics*, **49**, 303-314.
 ALLIK, J., & TUULMETS, T. (1991b). The perception of visual numerosity. In R. Watt (Ed.), *Vision and visual dysfunctions. Vol. 14: Pattern recognition by man and machine* (pp. 125-142). London: Macmillan.
 ALLIK, J., TUULMETS, T., & VOS, P. G. (1991). Size invariance in visual number discrimination. *Psychological Research*, **53**, 290-295.
 ALSOP, B., & HONIG, W. K. (1991). Sequential stimuli and relative numerosity discriminations in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, **17**, 386-395.
 BOYSEN, S. T., & CAPALDI, E. J. (Eds.) (1993). *The development of numerical competence: Animal and human models*. Hillsdale, NJ: Erlbaum.

BROWN, P. L., & JENKINS, H. M. (1968). Auto-shaping the pigeon's key peck. *Journal of the Experimental Analysis of Behavior*, **11**, 1-8.
 DAVIS, H. (1984). Discrimination of the number three by a raccoon (*Procyon lotor*). *Animal Learning & Behavior*, **12**, 409-413.
 DAVIS, H., & MEMMOTT, J. (1982). Counting behavior in animals: A critical evaluation. *Psychological Bulletin*, **92**, 547-571.
 DAVIS, H., & PÉRUSSE, R. (1988). Numerical competence in animals: Definitional issues, current evidence, and a new research agenda. *Behavioral & Brain Sciences*, **11**, 561-615.
 EMMERTON, J., & DELIUS, J. D. (1993). Beyond sensation: Visual cognition in pigeons. In H. P. Zeigler & H.-J. Bischof (Eds.), *Vision, brain, and behavior in birds* (pp. 377-390). Cambridge, MA: MIT Press.
 EMMERTON, J., LOHMANN, A., & NIEMANN, J. (1997). Pigeons' serial ordering of numerosity with visual arrays. *Animal Learning & Behavior*, **25**, 234-244.
 FERNANDES, D. M., & CHURCH, R. M. (1982). Discrimination of the number of sequential events by rats. *Animal Learning & Behavior*, **10**, 171-176.
 FETTERMAN, J. G. (1993). Numerosity discrimination: Both time and number matter. *Journal of Experimental Psychology: Animal Behavior Processes*, **19**, 149-164.
 FRIEDMAN, M. B. (1975). How birds use their eyes. In P. Wright, P. G. Caryl, & D. M. Vowles (Eds.), *Neural and endocrine aspects of behavior in birds* (pp. 181-204). Amsterdam: Elsevier.
 GELLERMANN, L. W. (1933). Chance orders of alternating stimuli in visual discrimination experiments. *Journal of Genetic Psychology*, **42**, 206-208.
 GOODALE, M. A. (1983). Visually guided pecking in the pigeon (*Columba livia*). *Brain, Behavior & Evolution*, **22**, 22-41.
 GOODALE, M. A., & GRAVES, J. A. (1982). Retinal locus as a factor of interocular transfer in the pigeon. In D. I. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 211-240). Cambridge, MA: MIT Press.
 HONIG, W. K. (1992). Emergent properties of complex arrays. In W. K. Honig & J. G. Fetterman (Eds.), *Cognitive aspects of stimulus control* (pp. 301-321). Hillsdale, NJ: Erlbaum.
 HONIG, W. K. (1993). Numerosity as a dimension of stimulus control. In S. T. Boyesen & E. J. Capaldi (Eds.), *The development of numerical competence: Animal and human models* (pp. 61-86). Hillsdale, NJ: Erlbaum.
 HONIG, W. K., & STEWART, K. E. (1989). Discrimination of relative numerosity by pigeons. *Animal Learning & Behavior*, **17**, 134-146.
 HONIG, W. K., & STEWART, K. E. (1993). Relative numerosity as a dimension of stimulus control: The peak shift. *Animal Learning & Behavior*, **21**, 346-354.
 JENKINS, H. M., & SAINSBURY, R. S. (1970). Discrimination learning with the distinctive feature on positive or negative trials. In D. I. Mostofsky (Ed.), *Attention: Contemporary theory and analysis* (pp. 239-273). New York: Appleton-Century-Crofts.
 KNORN, H.-J. (1987). *Visuelle Diskriminationsleistung von Tauben und Menschen beim simultanen Punktmengenvergleich* [Visual discrimination performance of pigeons and humans in simultaneous comparisons of dot quantities]. Unpublished Diploma thesis, Ruhr-Universität Bochum.
 KOEHLER, O. (1943). "Zähl"-Versuche an einem Kolkraben und Vergleichsversuche an Menschen ["Counting" experiments with a raven and comparative experiments with humans]. *Zeitschrift für Tierpsychologie*, **5**, 575-712.
 KOEHLER, O., MÜLLER, O., & WACHHOLTZ, R. (1935). Kann die Taube Anzahlen erfassen? [Can the pigeon comprehend numbers?] *Verhandlungen der Deutschen Zoologischen Gesellschaft, Zoologische Anzeige*, **8**, 39-54.
 KOEHLER, O., & WACHHOLTZ, R. (1936). Weitere Versuche an der Taube Nichtweiß zur Frage des Zählvermögens [Further experiments with the pigeon "Nichtweiß" concerning the question of counting ability]. *Verhandlungen der Deutschen Zoologischen Gesellschaft*, **38**, 211-236.
 KRUEGER, L. E. (1972). Perceived numerosity. *Perception & Psychophysics*, **11**, 5-9.
 MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.

