Comparative cognition: Human numerousness judgments

ROGER K. THOMAS, JULIA A. PHILLIPS, AND CHERYL DONALDSON YOUNG
University of Georgia

Previous studies regarding numerousness judgments of dot displays showed humans to be accurate to 6 dots and 2 monkeys to be accurate to 8 dots and 1 monkey to 9 dots. However, the monkeys performed numerousness discrimination involving pairs of dot arrays and the humans made absolute numerousness judgments involving as many as 35 dot arrays. With 140 humans we replicated as closely as possible the discrimination method used with monkeys. On the three most difficult tasks, 5 of 20 subjects discriminated 8 versus 9 dots and 2 of 20 discriminated 9 versus 10 and 10 versus 11 dots. There were no significant differences in response time (RT) for successive numerousness pairs and no correlation between mean RT and dot arrays over the range from 3 to 11 dots. Results were discussed in relation to human and nonhuman animal investigations of numerical competence and were considered consistent with a noncounting process interpretation.

Stevens (1951, p. 22) proposed that the term *numerousness* be applied to "the cardinality attribute of groups of objects . . . that we observe when we look at, but do not count, a collection of objects." Stevens proposed that the term *numerosity* be used when the cardinality attribute was determined by counting. Thus, Stevens (1951) distinguished between counting and noncounting explanations for determining cardinality. "Subitizing" and "estimation" were noncounting processes suggested by Kaufman, Lord, Reese, and Volkman (1949) in their well-known study of human subjects' numerousness judgments of dot displays.

Strong evidence suggests that several species of nonhuman animals have varying degrees of numerical competence (Davis & Pérusse, 1988; Boysen & Capaldi, 1993), and the possible ecological relevance and adaptive significance of such competence has been discussed (Thomas, 1992). Less clear are the processes animals might use to manifest their numerical competence. For example, how might one explain the results of research that has shown that squirrel monkeys (*Saimiri sciureus sciureus* and *Saimiri boliviensus boliviensus*) can discriminate as many as

AMERICAN JOURNAL OF PSYCHOLOGY
Summer 1999, Vol. 112, No.2, pp. 215–233

seven from eight entities whether they are separated, such as arrays of dots (Thomas, Fowlkes, & Vickery, 1980), or joined, such as the number of sides or angles of randomly constructed polygons (Terrell & Thomas, 1990)? Thomas et al. (1980) offered no process explanation for their monkeys' judgments of dot displays, except that it was *not* counting; they merely referred to the monkeys as making "numerousness judgments."

Several processes to explain numerical competence in animals have been proposed, including counting (Boysen, 1993; Boysen & Capaldi, 1993; Boysen & Berntsen, 1989; Davis & Pérusse, 1988; Rumbaugh, Hopkins, Washburn, & Savage-Rumbaugh, 1989; Rumbaugh & Washburn, 1993). There is general agreement that the minimum evidence for counting should include evidence for use of the first three (of five) "principles of counting" as presented by Gelman and Gallistel (1978). These three principles are the one–one principle, which refers to applying "tags" to the items to be counted; the stable-order principle, which refers to how tags are applied (e.g., 1–2–3–4); and the cardinal principle, which means that the last tag applied represents the number of items in the set. Some studies using nonhuman animals appear to have provided this evidence (Boysen & Berntsen, 1989; Rumbaugh & Washburn, 1993).

Thomas et al. (1980; Terrell & Thomas, 1990; Thomas & Chase, 1980) ruled out counting to explain the squirrel monkeys' performances because the monkeys had not been trained to tag items in one–one correspondence in stable order. Yet the monkeys' accurate discrimination of seven versus eight dots (hereafter 7:8), for example, is evidence of the cardinal principle following Stevens's (1951) definition. Some might say that Thomas et al.'s (1980) monkeys were subitizing. However, based on the original definition of *subitizing*, it did not seem to apply. Kaufman et al. (1949) coined the term *subitizing* to denote the "sudden apprehension" of a number reported in their study of human visual discrimination of numbers of dots. Kaufman et al. defined subitizing as being based on accurate, confident, and rapid numerousness judgments that applied to dot arrays up to about six. Slower, less accurate, and less confident judgments of dot arrays greater than six were said to be based on "estimation." Kaufman et al.'s arrays were presented under conditions intended to preclude counting, namely, 200-ms stimulus presentation times. Kaufman et al.'s findings were replicated and extended by Mandler and Shebo (1982).

