

Scalar effects in the visual discrimination of numerosity by pigeons

JACKY EMMERTON and JENNIFER C. RENNER
Purdue University, West Lafayette, Indiana

Pigeons trained in a conditional discrimination procedure to respond to a visual array made a left or right choice, depending on which of two numbers of elements (i.e., anchor numerosities) the array contained. They were then tested with novel arrays at these anchor numerosities, as well as at interpolated and extrapolated numerosities. Various control conditions showed that the birds' discrimination performance was primarily based on stimulus numerosity, and not on other factors, such as brightness or area. Results from a series of tests, spanning a wide range of numerosities, conformed to scalar principles. Psychometric functions showed superposition, indicating that Weber's law applies to numerosity discrimination. The subjective midpoint between anchor values was at the geometric mean. Variability about this bisection point increased in proportion to the numerical value of the mean.

It is becoming increasingly apparent that animals and humans share some basic similarities in their abilities to assess and discriminate the number of items they perceive or the number of events that occur in their environment (Dehaene, 1997; Gallistel, 1990; Gallistel & Gelman, 1992). When animals or humans discriminate stimuli that differ in their numerosities, they show numerical distance and size effects (see Dehaene, Dehaene-Lambertz, & Cohen, 1998). The numerical distance effect is demonstrated when, for a given numerosity, discrimination becomes easier as the difference between this numerosity and another one increases. So, for example, it is easier to discriminate 5 items from 10 items than it is to discriminate 5 items from 6. One way of demonstrating the numerical size effect is to fix the numerical difference, but vary the numerosities themselves. For the same numerical distance, discrimination is harder when the numbers of to-be-discriminated items or events are increased. Thus, for a difference of 2, discrimination performance is better for 2 versus 4 things than it is for 7 versus 9. These distance and size effects in numerical discrimination indicate that Weber's law applies to the discrimination of number. In general, Weber's law tells us that if stimulus numerosity increases, then so must the difference between this stimulus and a comparison, for discrimination performance to be maintained at the same level.

However, the similarities between animals and humans in these simple performance effects point to a more fundamental level of continuity across species. It is now

thought (see, e.g., Dehaene et al., 1998; Gallistel & Gelman, 2000) that the foundations for all numerical abilities occur at a nonverbal level, even in adult humans. According to this view, numerosities (the objective number of stimuli, or integer quantities) are internally represented in an analog fashion by magnitudes that are ordered as if on a continuum. These magnitudes have been posited as the activity of units in a neural network (Dehaene & Changeux, 1993) or as representations of numerical quantities that are stored in memory (Gallistel & Gelman, 2000). Conceptually, the mean activity of each neural unit or memory representation is mapped to a given numerosity, so there is an internal signal corresponding to a particular numerosity. But the signal is said to be noisy, and can be thought of as a symmetrical function having a mean value corresponding to stimulus numerosity, and a distribution of noise.

In many respects, these models of nonverbal processing of number resemble an earlier mathematical model derived by Gibbon (1977, 1981, 1986) to account for timing ability in animals. Similarities in the processing of time and number that might even rely on the same underlying mechanisms were suggested by the results of an experiment by Meck and Church (1983). They demonstrated a coincidence in the psychometric data obtained when rats discriminated either time or number with stimuli consisting of on-off cycles of noise. Pigeons also discriminate both the duration and numerosity of events consisting of either external stimuli (Roberts & Mitchell, 1994) or their own emitted responses (Fetterman, 1993). Although recent studies (see Roberts, 2002) call into question the generality of identical effects with time and number of the type reported by Meck and Church, versions of Gibbon's model have been applied to numerosity discrimination, as well as to timing ability.

An important component of Gibbon's model is the proposition of scalar variability. When it is applied to numerical ability, the mean of each signal function is pro-

J.E. dedicates this article, with gratitude, to Professor Emeritus Juan D. Delius, in honor of his 70th year. We thank Terry Powley for permission to use the NeuroLucida system in his lab, and Bob Phillips for showing us how. We are also grateful to Peter Urcuioli for his comments on an earlier version of this article. Correspondence concerning this article should be addressed to J. Emmerton, Department of Psychological Sciences, Purdue University, 703 Third Street, West Lafayette, IN 47907-2081 (e-mail: jemmert@psych.purdue.edu).

portional to stimulus numerosity, and the variability of each function also increases in proportion to its mean. In discriminating among numerosities, a subject would have to judge whether the noisy signal representing one numerosity value was subjectively different from another signal representing a different numerosity. If the noise of each signal is scaled to its mean value, then as numerosities increase in size, so must the distance between them, for these numerosities to be equally discriminable. This numerical size–distance effect means that scalar variability also concurs with Weber’s law.

Although Dehaene and Gallistel and their colleagues agree that discrete quantities are internally represented as magnitudes, they disagree about the scaling of these magnitudes on a subjective continuum or “number line.” Dehaene (1997; Dehaene & Changeux, 1993; Dehaene & Mehler, 1992) argues that the representation of numerosity on the “number line” is logarithmic, with subjective compression of the scale as numerosity increases. Gallistel and Gelman (1992, 2000) argue, instead, for a linear representation of number. But, as Gibbon (1981) showed mathematically, and Dehaene and Changeux pointed out, the same performance effects can be achieved by assuming linear scaling of numerosity, with scalar variability in the internal representation of number at increasing magnitudes, or logarithmic scaling of numerosity, with constant log variability in the activity of processing units. So computationally, these two formulations are equivalent, as we shall discuss later.

Gibbon’s (1981) mathematical model—as well as Gallistel and Gelman’s (1992, 2000) model of memory variability—leads to several predictions about numerosity discrimination. If animals are trained to discriminate two stimulus quantities, and then are required to judge whether intervening stimulus values are more similar to the larger or smaller training value, the psychometric functions describing their distribution of “large” versus “small” choices should superimpose under certain conditions. If the numerosities to be discriminated are in the same ratios, these functions should coincide, even if the absolute stimulus values differ. This superposition effect follows from Weber’s law. Furthermore, if each psychometric function is normalized about the subjective mean of discrimination (the bisection point between training values, or the point of subjective equality, PSE), then these functions should superimpose for different numerosity ratios also (Gibbon, 1986). This form of superposition effect has been demonstrated in humans’ discrimination of tone durations in experiments on timing ability (Allan & Gibbon, 1991).

Another prediction derived from models of scalar variability concerns the value of the subjective mean in bisection tasks of the type just described. Gibbon argued that if scalar principles apply, and if the relative similarity of a novel stimulus to each training stimulus is judged by comparing the ratios of the test stimulus with each of these anchor values, then the subjective midpoint between the learned stimulus values should fall at the geometric mean.

In this case, the ratios between this midpoint and each of the training values are equal.

Scalar models predict that variability in judgments will increase as mean judgments of numerosity increase. If the absolute values of training numerosities increase, then obviously the subjective mean (PSE) of these values must increase accordingly. One way of assessing variability about the mean is to determine the just noticeable difference (or difference limen) between the PSE and other numerosities. So, for tests of numerosity discrimination across a range of stimulus values, there should be an increase in the size of the just noticeable difference that is proportional to the numerosity at the PSE between training values.

Fetterman (1993) demonstrated several of these effects when he trained pigeons in a bisection task. In that task, the birds had to discriminate sequences of pecking responses that they emitted. Fetterman showed that they relied on both the number of pecks they made and the duration of each series of pecks. The present experiments also tested for scalar properties of numerosity discrimination, and were modeled on some aspects of Experiment 1 in Fetterman’s study. However, instead of having to keep track of series of their own responses, our pigeons had to assess the numbers of items in visual arrays. Since all the items in a stimulus array appeared simultaneously, time was not a factor in our experiments. But with visual stimuli, there are other factors that often covary with numerosity. These can include the summed area of the items, the overall brightness of the stimulus display, and the spacing of items. Our strategy was either to vary these common confounds of numerosity or to specifically control them.

However, when humans or animals make judgments about the number of things they see in the natural environment, numerosity rarely occurs as a “pure” stimulus factor. For instance, say we judge that there are two oranges in one bowl and six in another. The overall visual area of objects that we perceive, or the total stimulation by the color orange, is usually greater for six fruits than it is for two. Nevertheless, we say that our judgment is based on number. So the question throughout these experiments was whether stimulus numerosity was the primary factor that influenced pigeons’ judgments. Animals can base their decisions on multiple stimulus parameters (see, e.g., Chatlosh & Wasserman, 1993; Fetterman, Stubbs, & MacEwen, 1992), but if the number of items is the most important discriminative cue for the birds, then other stimulus properties that may covary with number should not be the sole or the major basis for their discrimination performance.

So we designed a series of experiments with the primary aim of finding out whether or not numerosity discrimination with visual stimuli conforms to Weber’s law and to other predictions of scalar theory. We tested birds’ discrimination across a range of large (Experiment 1) and small (Experiment 4) numerosities. In these experiments, we (1) looked for evidence of superposition effects, indicative of Weber’s law, (2) examined whether or not the point of subjective equality between training numerosi-

ties occurred at the geometric means of these values, and (3) examined whether or not variability in numerosity judgment increased in proportion to the value of the subjective midpoint, as predicted by scalar theory. Experiment 2 controlled for luminance and area of the stimulus elements themselves, to ensure that discrimination was not based solely on these cues. Experiment 3 examined the effect of stimulus spacing on discrimination performance, and tested whether the area filled by clusters of elements affected birds' numerosity judgments in the same way as it does humans'.

EXPERIMENT 1

In Experiment 1, pigeons were trained to discriminate arrays consisting of different numbers of items. The two numerosities that were used in training served as anchor values. The birds were then tested with numerosities that were distributed between the training values, as well as above and below those values. One purpose of these tests was to see if pigeons would show both interpolation and extrapolation in their judgments about novel numerosities. Interpolation has been shown in rats (see, e.g., Meck & Church, 1983) and in pigeons (see, e.g., Fetterman, 1993) when they were tested for their ability to judge novel numerosities interspersed between the training values. In those experiments, discrimination was based on series of events (auditory stimuli or the animals' own responses, respectively). Pigeons' ability to interpolate has also been demonstrated with visual arrays in which the number of items was fairly small (Emmert, Lohmann, & Niemann, 1997). Extrapolation to novel numerosities has been demonstrated in macaque monkeys when they were tested with visual displays (Brannon & Terrace, 2000) but has not yet been demonstrated in birds.