- MECHNER, F. (1958). Probability relations within response sequences under ratio reinforcement. *Journal of the Experimental Analysis of Behavior*, **1**, 109-122.
- MECK, W. H., & CHURCH, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, **9**, 320-334.
- MOKRE, H. (1928). Über den Einfluß von Größe und Abstand der Elemente auf die Mengenauffassung [Concerning the influence of size and spacing of elements on the perception of quantity]. *Zeitschrift für Psychologie*, **105**, 195-225.
- PEPPERBERG, I. M. (1994). Numerical competence in an African Gray parrot (*Psittacus erithacus*). *Journal of Comparative Psychology*, **108**, 36-44.
- PIAGET, J. (1952). *The child's conception of number*. London: Routledge & Kegan Paul.
- RILLING, M. (1993). Invisible counting animals: A history of contributions from comparative psychology, ethology, and learning theory. In S. T. Boysen & E. J. Capaldi (Eds.), *The development of numerical competence: Animal and human models* (pp. 3-37). Hillsdale, NJ: Erlbaum.
- ROBERTS, W. A., MACUDA, T., & BRODBECK, D. R. (1995). Memory for number of light flashes in the pigeon. *Animal Learning & Behavior*, **23**, 182-188.
- ROBERTS, W. A., & MITCHELL, S. (1994). Can a pigeon simultaneously process temporal and numerical information? *Journal of Experimental Psychology: Animal Behavior Processes*, **20**, 66-78.
- SIEGEL, S., & CASTELLAN, N. J., JR. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.
- THOMAS, R. K., FOWLKES, D., & VICKERY, J. D. (1980). Conceptual numerosness judgments by squirrel monkeys. *American Journal of Psychology*, **93**, 247-257.
- VOS, P. G., VAN OEFFELEN, M. P., TIBOSCH, H. J., & ALLIK, J. (1988). Interactions between area and numerosity. *Psychological Research*, **50**, 148-154.
- WRIGHT, A. A. (1990). Markov choice processes in simultaneous matching-to-sample at different levels of discriminability. *Animal Learning & Behavior*, **18**, 277-286.
- WRIGHT, A. A. (1992). Learning mechanisms in matching to sample. *Journal of Experimental Psychology: Animal Behavior Processes*, **18**, 67-79.
- WRIGHT, A. A., & SANDS, S. F. (1981). A model of detection and decision processes during matching to sample by pigeons: Performance with 88 different wavelengths in delayed and simultaneous matching tasks. *Journal of Experimental Psychology: Animal Behavior Processes*, **7**, 191-216.
- XIA, L., WYNNE, C. D. L., VON MÜNCHOW-POHL, F., & DELIUS, J. D. (1991). PSYCHOBASIC: A BASIC dialect for the control of psychological experiments with the Commodore-64 and DELA interfacing. *Behavior Research Methods, Instruments, & Computers*, **23**, 72-76.

(Manuscript received October 7, 1996;
revision accepted for publication February 26, 1998.)