Thomas et al.'s (1980) squirrel monkeys met a rigorous performance criterion (45 correct in 50 successive trials) discriminating between successive dot arrays, which meets the accuracy criterion for subitizing. However, they did not assess speed or confidence of judgment, and the monkeys were accurate with displays of dots greater than the limit of six that Kaufman et al. had said applied to human subitizing. Because the monkeys lacked the requisite training to count and because the methods required them to discriminate between simultaneously presented arrays of dots, 200-ms presentation times were not used and the monkeys were allowed to view the discriminanda until they made a response. Therefore, it was possible that the "sudden apprehension" aspect of subitizing and, perhaps, the upper limit of six that was seen in Kaufman et al.'s (1949) research was related to the need to use 200-ms displays with their human subjects. There were other major differences between the methods used by Kaufman et al. and Mandler and Shebo (1982) that may explain the differences between their findings with humans and Thomas et al.'s findings with monkeys (see Discussion).

Subsequently, Terrell and Thomas (1990) and Thomas and Lorden (1993) proposed that the monkeys' performances might be explained by prototype matching, which has been used to explain other perceptual and conceptual category judgments (e.g., Rosch, 1973). Practice with trial-unique arrays of the same numbers of dots (e.g., seven dots) might enable monkeys to acquire a prototype (e.g., "sevenness") that could be used to identify accurately new exemplars of such arrays. Just as one's prototype in other examples of category judgments need not be an exact match of previously unseen exemplars, a numerousness prototype need not be an exact match of trial-unique displays of items in order to make accurate numerousness judgments.

To determine whether there might be further support for the monkeys' performances and an alternative interpretation to subitizing to explain the results with squirrel monkeys (Thomas et al., 1980; Terrell & Thomas, 1990), especially in view of the human data where the upper limit for accurate numerousness judgments was reported to be six (Kaufman et al., 1949), it was deemed useful to approximate the monkey experiments using human subjects. To try to prevent counting, we used 200-ms stimulus presentations followed by masking stimuli to interfere with afterimages. However, the 200-ms stimulus presentation times and the masking stimulus precluded using the simultaneous discrimination procedures that were used with the monkeys. Therefore, we used successive presentations of the two numerousness arrays to be discriminated. We informed each subject that half the arrays would have n dots (we specified n) and the other half would have $n + 1$ dots. Subjects were instructed to press one button to indicate when an array was perceived to be n and another button to indicate when the array was perceived to be $n + 1$.

Although the emphasis here was on the dot displays for comparison with previous research using human subjects, we also had each subject perform numerousness judgments based on the number of sides (or angles) of randomly constructed polygons for comparison with the previous monkey research (Terrell & Thomas, 1990). Response times (RTs) were recorded because they were important to the process interpretations suggested by Kaufman et al. (1949) and Mandler and Shebo (1982). Consistent with a noncounting interpretation, we hypothesized that the response times for the two discriminanda of a successive-number-pair (dots or sides) would not differ significantly. We made no a priori prediction concerning possible differences in RT over the range from 3-dot to 11-dot arrays used here.

EXPERIMENT

METHOD

Participants

Subjects were 140 undergraduate volunteers (79 female, 61 male) ranging in age from 18 to 26 years. Their participation was gained via the Research Participant Pool in the University of Georgia's Department of Psychology, and each subject received class credit. The subject sign-up chart noted the following restriction: "Must have normal or corrected-to-normal vision. If you wear glasses, please bring them to your appointment." Although no tests were conducted to ensure adequate vision, no subject reported any difficulty perceiving the discriminanda and all subjects met criterion on the practice tasks. All subjects were treated in accordance with the "Ethical Principles of Psychologists" (American Psychological Association, 1981), and the research protocol was approved by the University of Georgia's Institutional Review Board.