The numerosity values used in the various phases of this experiment were based on those used by Fetterman (1993) so that we could compare our data with his. Another aim of our experiment was to see whether, as in Fetterman's study, the birds' test results yielded similar scaling effects for a particular ratio of numerosity values, even though the absolute numerosities varied across different phases of the experiment.

Method

Subjects

Three adult pigeons (*Columba livia*) of mixed breed were used. They were food deprived and maintained at about 85% of their free-feeding weight. The birds were housed individually in a colony room that was lit on a 14:10-h light:dark cycle. They had unrestricted access to water and health grit in their home cages. The birds were experimentally naive at the start of the study.

Materials

Apparatus. The operant conditioning chamber measured $33 \times 34.5 \times 32$ cm. An opening (26.5×15 cm) on one side gave access to a VGA monitor (Zenith 1492 FTM, flat screen), on which the stimuli were presented. The monitor was equipped with an infrared scanning touch screen (CarrollTouch) that sensed a bird's responses. A 1-mm-thick clear plastic sheet protected the computer screen and cushioned the impact of a bird's pecks. The chamber was equipped

with two grain feeders (Coulbourn Instruments, Model E14-10). There was an opening (3.5×5.5 cm) to each feeder on the walls to the left and to the right of the side facing the monitor. Each feeder could be illuminated separately, and independently of food delivery. An infrared lightgate, added to each unit, sensed when a bird inserted its head in the opening. A dim (1.2-W) houselight was located toward the rear of the box, and was shielded to minimize light reflection on the computer screen. A PC controlled experimental events via an interface. Visual stimuli were generated using Borland Turbo Pascal.

Stimulus arrays. Arrays consisted of small white rectangles against a dark background. Within an array, these elements were all of the same size, but three different sizes (small, medium, and large) were used across different arrays. The sizes of these items were programmed to be nominally 2×2 , 3×3 , or 4×4 pixels, but since the arrays had to be presented in EGA mode, the array elements appeared as rectangles. Based on calibrations of blocks of these rectangles, the three on-screen sizes of the individual elements were calculated to be 1.19×1.55 mm, 1.54×2.06 mm, and 1.96×2.54 mm, respectively. Coding files specified the number of items to appear in a given array, and the size of the rectangles. The locations of the items were randomly generated from an irregular matrix of possible locations, so the arrays varied across trials to produce configurations that were effectively unique for each trial. Different matrices, consisting of 30, 60, or 110 locations, were used, depending on the range of stimulus numerosities that were required within a particular series of training and test sessions.

Procedure

Pretraining. The birds were magazine trained. They each completed 9 or 10 sessions, with 48 trials per session, of modified auto-shaping. The conditioned stimulus was a white circle filled with a crosshatch pattern, which was presented in the middle of the screen for 6 sec. If a bird pecked this stimulus, or the presentation period expired without a peck, food access followed. The hopper was activated, and the hopper recess lit up for a maximum of 20 sec, to allow the bird adequate time to approach the food location. The effective reinforcement interval (feeding time) of 3 sec was initiated when the bird inserted its head into the recess, and interrupted an infrared lightgate. After this 3 sec expired, the hopper and hopper light were switched off. (All reward times that are described subsequently refer to this effective feeding time.) Either the right or left food hopper was activated randomly across trials. The intertrial interval (ITI) was 15 sec.

The birds then underwent instrumental pretraining. The same stimulus was used in these sessions. From trial to trial, it was randomly presented on the left or right side of the screen. A single peck switched off the stimulus and activated the food hopper on the corresponding side. The reward time was 3 sec, the ITI was 10 sec, and there were 48 trials per session. The birds completed 6 to 10 sessions to ensure that they responded fairly promptly when the trials began.

Training. The basic procedure was a conditional discrimination procedure. Each trial began with the presentation of a numerosity array in the center of the screen. The birds were trained to complete a fixed ratio of five pecks (FR5) to the array. With the array still visible, a yellow and a blue choice patch (25-mm diameter; separated by 140 mm) appeared to the left and right of the array. A single peck to one of these patches extinguished the stimuli and led to food reward (1.5–1.8 sec) if the array consisted of the larger training numerosity. Responding to the other patch was rewarded in conjunction with the smaller numerosity value. Choices of the incorrect patch also extinguished the stimuli but led to a 5-sec time-out. These incorrect choices initiated a correction procedure in which the trial was repeated (with the identical stimulus array) until the bird made a correct choice. Correction trials were recorded but not used in the calculation of discrimination scores. Which color was the correct one to choose for the larger as opposed to the smaller

number of items varied across birds. Trials were separated by an ITI of 15 sec. The order of presentation across trials of the larger versus smaller training numerosities was quasi-random (Fellows, 1967). There were 48 noncorrection trials per session. Sessions were run daily, 5 days a week.

For each noncorrection trial, the array configuration was randomly generated, as described above. Other stimulus parameters were counterbalanced within each session. The smaller numerosity was presented on 24 trials, and the larger numerosity on the other 24 trials. For each numerosity value, there were 8 trials at each of the three sizes of array elements.

Training continued like this, with each trial terminating in reward or time-out, until a bird discriminated at a level of $\geq 80\%$ correct choices within a session. Then, in preparation for testing the birds on extinction trials, partial reinforcement was gradually introduced. If performance was maintained at a level of $\geq 80\%$ correct, the probability of a trial ending in reward or time-out decreased in the next session by .1, until the probability of reinforcement had reached .5. After this, the test sessions for that pair of anchor values began.

The pairs of anchor values presented in the various training phases of Experiment 1 were as follows: 5 versus 10, 10 versus 20, 20 versus 40, 5 versus 15, 10 versus 30, 20 versus 60, 5 versus 20, 10 versus 40, and 20 versus 80. Thus, at each of three numerosity ratios (1:2, 1:3, and 1:4) there were three pairs of values that differed in their absolute numerosities. The order in which the birds were trained (and then tested) with these different anchor values is shown in Table 1.

Test sessions. For each pair of anchor values, there were five test sessions, each consisting of 96 trials. Forty-eight trials were reinforced trials, with 24 presentations each of the larger and smaller numerosity values used in training. For each of these values, the array elements were either small-, medium-, or large-sized on 8 trials. Reward and time-out parameters were the same as in training sessions, but there was no correction procedure after incorrect choices. The other 48 trials were nonreinforced test trials. On these, choice of one of the colored patches extinguished the stimulus display and initiated the ITI. There were 6 test trials at each of eight numerosity values, with three array-element sizes used at each value. Two test values were the large and small anchor numerosities used in training. Four values were numerosities that were evenly distributed between the anchor values (interpolated numerosities). One numerosity was greater than the large anchor value, and one was less than the small anchor value (extrapolated numerosities). The arithmetic interval between each test numerosity value was given by the difference between the anchor values divided by 5. The order of presentation of training trials and the various test numerosities was quasi-random in each test session.

Results

Training

For each of the training phases, Table 1 shows the number of sessions (including those with partial reinforcement) that each bird completed. Although Bird 196 required many sessions to acquire the discrimination of the first pair of anchor values, all birds showed rapid transfer to subsequent stimulus numerosities.

Test Data

Throughout the test phases, discrimination scores on the reinforced trials were maintained at a high level. The mean percent correct score, across all birds and sessions, was 92.2% ($SD = 6.8$). Scores on test trials were expressed as the proportion of choices made to the color patch that designated the larger of the anchor values for that series of test sessions. For each series, the test scores were analyzed with three-way repeated measures ANOVAs, with numerosity value, array-element size, and sessions as the main factors. Numerosity had a highly significant effect ($p < .00001$) on choices in all tests, with $F(7,14) = 36.4$ (5 vs. 10); 39.2 (10 vs. 20); 41.9 (20 vs. 40); 78.3 (5 vs. 15); 74.9 (10 vs. 30); 360.7 (20 vs. 60); 331.5 (5 vs. 20); 170.2 (10 vs. 40); and 157.8 (20 vs. 80), where the numbers in parentheses indicate the anchor values for each test series. There was no significant effect ($p > .05$) of sessions. Element size yielded a significant effect only in the following tests: $F(2,4) = 13.2, p < .05$ (5 vs. 15); 8.4, $p < .05$ (10 vs. 30); 31.9, $p < .01$ (20 vs. 60); 18.1, $p < .05$ (10 vs. 40). Scheffé post hoc tests showed significant differences between scores with small and large elements on all these tests ($p < .05$ for 5 vs. 15, 10 vs. 30, and 10 vs. 40; $p < .01$ for 20 vs. 60 data). There were also significant differences ($p < .05$) for scores with small- versus medium-sized elements in the 20 versus 60 test, and with medium versus large items in the 10 versus 40 test. So in general, for those tests in which element size influenced performance, birds were slightly more likely, on average, to choose the patch for “large” value as element size increased from small to large.

Table 1
Order of Training and Number of Training Sessions for Each Bird in Experiment 1

Order of Training	Bird 196		Bird 197		Bird 198	
	Anchor Values	No. of Sessions	Anchor Values	No. of Sessions	Anchor Values	No. of Sessions
1	5 vs. 10	51	10 vs. 30	28	10 vs. 30	28
2	20 vs. 80	7	20 vs. 80	6	20 vs. 80	6
3	5 vs. 15	6	10 vs. 20	7	10 vs. 20	7
4	10 vs. 40	8	5 vs. 10	10	5 vs. 10	10
5	10 vs. 20	6	10 vs. 40	6	10 vs. 40	6
6	20 vs. 40	6	20 vs. 60	6	5 vs. 15	6
7	10 vs. 30	6	5 vs. 15	7	20 vs. 40	8
8	5 vs. 20	8	5 vs. 20	6	5 vs. 20	6
9	20 vs. 60	6	20 vs. 40	7	20 vs. 60	6

Note—Number of training sessions includes a minimum of five that were conducted with partial reinforcement, after each bird reached a criterion performance of $\geq 80\%$ correct.

For each set of test sessions, the data were averaged across the 3 birds, and then these means were plotted to give a psychometric function. The proportion of “large numerosity” choices the birds made was plotted on the y -axis, and the ratios of a given test numerosity (N) relative to the value of the smaller anchor numerosity (S), on the x -axis. This allowed us to compare functions within a ratio category of anchor values when absolute numerosities differed, and also to compare our results with Fetterman’s. For each psychometric plot, a logistic distribution function was fitted to the data with an iterative least squares estimation procedure. [When plotted on the linear coordinates used here, the logistic function approximates the Gaussian function as a cumulative function that may be fitted to ogivally distributed psychometric data (Macmillan & Creelman, 1991). For practical purposes, the two functions are regarded as extremely similar (Luce, 1959). Although the cumulative

Gaussian function has more traditionally been used for fitting psychometric data, the logistic function is mathematically simpler to apply.] These fitted functions accounted for over 99% of the variance in the data, on average (ranging from 95.3%–99.9%). These functions, grouped according to anchor value ratios, are shown in Figure 1.