Apparatus

Two Kodak Ektagraphic IIIB slide projectors, fitted with shutters controlled by a Lafayette 42011–A Constant Illumination Tachistoscope, were used to project photographic slide stimuli to a standard slide projection screen (97.5 cm × 95 cm; Day-Lite Picture King). One projector was used to present a discriminandum, and coincident with its offset, the second projector presented a masking stimulus. Subject responses were made on a locally constructed, enclosed response box (38 mm × 51 mm × 102 mm), the top of which was fitted with two pushbuttons (9.5 mm × 13 mm) spaced 40 mm apart and a green lamp (9.5 mm diameter) located midway between the buttons (box, buttons, and lamp were purchased from Radio Shack). A label beneath the button on the left read "LOW" and a label below the button on the right read "HIGH" to denote which was to be pushed for the low and high number in a pair. The response box was connected by a cable to an interface apparatus and the subject could hold the box on the writing surface of a typical, movable student classroom desk at which she or he was seated. A Commodore 8032 microcomputer programmed by locally written software was used to initiate trials, record subject responses and RTs, and generate summary statistics. An electronic interfacing apparatus constructed by the University of Georgia's Electronics Design and Maintenance Shop was used to interface the microcomputer, the tachistoscope, and the response box, and it also debounced input from the response box buttons.

Discriminanda

The discriminanda were photographic slides of displays of black-filled circles ("dots") or black-filled irregularly shaped polygons that had been drawn on plain white index cards. To control against the use of cumulative area or brightness cues, dots were selected randomly from three sizes (visual angle data follow). Dot clusters were determined by random assignment to loci on a 4 × 4 grid with the restriction that no dot was separated from another dot by more than one locus site (this was to prevent distinguishable clusters such as two dots in one corner and three dots in another corner); the grid loci were sufficiently separated that no two dots would be closer than the distance of the diameter of a small dot. Numerousness was represented by the number of dots in a display or by the number of sides or angles of a polygon. Because the number of sides and the number of angles of a polygon are the same, for convenience, we will refer to the numerousness cues from the polygons as being based on the number of sides. Sizes of polygons were varied to preclude the use of cues based on area or brightness differences. (For further details about the stimuli and controls incorporated, see Thomas et al., 1980; Terrell & Thomas, 1990; Thomas & Lorden, 1993.) The illuminated field on which the dots or a polygon appeared subtended a visual angle of approximately 5.7° The large, medium, and small dots subtended visual angles of approximately 0.86°, 0.57°, and 0.29°, respectively. Testing was done between 9:00 a.m. and 4:00 p.m. During stimulus presentation phases, the room lights were extinguished and the ambient room illumination was daylight from a single window (aperture of 88 cm × 152 cm) with a closed adjustable blind. The subject and viewing screen were approximately 5 m from the window.

One hundred unique slides in each stimulus category were used. If a subject met criterion within 100 trials, all trials were unique; otherwise, the same set of 100 slides was repeated once. Masking stimuli were presented immediately following each presentation of a numerousness discriminandum to prevent counting based on afterimages. The masking stimulus filled the viewing field with various sized dots or various sized, irregularly shaped polygons, depending on whether a subject was in the dots or polygons phase of the study.

Procedure

Subjects were randomly assigned to one of seven groups, each with 20 subjects. Each subject was tested on dots and polygons. Randomly determined, half the subjects within a group were tested on dots first and half were tested on polygons first. All subjects, regardless of group, received practice experience that involved three versus four discriminations (3:4). Within a group, the subjects were tested on one additional number pair from the following: 4:5, 5:6, 6:7, 7:8, 8:9, 9:10, or 10:11. Because practice with the 3:4 tasks was administered under standardized conditions, the data are included here.

Subjects were tested individually. After reading and signing a consent form, the subject was seated in the student desk on which the response box was placed. The desk was 2 m from the projection screen. Instructions were read aloud by the experimenter to each subject, with appropriate changes in the instructions to reflect whether the dots or polygons were presented first and which numerousness task the subject received.

For the practice task, slides depicting either three or four dots or polygons with three or four sides, depending on group assignment, were presented, one at a time for 200 ms each. The order of presentation of stimuli for depicting three or four was determined by the Fellows (1967) series. A masking stimulus immediately replaced the discriminandum and remained on the screen until the next stimulus was presented, including a 2-s interval between the subject's response to one stimulus and the presentation of the next stimulus. The subject pressed the button on the left side of the response box if the slide depicted three entities (the button was labeled "LOW," although the subject was always told the actual number to be judged) and pressed the button on the right if four entities were depicted (the button was labeled "HIGH," but the subject was always told the actual number to be judged). If the subject responded correctly, the lamp on the response box was illuminated. Practice continued until the subject attained 90% correct in a block of 20 trials. After completing the first practice task, the subject received the first experimental task, either 4:5, 5:6, 6:7, 7:8, 8:9, 9:10, or 10:11. General procedures were the same as for the practice task; that is, a discriminandum was presented for 200 ms, a masking stimulus was projected, the subject pressed the button on the left if the slide depicted the lower number and the button on the right if the higher number was depicted, and the lamp on the response box was illuminated when the subject responded correctly. Each presentation of a discriminandum constituted a trial, and trials were presented until the subject attained 90% correct responses in 20 successive trials or until 200 trials had been administered. If the subject did not reach 90% correct responses within the first 100 trials, she or he was given a rest break of approximately 5 min, and the slide tray was repositioned for the second run of 100 trials.

After completing the first experimental task, the subject was given the second 3:4 practice task (either dots or polygons, depending on which had not yet been seen), followed by the second experimental task. For example, if a subject was assigned to the 7:8 group and had dots first, the order of testing would be practice on 3:4 dots, test on 7:8 dots, practice on 3:4 polygons, and test on 7:8 polygons. After completing the second experimental task, the subject was asked to complete a brief questionnaire that asked whether she or he had counted. Response categories were "yes," "no," "sometimes," and "almost always." Before leaving, each subject was asked to read and sign a debriefing statement that explained the experiment, and she or he was given an opportunity to ask questions about the experiment.

Instructions

Although the instructions iterate some of the information already given and because our results may appear to conflict with previous studies (but see *Dis-*

cussion), any who may replicate this experiment should be informed about the precise instructions our subjects received. The following instructions were read to subjects who had the dots task first. An appropriately altered set of instructions (available on request) was read to subjects who had the polygons task first. The instructions were also adjusted appropriately to denote the specific numerousness group to which the subject was assigned; this is indicated by n and $n+1$ in brackets.

> This study is concerned with the ability to make number judgments without counting. Therefore, please try not to count. We are going to show you two kinds of things. First, we will show you slides with dots. In order to let you become familiar with the task, we will have you practice with an easy task, then you will receive a slightly harder task for the main experiment. The practice task involves slides which have only 3 dots or 4 dots. If there are 3 dots, press the button on the left. If there are 4 dots, press the button on the right. Whether a slide shows 3 or 4 dots will be determined randomly. When you respond correctly, the light on the box with the buttons will be illuminated. When you are getting 90% of your responses correct, the practice on the dots will end, and we will move on to a somewhat harder task. How hard your next task will be is determined by random assignment. We are trying to learn how well people can do on all the tasks, so if yours is a hard one, don't worry about how you do, just try to do your best. We think you can do your best if you will try to relax and not think too much about it. Just pay attention to the dots as best you can and give us your best guess. The slide will be present for only a fraction of a second and it will be followed immediately by a slide which is intended to interfere with your ability to count the dots, but it won't interfere with your ability to estimate the dots. We also want you to respond as quickly as you can. Finally, there will be only a few seconds between slides, so pay attention to the screen at all times. Do you have any questions?

After the subject had completed the first practice task, the experimenter read the following instructions.

> Good, you did very well. Now, we are going to give you the somewhat harder task. Remember, the task you receive was determined randomly. In your case, we are going to give you [n] dots and [$n+1$] dots. Press the button on the left if you think there were [n] dots and press the button on the right if you think there were [$n+1$] dots. Remember to respond as quickly as you can.