Further analyses that examined these data for scalar effects will be presented later in this article.

Discussion

Across all the tests, covering a large range of stimulus quantities, the birds’ performance was strongly dependent on the numbers of items in the arrays. After each training phase with two anchor values, the birds’ test performance showed that they could generalize their numerosity discrimination to novel values. They showed clear evidence of interpolating the four numerosities that were between the

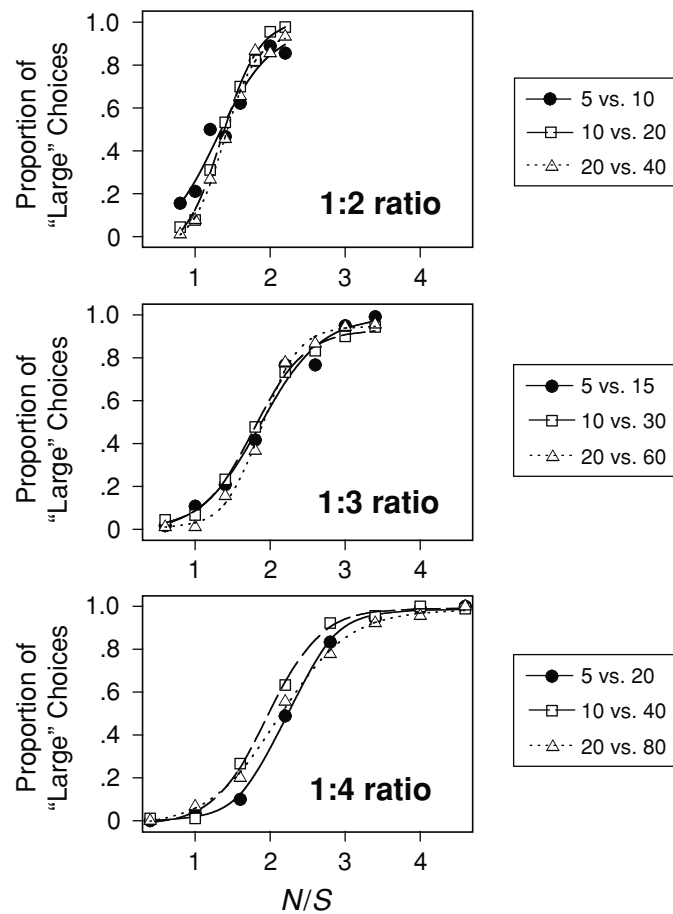


Figure 1. Psychometric data from test trials, averaged across birds and sessions, in Experiment 1. Data are plotted as the proportion of choices made to the color patch designated correct for the “large” training numerosity. N/S is a relative scale, where N is the stimulus numerosity for a given test trial, and S is the value of the “small” training numerosity. A cumulative logistic function is fitted to each set of test data. The results are displayed in groups, according to the ratios of the training values used with each test phase. These training values are shown in the legend beside each graph.

anchor values with which they were trained. These interpolation results thus extend the earlier findings with small numerosities (Emmerton et al., 1997). There was also evidence of their ability to extrapolate to numerosities beyond the training values. In Figure 1, the upper and lower end points of each plot show discrimination scores at these extrapolation values. On all but one of the tests, the birds were either more likely or—due to a ceiling effect—as likely to choose “large” when the test numerosity exceeded the larger of the anchor values. Similarly, they were more likely or as likely to choose “small” with arrays containing fewer items than the smaller anchor value. So birds, like primates (Brannon & Terrace, 2000), extend their numerosity judgments to values that are ordered outside the anchor points defined by the training numerosities.

One of the predictions derived from Weber’s law and scalar theory is a superposition effect. So at the very least, psychometric functions should superimpose when the ratio of their anchor values is the same, although absolute numerosity values differ. This form of superposition effect was obtained (Figure 1). These results are similar to those reported by Fetterman (1993).

The effect on the birds’ choices of the sizes of array elements was inconsistent across tests. On five of the tests, element size had no reliable effect on performance. On four of the tests, the element size had some influence on the birds’ choices, as indicated by the significant effect of this factor in the ANOVAs. But even on these tests, the effect of element size was much less than that of stimulus numerosity. Nevertheless, the average tendency was for the birds to choose “large” slightly more often with larger item size.

With homogeneous arrays, the summed area of whiteness (and, correlated with this, the overall luminance of the array) will increase as element size increases, for a fixed number of elements. Conversely, for a fixed element size, the area and brightness of an array will increase as numerosity increases. However, initial calculations of summed areas at different numerosities indicated that the birds’ choices must still have been based predominantly on the number of items. For instance, in tests for the anchor values 5 versus 15, the summed area of 13 small elements was virtually the same as the area of 5 large ones (24.1 vs. 24.8 mm², respectively), although the proportion of “large” choices differed. In the 20 versus 60 tests, the area of 68 small elements (125.8 mm²) was less than that of 28 large elements (138.9 mm²). But choice of “large” was related to greater numerosity rather than greater area. Similar examples could be given for all the other tests. So it was not really larger area (or greater luminance) that primarily determined the distribution of the birds’ choices that are plotted in Figure 1. Instead, it was the number of elements in the arrays. Nevertheless, to support this claim, we conducted further tests to see if the birds still responded to numerosity when other factors were controlled.

EXPERIMENT 2

In order to make sure that the birds’ choices in these tests were indeed primarily dependent on variations in numerosity,

rather than on differences in overall stimulus brightness or area, we created two new sets of stimulus arrays. In one condition, arrays were matched, across numerosity values, for their overall luminance, but the summed areas of elements could still vary across arrays. In the second condition, all the arrays were equated for the summed areas of their elements, and, as a corollary, were equal in brightness too. These control tests were conducted with numerosities that went with the 10 versus 20 anchor values, because, for these stimulus values, brightness and area matches could be achieved across all the numerosities by creating combinations of the three element sizes that had been used in Experiment 1. If variations in brightness and/or summed area were not essential to the birds, they should still be able to discriminate arrays according to their numerosity values. If, instead, the birds had been basing their discrimination mainly on these other cues, rather than on the number of elements in a stimulus, their discrimination ability should collapse, and they should perform at chance level.

Method

Subjects

Chronologically, this experiment was conducted at the end of this whole study. For the first part of the experiment (equal-brightness control), the subjects were the same 3 birds that had served in the first experiment. After the first part of Experiment 2 was completed, 1 of the birds died. Another bird, which had previous conditioning experience in the lab, was substituted. This bird and the remaining 2 pigeons were used in the equal-area control phase.

Materials

Apparatus. The apparatus was the same as in Experiment 1.

Stimulus arrays. Stimulus brightness was calibrated by measuring the overall luminance of arrays consisting of a cluster of 1–16 elements. The measurements were repeated for each of three sizes of elements (small, medium, and large, as defined in Experiment 1), and for each of two colors (white or light gray, as defined by Turbo Pascal’s color palette). A Mavolux electronic luxmeter was used to measure luminance in cd/m². The measurements were plotted across numerosity for each element size and color combination, and linear regression functions were fitted to these six sets of data. The linear functions were then used to calculate the luminance of the arrays for the first phase of this experiment.

In order to control the parameters of each array, coding files specified the fixed locations of array elements, the size of each element, and the color of each element. Arrays used in the first part of the experiment were equated in luminance to a value of 5.01 cd/m² (0.7 log cd/m²). The difference in luminance of each array from this standard value was less than or equal to 0.1 log cd/m², which is the pigeons’ threshold for brightness discrimination with successively presented stimuli (Hodos, Bessette, Macko, & Weiss, 1985). Arrays were matched for brightness by choosing combinations of element sizes in either white or gray. Although summed areas of elements increased, on average, across numerosities (34.2–68.8 mm²), areas of individual arrays varied both within and across numerosity values. The configurations of elements varied across all arrays.

Stimulus files were also prepared for the arrays used in the second part of this experiment. The files encoded the fixed locations of elements that were all white. Different element sizes were combined in order to equate arrays, across numerosities, to be within 0.8 mm² of 40 mm². As a corollary, luminance was also effectively matched across arrays, ranging from 5.77–6.77 cd/m² (or 0.76–0.83 log cd/m², so that the brightness difference was below pigeons’ discrimination threshold).

Procedure

Equal-brightness control. Procedurally, training sessions were conducted in the same way as for Experiment 1. Arrays consisting of 10 items were presented, in quasi-random order, on half of the 48 trials per session. On the other trials, arrays of 20 items were shown. On each trial, the bird pecked on an FR5 to the stimulus array, and then had to choose the blue or yellow color patch, depending on the stimulus numerosity. In initial training sessions, until an individual bird's discrimination score was $\geq 80\%$, each trial terminated with food reward or with a time-out followed by a correction procedure. Partial reinforcement, down to a reinforcement probability of .5, was introduced over subsequent training sessions. Testing then began. There were five sessions, consisting of 96 trials each. Forty-eight trials, on which the anchor stimulus arrays were presented, terminated in either reward or a time-out (but no correction procedure followed). Intermixed with these trials were the nonreinforced test trials, in which novel arrays were introduced. Interpolated and extrapolated numerosities on these test trials were the same as for the 10 versus 20 anchor values in Experiment 1, so there were 6 test trials at each of eight numerosities. In both training and testing, a particular stimulus array was presented only once within a session, but the same array was used across sessions. Trial order was changed daily in the test sessions.

Equal-area control. In this phase, the birds underwent training sessions, as for the equal-area control phase, but with 48 stimulus arrays that were equated for summed area. The same arrays were used on reinforced trials in subsequent test sessions. Six different arrays at each of the eight test numerosities were presented on nonreinforced test trials. One of the birds that had been used in Experiment 1 became ill after completing only three of the five test sessions. The other 2 birds successfully completed all five sessions.

Results and Discussion

The mean results for the equal-luminance tests, fitted with a logistic distribution function, are shown in Figure 2A. The birds' choices were still related to stimulus numerosity when arrays were equated for their overall luminance but summed area varied.

Mean results for the equal-area tests are shown in Figure 2B. Since 1 bird finished only 3 out of 5 sessions, average scores were taken from the total of 13 sessions that were completed by the 3 birds used for these tests. The results again indicate that stimulus numerosity was the most likely factor underlying the birds' choices in Experiment 1. Their ability to discriminate the various arrays certainly did not depend on differences in stimulus area (or correlated differences in brightness).