After the subject completed the first experimental task, the experimenter read the following:

> Now, we want to change to a slightly different task. Instead of dots, we are going to show you some randomly constructed polygons and ask you to estimate the number of sides and angles. The number of sides and angles is the same for each polygon. As we did before, we will let you practice on three-sided polygons and four-sided polygons. Other than that, everything else is the same. That is, please try not to count. Press the button on the left if you think the answer is 3 and press the button on the right if you think the answer is 4. Respond as quickly as you can. Do you have any questions?

After the second practice task was completed, the experimenter read the following instructions.

Finally, as we did with the dots, we are going to give you polygons with [n] sides and polygons with [$n + 1$] sides. As you did before, try not to count. Please press the button on the left if the polygon had [n] sides and the button on the right if it had [$n + 1$] sides. Try to respond as quickly as you can.

RESULTS

Accuracy for numerousness judgments was found to be greater than the six-dot accuracy that has been reported in previous studies using humans; accuracy here was extended to 11-dot arrays for 2 of 20 subjects tested with on 10:11 dots. Based on monkey research, it was expected that accuracy with higher-number dot arrays would be seen with humans when a numerousness discrimination procedure is used. Additionally, as indicated by the data in Table 1 (dots and polygons), Figure 1 (dots), and Figure 2 (polygons), the hypothesis that RTs would not differ significantly for numerousness judgments of successive numer-

Table 1. Numbers of subjects trained (NT) and reaching criterion (NRC), mean trials to criterion (TC), mean response times (RT), and standard errors of response times (RTSE) in seconds for exemplars of low and high numerousness judgments

				Low		High	
Tasks	NT	NRC	TC	RT	RTSE	RT	RTSE
Dots tasks							
3:4	140	140	30	0.79	0.03	0.75	0.01
4:5	20	20	39	0.77	0.04	0.82	0.05
5:6	20	13	75	0.92	0.06	0.92	0.06
6:7	20	17	114	0.91	0.07	0.92	0.06
7:8	20	8	140	0.82	0.06	0.83	0.08
8:9	20	5	152	0.81	0.02	0.83	0.03
9:10	20	2	153	0.74	0.10	0.72	0.07
10:11	20	2	170	0.96	0.14	1.01	0.09
Polygons tasks							
3:4	139[a]	139	22	0.75	0.02	0.70	0.01
4:5	20	20	52	0.78	0.05	0.92	0.09
5:6	20	5	120	1.22	0.02	1.08	0.24
6:7	20	6	163	1.02	0.07	0.95	0.06
7:8	20	3	160	0.83	0.04	0.94	0.07
8:9	20	2	110	1.02	0.21	0.99	0.09
9:10	19[a]	1	120	1.07		1.02	
10:11	20	1	160	1.06		1.13	

[a]Apparatus malfunctioned during the polygon test for one subject in the 9:10 group who had attained criterion on the dots test.

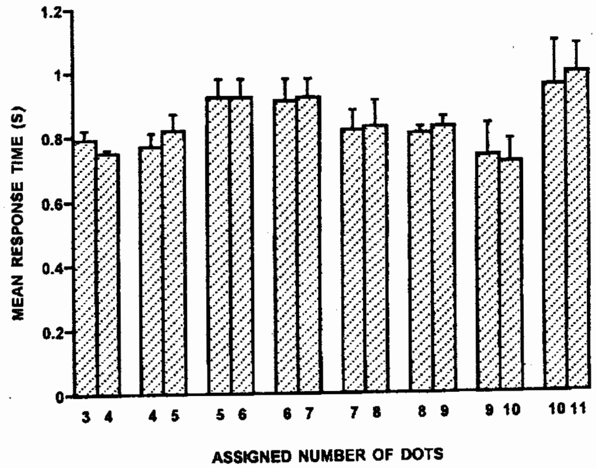

Figure 1. Mean response times and standard errors in seconds for subjects who met criterion on the dots discrimination tasks (see Table 1) as a function of their assigned numbers of dots