EXPERIMENT 3

When human subjects have to estimate the numerosity of visual displays containing large numbers of items, the spatial distribution of these items can influence numerosity judgments in a variety of ways (see Allik & Tuulmets, 1991, for a review). An example of the effect of spatial distribution on these judgments is the "regular-random illusion" reported by Ginsburg (1976). He found that human observers estimated a display of evenly spread-out dots to be more numerous than a random distribution of the same number of dots. Others (see, e.g., Krueger, 1972; Vos, van Oeffelen, Tibosch, & Allik, 1988) have reported that random configurations of dots are judged to be more numerous than displays consisting of clumps

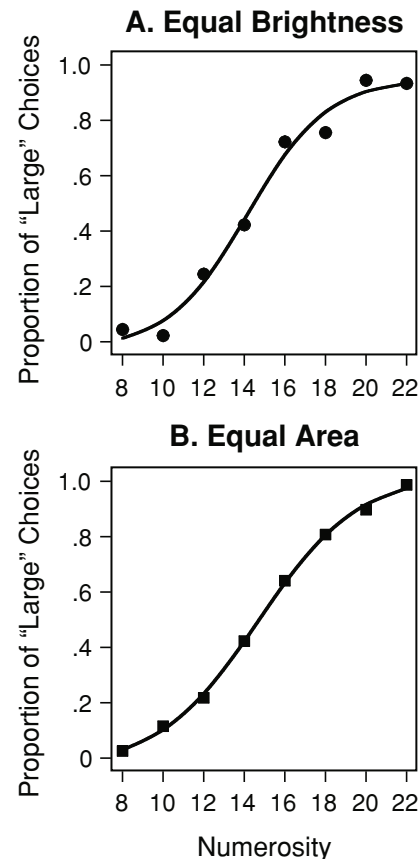


Figure 2. Test results from Experiment 2, fitted by logistic functions. (A) Stimulus arrays were equated for luminance across numerosities, but the summed areas of array elements varied. (B) Arrays were equated for the summed areas of their component elements, and were also of equal luminance.

of dots, when the total quantities are actually the same. Vos et al. (1988) attributed these types of effects to the perceptual grouping of items within an array to form "filled" areas. They have modeled their own results, as well as the "regular-random illusion," with an algorithm called CODE (COntour DETector), developed earlier by van Oeffelen and Vos (1983) to show how dots group to form perceptual gestalts.

Basically, the algorithm relies on the relative proximity of array items to determine which ones group, and then calculates the extent of the clusters, or "filled" areas, within an array. When this algorithm is applied to judgments of numerosity, Vos et al. (1988) have proposed that humans sum the perceived clusters. If the summed area of these clusters is larger in one array configuration than in another one, then human observers are biased to judge numerosity to be greater in the array with the larger cluster area.

In an experiment on the effects of array density with small numbers of dots (Emmerton, 1998), pigeons showed effects that tended to be the opposite of those found with humans. Although the birds' discrimination performance

was primarily dependent on the relative numerosities of dots in the stimulus arrays, the results indicated that the spacing of dots biased this performance in such a way that numerosity was judged to be less when the dots were spread out than when they were more closely spaced. To our knowledge, there have been no other studies examining the effects of such spatial factors on animals' visual assessment of numerosity.

So in our third experiment, we created fixed arrays in which, for each numerosity value, there were three types of spatial configuration. In one type of array, the items were spaced close together to form what looked like a single clump. In another array type, there were two clumps, with the items within each clump closely spaced. This type of array was designed to see if pigeons' estimates of overall numerosity would be affected by spatial separation between groups of elements. In the third type of array, the items were more spread out. Their spacing was irregular, so that although the experimenters chose the item locations, the arrays resembled the "random" arrays that Ginsburg (1976) used, rather than his "regular" arrays. Since the parameters of the stimulus arrays were all specified in coding files, van Oeffelen and Vos's (1983) CODE algorithm could subsequently be applied to the stimuli. The summed "filled" areas that this cluster algorithm generates were then measured. So the aim of this experiment was to see if spatial distribution factors would influence the pigeons' numerosity judgments and, if so, whether any biases that emerged would conform to the effects reported in human subjects.

Method

Subjects

The same 3 pigeons that were used in Experiment 1 served as subjects.

Materials

Apparatus. The apparatus was the same as in the previous experiments.

Stimulus arrays. Coding files for the stimulus arrays specified the locations of elements. These locations were chosen by the experimenters to produce three types of array configurations (single clump, two clumps, and spread-out elements) on both training and test trials. All the elements were white, medium-sized rectangles.

Cluster calculations and measurements. The main assumption in van Oeffelen and Vos's (1983) CODE algorithm is that an array element does not have a discrete visual effect, but rather produces a distribution of activation, or a spread function, that can be represented by a binormal distribution. The standard deviation of this normally distributed activation function is determined by taking half the distance between that element and its nearest neighbor. These calculations are repeated for every element in the array, and then the overlapping patterns of activation among elements are summed. "Clusters" or "filled" areas are formed where summed activity exceeds a threshold level, set at the maximum activation produced by a single element. In our computations, we took into account the area of the array elements themselves, so that the smallest "filled" area corresponded to the area of an isolated, medium-sized rectangle.

For each of the test stimulus arrays, filled areas were computed and displayed on screen. These displays were captured as bitmap images that were scaled to the sizes of the actual stimulus presentations. The areas of all the clusters that were formed in these bitmap images were then measured using NeuroLucida software. This system is actually designed to compute the area encompassed by

an irregular polygon of the type formed by tracing the outline of a digitized image of a nerve cell, or some other histological structure. Examples of stimulus arrays and the clusters computed with the CODE algorithm are illustrated in Figure 3.

Procedure

Each bird was trained and tested at three sets of numerosity values, taking one set from each of the numerosity ratios 1:2, 1:3, and 1:4. The anchor values for the three sets were 20 versus 40, 5 versus 15, and 10 versus 40. For each pair of anchor values, training sessions were conducted in the same way as in the previous experiments. There were 48 trials per session. On half of these trials, the stimulus array had the smaller numerosity, and on the other 24 trials, it had the larger numerosity. At each of these numerosities, the three types of array configuration (single clump, two clumps, and spread-out elements) were presented on one third of the trials. Each bird was trained to a criterion of $\geq 80\%$ correct choices of the color patch for the larger or smaller stimulus numerosity. Reinforcement probability was then reduced to .5 before test sessions began.

In test sessions, 48 nonreinforced trials were intermixed with the 48 familiar, reinforced trials. For each pair of anchor values, the novel stimuli included eight numerosities, as in Experiment 1 (six novel stimulus arrays for the two anchor values, the four interpolated numerosity values, and the two extrapolated values). For each numerosity value, there were 2 trials with each of the three types of array configurations as stimuli. For each type of array, the specific locations of elements varied across stimuli. The order of reinforced and nonreinforced trials was quasi-random, and the order was changed daily across the five test sessions.

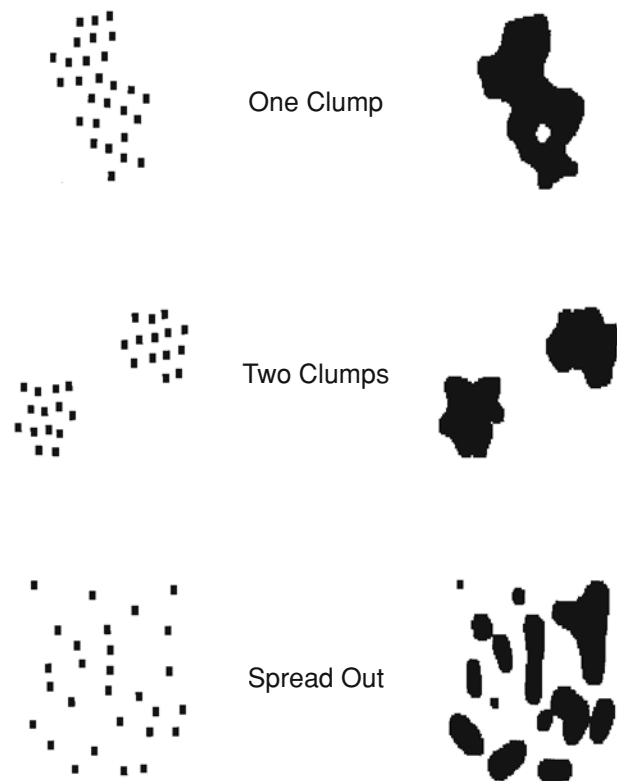


Figure 3. Examples of the three types of spatial configuration of stimulus arrays (left) used in Experiment 3, and the corresponding clusters of "filled" areas (right), computed according to the CODE algorithm. For illustration, arrays are shown in reversed contrast.

Results

Each set of test results was analyzed with three-way repeated measures ANOVAs, with numerosity, type of array configuration, and sessions as variables. The test scores with each type of array and for each numerosity, but averaged across sessions, are illustrated in Figure 4.

For the tests associated with anchor values 20 versus 40, there were significant main effects of numerosity [$F(7,14) = 54.5, p < .00001$] and of sessions [$F(4,8) = 4.2, p < .05$]. Mean scores varied across test sessions, but not in a consistent direction (i.e., there was no obvious extinction effect). The effect of array configuration approached but did not reach significance [$F(2,4) = 5.6, p = .07$]. But there was a greater tendency for birds to choose “large” when arrays were of the one-clump type as opposed to the other types.

For the 5 versus 15 test set, the only significant main effect was of numerosity [$F(7,14) = 36.7, p < .00001$]. The interaction of numerosity \times array configuration was also significant [$F(14,28) = 2.5, p < .05$]. As stimulus numerosity increased, the proportion of “large” choices

with spread-out arrays tended to diverge from and be lower than the scores with other types of arrays.

For the 10 versus 40 test set, numerosity produced a significant effect [$F(7,14) = 308.9, p < .00001$] and so did array configuration [$F(2,4) = 33.1, p < .01$], but not sessions. A Scheffé post hoc test on the variable of array configuration revealed significant differences in the scores for one-clump versus spread-out arrays ($p < .01$) and between two-clump and spread-out arrays ($p < .05$), but not between one-clump and two-clump arrays. The birds were more likely, on average, to choose “large” when array items were clumped together than when they were spread out. There were also several significant interactions in the ANOVA: numerosity \times array configuration [$F(14,28) = 7.3, p < .0001$], numerosity \times sessions [$F(28,56) = 2.1, p < .01$], and numerosity \times array configuration \times sessions [$F(56,112) = 1.6, p < .05$]. The probable reason for the significant numerosity \times array-type interaction was that test scores with different array types deviated from one another at the middle, interpolated stimulus numerosities. The other significant interactions are not easy to interpret, but seem to be attributable to various crossover effects in the data.

The scores for each type of array configuration were also plotted against the summed “filled” areas yielded by the CODE algorithm applied to each array. The measurements of these summed areas were averaged for each stimulus numerosity value, and plotted against the mean proportion of “large” choices the birds made. Because just one set of test results showed a significant main effect of array configuration on choices, only these results are illustrated in Figure 5, but the general pattern of results was similar for the other test sets. As the figure shows, the “filled” areas computed for arrays consisting of one or two clumps were very similar, whereas the summed areas of all the clusters for the spread-out arrays extended to higher values. The birds’ choices were not simply related to the filled areas computed for the stimuli; dissimilar behavioral scores were obtained even when area values were similar.

Discussion

The main factor influencing the birds’ choices was once more the numerosity value of the stimulus array. Array configuration appeared to have some influence on the birds’ choices, but the effect of this factor was not clear-cut, and was, in any case, subsidiary to the effect of numerosity. In only one out of the three tests did the spatial distribution of stimulus elements have a significant effect on the birds’ judgments, although in another test, this factor also approached significance. In the remaining test, although array configuration itself did not reliably affect performance, there was a significant interaction between numerosity and stimulus configuration. But as Figure 4 shows, any effects of spatial factors on the birds’ judgments of numerosity were in the opposite direction compared with those reported in humans. For the same number of items, birds were less likely to make a “large number” judgment when array elements were spread out

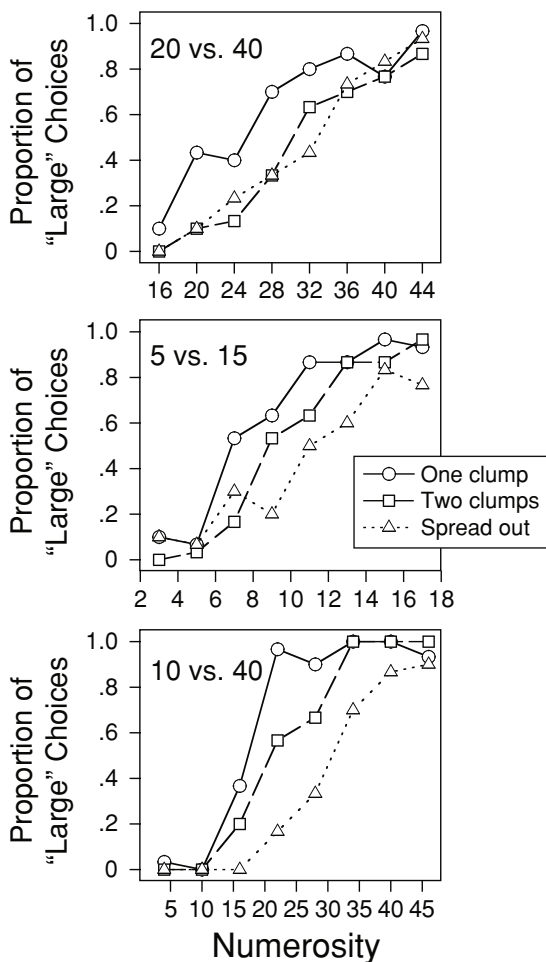


Figure 4. Test results from Experiment 3, plotted according to the type of stimulus array configuration.

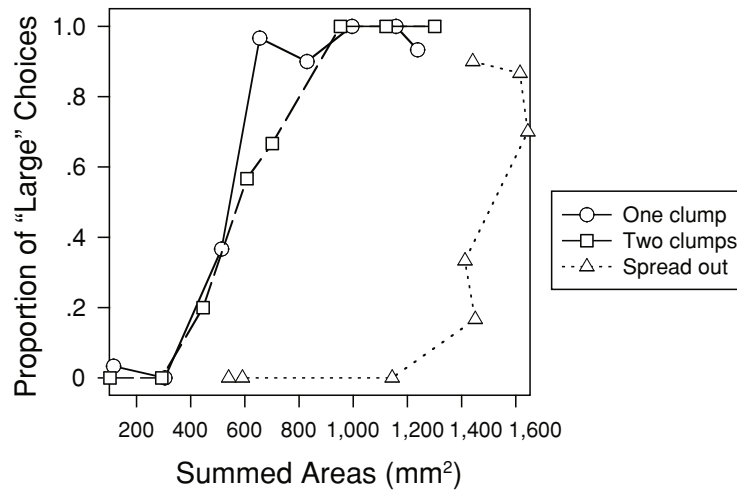


Figure 5. Results from Experiment 3 for the 10 versus 40 test, for which the effect of type of array configuration was significant. The x-axis shows the mean summed areas, measured from calibrated digital images of the clusters of “filled” areas computed from the three types of array used at each of the test numerosities. The leftmost data point on each curve is the score for the smallest test numerosity (extrapolated value below “small” training value). Subsequent data points on each curve are plotted in order of increasing test stimulus numerosity.

than when they were closely spaced. As indicated by Figure 5, birds did not seem to assess “filled” areas formed by these elements either. (Other measurements, not shown here, were made of the areas encompassed by the irregular outer perimeters of the arrays. These measurements simply showed that the overall areas of one- and two-clump arrays were similar, and that the areas of spread-out arrays were much larger. So the birds did not choose “large” on the basis of larger extent of the arrays, however that is measured.) The secondary effects on performance of manipulating the spacing and arrangement of array items more closely resembled earlier findings obtained when pigeons were tested with small numbers of dots (Emmerton, 1998) than the effects reported in humans’ estimation of large numbers of dots.

That is not to say that van Oeffelen and Vos’s CODE algorithm may not apply to humans’ estimation of numerosity. But humans and pigeons may perceive these types of stimulus arrays rather differently. First of all, the effective viewing conditions differ for human observers and pigeons. Humans view stimulus arrays at arm’s length, and the overall extent of the arrays is typically quite small. For instance, the dot displays used by van Oeffelen and Vos (1982) subtended a visual angle of 8° when viewed at a distance of 75 cm. Although the birds in our experiment were free to move around, pigeons preferentially view static stimuli binocularly, in their frontal visual field, when they are going to peck at them (Maldonado, Maturana, & Varela, 1988). In preparation for pecking, they fixate a stimulus from a distance of about 6 cm (Goodale, 1983; Macko & Hodos, 1985). At that distance, the largest arrays in this experiment, which extended over an area of 50 × 60 mm, at a maximum, would have subtended a

visual angle of about 45° × 53°. Although optically, the pigeon’s binocular visual field covers 80° × ~90°, when convergent eye movements are taken into account (Nalbach, Wolf-Oberhollenzer, & Remy, 1993), other factors most likely come into play to affect the way pigeons actually perceive or process information within these arrays.

Several studies have examined how humans, pigeons, and other species process visual stimuli in which they could either use global information about the overall configuration of the display or, alternatively, focus on local information about detailed features or elements that make up that configuration. A common finding is that humans tend to show a “global precedence effect,” focusing on the overall stimulus configuration (see, e.g., J. Miller, 1981; Navon, 1977), whereas pigeons have shown a bias toward discriminating local features over global ones (Cavoto & Cook, 2001; Fremouw, Herbranson, & Shimp, 2002). There are exceptions to these general findings in both species. For instance, Goto, Wills, and Lea (2004) demonstrated a global precedence effect in pigeons. However, in their experiment, the stimuli were viewed at a distance of 17.5 cm and were small in extent, subtending 6° of visual angle. By contrast, a tendency of pigeons to base their discrimination primarily on local features of the stimulus display was reported by Cavoto and Cook (2001), who presented their stimuli on a monitor equipped with a touch screen. The largest of their stimuli measured 42 × 42 mm globally, with component features measuring 7 × 7 mm. If we assume a fixation distance of 60 mm, this gives a global extent of 40° × 40° of visual angle, and about 6.8° × 6.8° for their local features. The overall size of their stimulus display was comparable to that of our arrays; their “local features” were larger than the rectangles

in our arrays, which subtended about $1.9^\circ \times 2.4^\circ$ for the largest size of rectangle. So if global versus local precedence effects depend on viewing conditions and stimulus sizes, it is also more likely that the birds in our experiments focused on array items, rather than the global parameters of the array.

Which level of information is processed, whether global or local, is presumed to depend on attentional factors, rather than purely sensory ones (Cavoto & Cook, 2001; Fremouw et al., 2002). We suspect that the birds in our experiment most likely employed a visual search strategy of scanning localized regions of the stimulus display for individual items, along the lines suggested by Emmerton (1998), rather than integrating information across the whole array. Any bias that pigeons might have to process local stimulus components rather than a global pattern would also be enhanced by the variability in overall array configurations the birds encountered in all our experiments. It appears, though, that with a visual search strategy, it may become more difficult to track items when array elements are spread out, leading to some decrease in the birds' estimates of the number of items in the stimulus.

One alternative we considered was that birds might sample the density of a local region of the stimulus, and then be biased to judge the overall quantity of items as "large" if the local density is high. However, this type of bias still would not account for the birds' judgments of different numerosities, because, in the one-clump and two-clump arrays, there was little variation across numerosity in the average distances between elements. So, although density was higher in these arrays than in spread-out arrays, the high density was almost invariant across numerosity values. In any case, a strategy of sampling local density would not really concur with predictions derived from the cluster algorithm.

Overall, effects on performance of array configuration or spacing seem to be rather subtle, but do not fully account for the relationship between the birds' choices and the numerosity of a stimulus display.

EXPERIMENT 4

Most of the stimuli used in the training and testing phases of the previous experiments consisted of relatively large numbers of items. In the next experiment, we wanted to see if birds' discrimination of small numerosities was similar to their performance with larger quantities. Kaufman, Lord, Reese, and Volkman (1949) suggested that adult humans can accurately judge the numerosity of a few visual items through a process called "subitizing." Originally, the range in which this subitizing process appeared to operate was between 1 and 6 or 7 items. Kaufman et al. contrasted subitizing with a less accurate process, by which larger numerosities are assessed. They called this process "estimation." Later studies appeared to show a change in the slope of a reaction time (RT) function when the number of items to be assessed exceeded about 4, so this was taken as the upper limit for the subitizing range

(see, e.g., Chi & Klahr, 1975). The rapid RTs at small numerosities led to the assumption that subitizing involves a form of pattern recognition based on the configurations that small numbers of items typically form (e.g., three dots form a triangle). However, a more recent and extensive study of RT functions in adults showed no discontinuity that would indicate separate processes (Balakrishnan & Ashby, 1992). Also, D. J. Miller (1993) pointed out that there is no empirical evidence for subitizing in animals.

Another proposed mechanism by which small numbers of items may be differentiated (albeit indirectly) involves "object file" representations. For each item, a separate internal file is opened that stores information about an individual object's properties (e.g., its surface area, color, etc.). However, there is a limit to the number of object files that can be deployed simultaneously. This limit seems to be 3 for human infants (Feigenson, Carey, & Hauser, 2002) and 4 for macaque monkeys (Hauser, Carey, & Hauser, 2000). Both infants and monkeys have successfully discriminated successively presented sets of items when set sizes were small (e.g., 2 vs. 3) but failed to do so when either set exceeded the limit for object file representations (e.g., 3 vs. 4 for infants, and 4 vs. 5 for monkeys). Also, discrimination failed with numbers of items that exceeded the object file limit, even if the numerical ratio between sets was maintained. Thus, infants could discriminate 1 vs. 2 items but not 2 vs. 4, so Weber's law does not apply to the object file system.

Furthermore, even with small set sizes, infants may base their choice of quantity on properties such as the total surface area of items rather than on their number per se (Feigenson et al., 2002). Although there is empirical support for an object file representation system operating in infants and monkeys, it seems unlikely that pigeons rely on such a system, because they have already demonstrated their ability to discriminate arrays with much larger numbers of items than 3 or 4.

Yet another possibility is that avian numerosity discrimination resembles the performance of adult humans when they have to rely on nonverbal encoding of numerosity. Cordes, Gelman, Gallistel, & Whalen (2001) have examined humans' mean judgments of numerosity across a wide range of values, as well as the variability of these judgments, under conditions that would preclude enumeration, or verbal counting. These researchers found no evidence for the type of discontinuity that would be expected if small (<4 or 5) numerosities were subitized (or were represented as object files) and larger values were not. Instead, they argued for continuity in nonverbal numerical processing across both large and small values.

So the purpose of the next experiment was to compare the results of testing for interpolation and extrapolation with small numerosities with the results obtained within the larger numerosity range. If, as Cordes et al. (2001) have argued, there is a continuum of nonverbal numerical processing, then the psychometric function for test data in the small numerosity range should closely resemble functions from the previous experiments with larger numerosities.

Method

Subjects

Subjects were the same 3 birds that had served in Experiments 1 and 3.

Materials

Apparatus. The apparatus was the same as that used in the previous experiments.

Stimuli. For each numerosity value, specified in coding files, the locations of items in the stimulus arrays were randomly generated from an irregular matrix of 30 possible locations. The size of each item was also randomly generated from the three sizes (small, medium, and large), as defined in Experiment 1, so that most arrays consisted of random mixtures of element sizes.

Procedure

Birds were trained to discriminate the anchor values 2 versus 7. Training was conducted according to the procedures described in the previous experiments. Then five test sessions began. Testing was performed in the same way as in the previous experiments, with 96 trials per session, half being the reinforced trials with anchor value stimuli, and the rest being nonreinforced trials to test the birds' performance with eight numerosities. As in the earlier experiments, two numerosities were the same as the anchor values (2 and 7), four were interpolated values (3, 4, 5, and 6), and two were extrapolated values (1 and 8).

Results and Discussion

Test data at each of the eight numerosity values were averaged across birds and sessions, and are presented in Figure 6, together with the cumulative logistic function that was fitted to these scores. This function accounted for over 99% of the variance in the data. At a descriptive level, the psychometric function of the birds' discrimination performance with these numerosities closely resembles the functions that were obtained in Experiment 1. The "small" numerosities in this test extended beyond 4, so neither subitizing nor object file representations readily account for the smooth ogival form of the psychometric function.

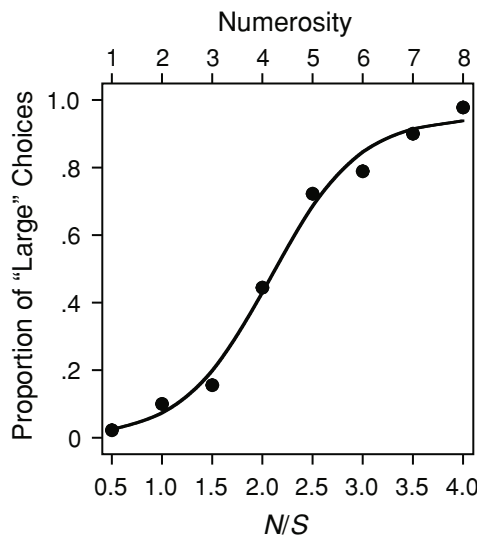


Figure 6. Test data, obtained with "small" numerosities, in Experiment 4. The data are fitted by a logistic function.

Although the stimulus arrays differed slightly between Experiment 1 (which used homogeneous arrays of same-size elements) and Experiment 4 (which used random mixtures of element sizes), this did not appear to affect performance. Since Experiments 1 through 3 showed that birds' judgments depended primarily on stimulus numerosity, with other stimulus properties playing a subsidiary role, we would not have expected these slight differences in stimulus arrays to have influenced the birds' judgments about the number of components.

GENERAL DISCUSSION

Psychometric Functions and Scalar Effects

In order to examine the data for other parameters that are consistent with both scalar principles and Weber's law, the psychometric functions from Experiments 1 and 4 were collated. One parameter, the numerosity value at the PSE, has been related to the subjective scaling of number, and to the type of decision process in the discrimination task (Gibbon, 1981). PSEs were derived from the psychometric functions obtained in Experiments 1 and 4. For each function, the numerical quantity corresponding to the midpoint between the maximum and minimum proportion of "large" choices was calculated.

Gibbon argued that if subjective scaling of numerosity is linear, and there is constant variability in judgment across numerosities, PSEs should occur at the arithmetic means of the anchor values in each test. If subjective scaling is approximately linear, but variability is scaled to mean numerosity judgments, then PSEs should correlate with geometric means of the anchor values, as long as judgments are based on a "similarity" rule—that is, an assessment of relative similarity between a test value and both anchor values, made by comparing their ratios. Alternatively, with linear subjective scaling, and scalar variability, the subjective midpoint between anchor values could occur at the harmonic mean of these training numerosities if a different discrimination process is assumed. This process involves a "proximity" rule, or an assessment of which anchor value most likely generated the noisy signal representing a test value. In bisection tasks on animals' timing ability, PSEs have often been reported at the geometric mean (Church & Deluty, 1977; Meck & Church, 1983; Platt & Davis, 1983; Stubbs, 1976). There has been less consistency in experiments on numerosity discrimination because PSEs sometimes occur at the geometric mean, sometimes at the harmonic mean, and sometimes in between (Fetterman, 1993; Fetterman, Dreyfus, & Stubbs, 1985; Fetterman, Stubbs, & Dreyfus, 1986; Martin-Iverson, Fibiger, & Wilkie, 1988; Roberts, 2005).

PSEs from the present experiments are plotted against each of the predicted types of means in Figure 7. Linear regressions were fitted to each plot. The correlation coefficients were all significant ($p < .05$). For each plot, the difference was tested between the regression coefficient (slope) for the obtained PSEs and a slope of 1, predicted if the PSEs coincided exactly with arithmetic, geometric, or harmonic means. These tests revealed no reliable differ-

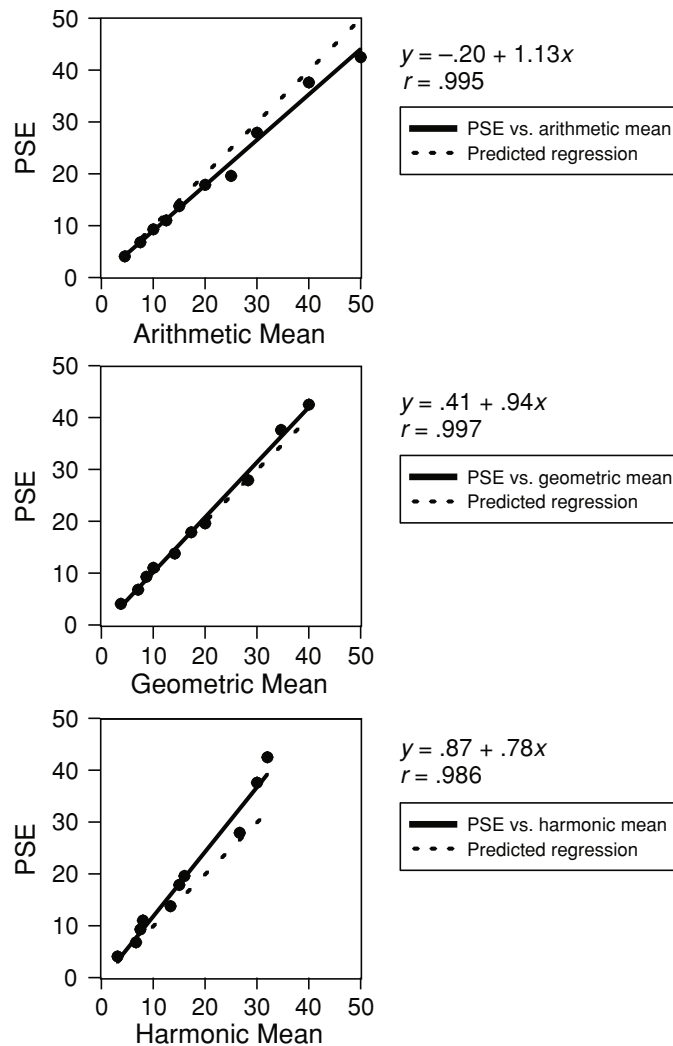


Figure 7. Numerosity values at the points of subjective equality (PSEs) calculated from the logistic functions fitted to test results in Experiments 1 and 4. PSE values are plotted against the arithmetic, geometric, and harmonic means, respectively, that were calculated based on the training numerosities for each test. Solid lines show the regressions that were fitted to each of these plots. Regression equations are shown to the right of each plot. Dashed lines show the regressions predicted if PSEs had occurred exactly at each of the three types of mean (i.e., if there had been a 1:1 correlation between PSEs and a given type of mean).

ence between the regression coefficient for PSEs versus geometric means and the predicted value of 1 [$t(16) = 1.62, p > .05$], whereas there was a significant difference from 1 of the regression slopes for PSEs versus arithmetic means [$t(16) = 3.51, p < .01$] and for PSEs versus harmonic means [$t(16) = 5.95, p < .001$]. Thus, the birds' PSEs best matched the geometric means of the various anchor values. This indicates that pigeons discriminate relative or ratio differences in numerosity with visual arrays of items.

The way in which the birds extrapolated their choices to stimulus values above and below the training values also suggests that birds discriminated numerosity ratios,

or based their choices on a similarity rule. For this type of decision rule to apply, not only must the PSEs be at the geometric means, but also stimulus values outside the anchor values must produce an equal or greater proportion of "large" choices for arrays with more items than the "large" anchor value, and an equal or smaller proportion of "large" choices for arrays with fewer items than the "small" anchor value. This was the general pattern of results seen in the psychometric functions.

This result is not always seen when "outside" probe values are included. In temporal bisection tasks with rats, Siegel (1986) reported that the PSE occurred at a duration that was less than the geometric mean of the anchor values

(see also Siegel & Church, 1984). Also, for outside values, the percentage of choices denoting a “long” duration did not concur with use of a similarity rule. Rats were more likely to choose “long” with durations that were shorter than the “short” anchor value, and less likely to choose “long” with durations that exceeded the “long” training stimulus. Siegel suggested that the rats applied a proximity rule, rather than a similarity rule. Meck (1997) reported similar effects in a bisection task in which rats had to judge the numerosity as well as the duration of auditory stimuli after they had been trained with stimuli in which time and number covaried. In both numerical and timing tests, the PSE was at a stimulus value that was less than the geometric mean. Furthermore, in these tests, the choices at outside probe values approached a 50% choice of the larger anchor value. Like Siegel, Meck attributed this pattern of results to rats’ use of a proximity rule. He argued that the rats learned the absolute values of the training stimuli, and that psychometric functions represented a combination of two different generalization gradients centered on the smaller and larger training values, respectively. It is not clear, at present, whether the presence or absence of extrapolation effects is due to species differences, types of stimuli, or other procedural factors. But the use of a similarity rule, although it accounts for the current data, is not universal.

Our finding of PSEs at the geometric means is consistent with scalar principles of a linear relationship between subjective scaling and objective numerosity, combined with scalar variability. So next, we assessed more directly whether variability in judgments about numerosity was proportional to the mean judgments. A measure of variability that can be derived from the logistic distribution functions is the difference limen (DL). This was calculated as half the difference in numerosity values corresponding to one quarter and three quarters of the maximum proportion of “large” choices. Figure 8A shows a plot of the

numerosity values for the DLs against those for the PSEs. The fit of the linear regression was significant ($r = .94$, $p < .05$), so these data indicate that variability in the birds’ judgments of numerosity increased in proportion to their mean subjective estimates of numerosity. Related to this, Weber fractions (the ratio of numerosity values for DL/PSE) were calculated and plotted against the subjective midpoints for numerosity (PSEs) (Figure 8B). The linear regression fitted to this plot had a slope of almost 0, so that Weber fractions were fairly constant across the numerosity range. The value of the Weber fractions, with the variability threshold used here, was approximately 0.2.

The choice data from tests in Experiments 1 and 4 were also replotted against a normalized scale of numerosity. Test numerosities were divided by the respective PSE numerosities so that the normalized value was 1 at the PSE, greater than 1 for a numerosity at which birds more often chose “large,” and less than 1 for a numerosity that was judged to be more similar to the “small” one. As Figure 9 shows, the test data superimposed, to a large extent.

Thus, on several measures, the results of these experiments conform to scalar principles. The finding of superposition effects complies with Weber’s law. The findings of (1) a linear relationship between numerosity and subjective midpoints, (2) a linear relationship between these midpoint values and measures of variability in numerosity judgments, and (3) PSEs being at the geometric means across a range of numerosity values, all concur with scalar theory.

Fetterman (1993) found evidence of superposition, at least for the same ratios of training values. He fitted his data with cumulative normal distributions, and derived the means and standard deviations from these functions as estimates of PSEs and variability in numerosity discrimination. In his study, the numerosities at the subjective midpoints also correlated best with the geometric mean, and the coefficient of variation (σ/μ of the Gaussian function)

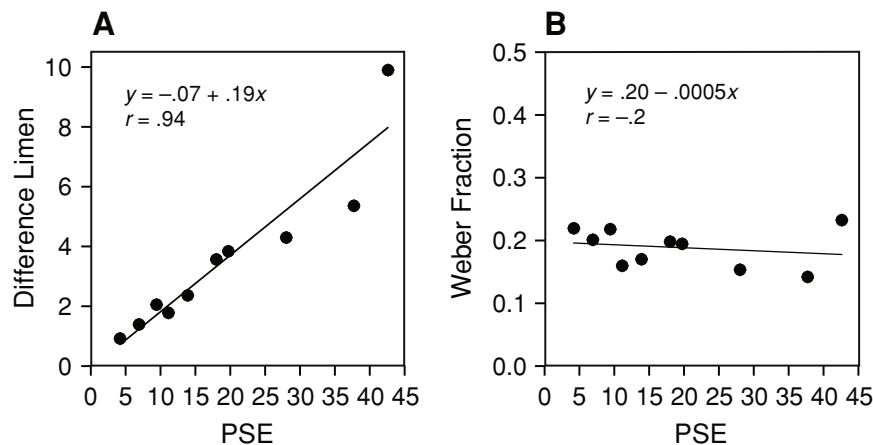


Figure 8. (A) Numerosity values for mean difference limens (DL) were calculated from logistic functions fitted to test results obtained in Experiments 1 and 4. DL values are plotted against numerosities at the subjective midpoints (PSEs) in each test. (B) Weber fractions (numerosities at DL/PSE) plotted against PSE numerosities. Solid lines show the regressions fitted to each plot.

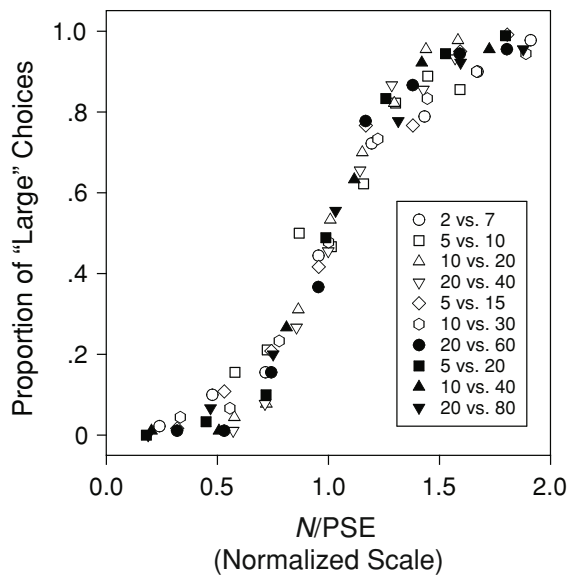


Figure 9. Test data from Experiments 1 and 4 plotted against a normalized numerosity scale to demonstrate superposition. The numerosity of each test stimulus (N) is divided by the numerosity value at the PSE for that test.

was roughly constant across the numerosities that were tested. Although mathematically different, the Weber fraction calculated in the present experiments provides similar information to the coefficient of variation. Thus, there are some close similarities between our data and Fetterman's. This suggests that both experiments tapped into the same type of numerical processing mechanism, in spite of the differences in the types of stimuli or events the birds had to assess, although a convergence of outcomes cannot be ruled out.

The addition of a test at small numerosities yielded no evidence of a discontinuity in measures of the subjective midpoint of discrimination, or of scaled variability about this mean point. In this respect, the data support the proposal of others who have studied humans' nonverbal numerosity estimation (Cordes et al., 2001; Gallistel & Gelman, 2000) that numerosity is subjectively represented on a continuum spanning small as well as large quantities, with no differences in underlying processes of the type suggested by "subitizing," or use of an object file system, when compared with "estimation."

Linear Versus Log Scaling

In interpreting our data as showing scalar variability, we make the computational assumption that subjective scaling of numerosity is linear. However, the main features of our results could equally well be obtained if we assume instead that subjective scaling is logarithmic, with constant log variability in each numerosity representation (see Dehaene, 1997). If scalar variability is assumed, the PSE at the geometric mean is obtained when ratios of the distributions representing numerosity are compared. Then, at the subjective midpoint, $PSE/S = L/PSE$, or $PSE = \sqrt{(S*L)}$,

where S and L are the small and large numerosities of the training stimuli. Variability at the PSE will increase in proportion to numerosity, since $PSE_{var} = \sqrt{(S_{var} * L_{var})}$, where var stands for a measure of variability in numerosity judgment, such as σ (see Gibbon, 1981) or DL. If constant log variability is assumed, the PSE at the geometric mean is given by the arithmetic mean of the log difference between anchor values, i.e., $\frac{1}{2}(\log L - \log S)$, or by equal intervals on a log scale, thus $(\log PSE - \log S) = (\log L - \log PSE)$. When DLs in our experiment were converted to log units of numerosity and plotted against log PSE values, the slope of the regression line was approximately 0, so variability was more or less constant, on a log scale. On the basis of the present results, it is thus not possible to distinguish between these two alternatives.

Elsewhere, though, Brannon, Wusthoff, Gallistel, and Gibbon (2001) interpreted their results in a "subtraction" task as evidence for a linear subjective scale of number in birds. But in his critique of that experiment, Dehaene (2001) demonstrated that Brannon et al.'s results could be simulated if either linear scaling with scalar variability or logarithmic scaling with constant log variability was assumed. More recently, Roberts (2005) trained pigeons to emit specific numbers of responses, and then tested them in a bisection task. He located the subjective midpoint at the geometric mean. In a second experiment, he obtained asymmetric results when the pigeons were trained to bisect the numerical range of emitted responses at the arithmetic mean. In simulations based on linear scaling with scalar variability versus log scaling with constant log variability, he found much better qualitative matches to his empirical data with log than with linear scaling. The differences in the simulation results are somewhat puzzling, though, if there is computational equivalence between both types of scaling and variability assumptions. As these examples illustrate, some aspects of birds' subjective representation of numerosity remain to be clarified.

Visual Search Strategy

Throughout the present experiments, the birds' choices were dependent primarily on the numbers of items they saw. Some other factors may have modulated their judgments of numerosity, as is sometimes the case in humans' visual assessment of numerosity. One factor that may affect pigeons' choices is the spacing of the items they have to process. Although the effect of spacing was not decisive, it was apparent that spreading out elements in visual displays does not have the same influence on birds' numerosity judgments as this spatial factor has on humans' estimation of numerosity. It is suggested that birds might use a visual search strategy to find items within localized regions of a stimulus display. If they do so, the partial effect of item size in the first experiment might also be related to this use of a search strategy. In that experiment, item locations varied randomly, so the effect of spacing could not be examined directly. But if pigeons have to search for items that may sometimes be farther away, then these items may become more salient if their size is larger, making them easier to find. This, in turn, could make it

slightly more likely that birds choose “large” rather than “small” for a given numerosity.

If this suggestion of a visual search strategy is correct, this could have some implications for the type of underlying mechanism that birds use to assess numerosity in these types of displays. If processing is via a neural network of the type that Dehaene and Changeux (1993) have proposed, then, for some aspects of avian perception of numerosity, their model might have to be modified to accommodate processing of elements at a local level, rather than parallel uptake across the entire retina of a global set of visual items. Regardless of details of processing mechanisms, however, the present experiments strengthen the case at a behavioral level for similarities across species in nonverbal abilities to discriminate numerosity.

REFERENCES

ALLAN, L. G., & GIBBON, J. (1991). Human bisection at the geometric mean. *Learning & Motivation*, **22**, 39-58.

ALLIK, J., & TUULMETS, T. (1991). The perception of visual numerosity. In R. J. Watt (Ed.), *Vision and visual dysfunction: Vol. 14. Pattern recognition by man and machine* (pp. 125-142). London: Macmillan.

BALAKRISHNAN, J. D., & ASHBY, F. G. (1992). Subitizing: Magical numbers or mere superstition? *Psychological Research*, **54**, 80-90.

BRANNON, E. M., & TERRACE, H. S. (2000). Representation of the numerosities 1–9 by rhesus macaques (*Macaca mulatta*). *Journal of Experimental Psychology: Animal Behavior Processes*, **26**, 31-49.

BRANNON, E. M., WÜSTHOFF, C. J., GALLISTEL, C. R., & GIBBON, J. (2001). Numerical subtraction in the pigeon: Evidence for a linear subjective number scale. *Psychological Science*, **12**, 238-243.

CAVOTO, K. K., & COOK, R. G. (2001). Cognitive precedence for local information in hierarchical stimulus processing by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, **27**, 3-16.

CHATLOSH, D. L., & WASSERMAN, E. A. (1993). Multidimensional stimulus control in pigeons: Selective attention and other issues. In T. R. Zentall (Ed.), *Animal cognition: A tribute to Donald A. Riley* (pp. 271-292). Hillsdale, NJ: Erlbaum.

CHI, M. T. H., & KLAHR, D. (1975). Span and rate of apprehension in children and adults. *Journal of Experimental Child Psychology*, **19**, 434-439.

CHURCH, R. M., & DELUTY, M. Z. (1977). Bisection of temporal intervals. *Journal of Experimental Psychology: Animal Behavior Processes*, **3**, 216-228.

CORDES, S., GELMAN, R., GALLISTEL, C. R., & WHALEN, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic Bulletin & Review*, **8**, 698-707.

DEHAENE, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.

DEHAENE, S. (2001). Subtracting pigeons: Logarithmic or linear? *Psychological Science*, **12**, 244-246.

DEHAENE, S., & CHANGEUX, J.-P. (1993). Development of elementary numerical abilities: A neuronal model. *Journal of Cognitive Neuroscience*, **5**, 390-407.

DEHAENE, S., DEHAENE-LAMBERTZ, G., & COHEN, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences*, **21**, 355-361.

DEHAENE, S., & MEHLER, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, **43**, 1-29.

EMMERTON, J. (1998). Numerosity differences and effects of stimulus density on pigeons’ discrimination performance. *Animal Learning & Behavior*, **26**, 243-256.

EMMERTON, J., LOHMANN, A., & NIEMANN, J. (1997). Pigeons’ serial ordering of numerosity with visual arrays. *Animal Learning & Behavior*, **25**, 234-244.

FEIGENSON, L., CAREY, S., & HAUSER, M. (2002). The representations underlying infants’ choice of more: Object files versus analog magnitudes. *Psychological Science*, **13**, 150-156.

FELLOWS, B. J. (1967). Chance stimulus sequences for discrimination tasks. *Psychological Bulletin*, **67**, 87-92.

FETTERMAN, J. G. (1993). Numerosity discrimination: Both time and number matter. *Journal of Experimental Psychology: Animal Behavior Processes*, **19**, 149-164.

FETTERMAN, J. G., DREYFUS, L. R., & STUBBS, D. A. (1985). Scaling of response-based events. *Journal of Experimental Psychology: Animal Behavior Processes*, **11**, 388-404.

FETTERMAN, J. G., STUBBS, D. A., & DREYFUS, L. R. (1986). Scaling of events spaced in time. *Behavioural Processes*, **13**, 53-68.

FETTERMAN, J. G., STUBBS, D. A., & MACEWEN, D. (1992). The perception of the extended stimulus. In W. K. Honig & J. G. Fetterman (Eds.), *Cognitive aspects of stimulus control* (pp. 1-20). Hillsdale, NJ: Erlbaum.

FREMOUW, T., HERBRANSON, W. T., & SHIMP, C. P. (2002). Dynamic shifts of pigeon local/global attention. *Animal Cognition*, **5**, 233-243.

GALLISTEL, C. R. (1990). *The organization of learning*. Cambridge, MA: MIT Press.

GALLISTEL, C. R., & GELMAN, R. (1992). Preverbal and verbal counting and computation. *Cognition*, **44**, 43-74.

GALLISTEL, C. R., & GELMAN, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences*, **4**, 59-65.

GIBBON, J. (1977). Scalar expectancy theory and Weber’s law in animal timing. *Psychological Review*, **84**, 279-325.

GIBBON, J. (1981). On the form and location of the psychometric bisection function for time. *Journal of Mathematical Psychology*, **24**, 58-87.

GIBBON, J. (1986). The structure of subjective time: How time flies. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 20, pp. 105-135). San Diego: Academic Press.

GINSBURG, N. (1976). Effect of item arrangement on perceived numerosity: Randomness vs. regularity. *Perceptual & Motor Skills*, **43**, 663-668.

GOODALE, M. A. (1983). Visually guided pecking in the pigeon (*Columba livia*). *Brain, Behavior & Evolution*, **22**, 22-41.

GOTO, K., WILLS, A. J., & LEA, S. E. G. (2004). Global-feature classification can be acquired more rapidly than local-feature classification in both humans and pigeons. *Animal Cognition*, **7**, 109-113.

HAUSER, M. D., CAREY, S., & HAUSER, L. B. (2000). Spontaneous number representation in semi-free-ranging rhesus monkeys. *Proceedings of the Royal Society of London: Series B*, **267**, 829-833.

HODOS, W., BESSETTE, B. B., MACKO, K. A., & WEISS, S. R. B. (1985). Normative data for pigeon vision. *Vision Research*, **25**, 1525-1527.

KAUFMAN, E. L., LORD, M. W., REESE, T. W., & VOLKMAN, J. (1949). The discrimination of visual number. *American Journal of Psychology*, **62**, 498-525.

KRUEGER, L. E. (1972). Perceived numerosity. *Perception & Psychophysics*, **11**, 5-9.

LUCE, R. D. (1959). *Individual choice behavior: A theoretical analysis*. New York: Wiley.

MACKO, K. A., & HODOS, W. (1985). Near point of accommodation in pigeons. *Vision Research*, **25**, 1529-1530.

MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A user’s guide*. Cambridge: Cambridge University Press.

MALDONADO, P. E., MATURANA, H. R., & VARELA, F. J. (1988). Frontal and lateral visual system in birds: Frontal and lateral gaze. *Brain, Behavior & Evolution*, **32**, 57-62.

MARTIN-IVERSON, M. T., FIBIGER, H. C., & WILKIE, D. M. (1988). Alteration in the perception of food quantity by rats induced by manipulations of hunger and food sweetness. *Learning & Motivation*, **19**, 44-65.

MECK, W. H. (1997). Application of a mode-control model of temporal integration to counting and timing behaviour. In C. M. Bradshaw & E. Szabadi (Eds.), *Time and behaviour: Psychological and neurobehavioural analyses* (pp. 133-184). Amsterdam: Elsevier.

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.

MILLER, J. (1981). Global precedence in attention and decision. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 1161-1174.

- MILLER, D. J. (1993). Do animals subitize? In S. T. Boysen & E. J. Capaldi (Eds.), *The development of numerical competence: Animal and human models* (pp. 149-169). Hillsdale, NJ: Erlbaum.
- NALBACH, H.-O., WOLF-OBERHOLLENZER, F., & REMY, M. (1993). Exploring the image. In H. P. Zeigler & H.-J. Bischof (Eds.), *Vision, brain, and behavior in birds* (pp. 25-46). Cambridge, MA: MIT Press.
- NAVON, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, **9**, 353-383.
- PLATT, J. R., & DAVIS, E. R. (1983). Bisection of temporal intervals by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, **9**, 160-170.
- ROBERTS, W. A. (2002). Mechanisms of "counting" in animals. In S. B. Fountain, M. D. Bunsey, J. H. Danks, & M. K. McBeath (Eds.), *Animal cognition and sequential behavior: Behavioral, biological, and computational perspectives* (pp. 153-174). Norwell, MA: Kluwer.
- ROBERTS, W. A. (2005). How do pigeons represent numbers? Studies of number scale bisection. *Behavioural Processes*, **69**, 33-43.
- 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. F. (1986). A test of the similarity rule model of temporal bisection. *Learning & Motivation*, **17**, 59-75.
- SIEGEL, S. F., & CHURCH, R. M. (1984). The decision rule in temporal bisection. In J. Gibbon & L. Allan (Eds.), *Timing and time perception* (Annals of the New York Academy of Sciences, Vol. 423, pp. 643-645). New York: New York Academy of Sciences.
- STUBBS, D. A. (1976). Scaling of stimulus duration by pigeons. *Journal of the Experimental Analysis of Behavior*, **26**, 15-25.
- VAN OEFFELEN, M. P., & VOS, P. G. (1982). Configurational effects on the enumeration of dots: Counting by groups. *Memory & Cognition*, **10**, 396-404.
- VAN OEFFELEN, M. P., & VOS, P. G. (1983). An algorithm for pattern description on the level of relative proximity. *Pattern Recognition*, **16**, 341-348.
- VOS, P. G., VAN OEFFELEN, M. P., TIBOSCH, H. J., & ALLIK, J. (1988). Interactions between area and numerosity. *Psychological Research*, **50**, 148-154.

(Manuscript received August 18, 2005;
revision accepted for publication November 9, 2005.)