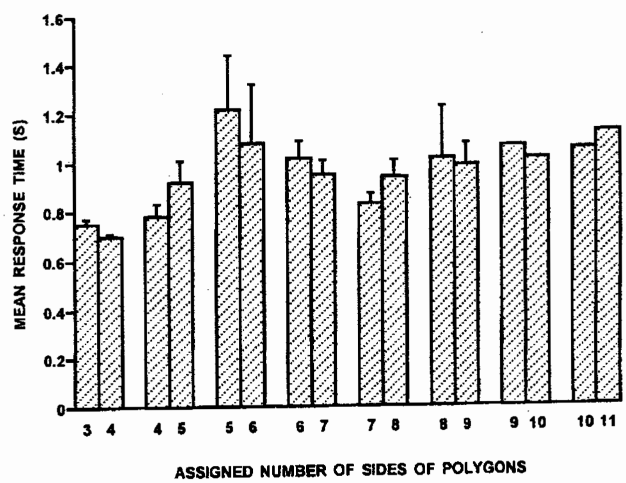

Figure 2. Mean response times and standard errors in seconds for subjects who met criterion on the polygons discrimination tasks (see Table 1) as a function of their assigned pairs of polygons. There is no standard error for 9:10 and 10:11 because only one subject met criterion on each of those tasks

ousness pairs was confirmed. Because we hypothesized only that the RTs in each successive pair of numerousness discriminanda would not differ significantly, it was most appropriate to assess each number pair independently and to do so only using the data from the subjects who met criterion. An appropriate inferential statistical test is the *t* test for correlated means (e.g., Meyer, 1976). There were no significant differences between the mean RTs of any of the successive number pairs, 4:5 through 10:11, for dots or polygons (all $ps > .05$).

We made no hypothesis regarding RTs other than that based on comparisons within numerousness pairs (e.g., 7:8), and a post hoc analysis such as analysis of variance over all numerousness arrays would be problematic because the numbers of subjects reaching criterion on the different tasks varied from 2 to 20 (Table 1). However, to address the possible relationship between numerousness arrays and RTs over the range of numerousness displays used, it might be useful to examine the correlation coefficient for these variables. Because exemplars of most dot arrays were used as the low number in one pair and the high number in another pair, we pooled the weighted mean RTs to arrive at a single mean response time for each number. It seems apparent from the data shown in Figures 1 and 2 that such pooling did not misrepresent the means. The Spearman correlation coefficient (Siegel, 1956) was 0.50 ($p > .05$). Therefore, the relationship, if any, between RT and the number of dots for the dot arrays used in the present investigation was not statistically significant. Similarly, the correlation coefficient was computed for RTs as a function of the number of sides of the polygons, and the Spearman r was 0.72 ($p < .025$). It is not obvious why the polygon discriminanda might affect subjects differently from the dots discriminanda with regard to numerousness RT, as suggested by the significant correlation coefficient versus the nonsignificant correlation coefficient, but it is not believed to be caused by differences in the "complexity" of polygons for reasons that are extensively discussed in Terrell and Thomas (1990).

Despite the comparable RTs for the dots, two kinds of evidence suggested that the tasks became increasingly difficult as a function of the numbers of items to be judged. First, as can be seen in Table 1, the number of trials to criterion for subjects who met criterion on the dots tasks increased systematically. There is a perfect positive Spearman correlation coefficient between increasing numbers of dots to be discriminated and the numbers of trials to criterion ($r_s = 1.00$, $df = 7$, $p < .01$). Second, there was a strong negative correlation between the increasing numbers of dots to be discriminated and the percentages of subjects who reached criterion ($r_s = -.84$, $df = 7$, $p < .01$). For the polygons tasks, there was a moderately strong correlation between increasing numbers of sides to be discriminated and the number of trials to criterion (r_s = 0.68, $df = 7$, $p < .05$) and a strong negative correlation between the increasing numbers of sides to be discriminated and the percentages of subjects reaching criterion ($r_s = -.94$, $df = 7$, $p < .01$). Thus, these correlations indicate that the numerousness judgment tasks became increasingly difficult as a function of increasing numerousness.

The results of our postexperiment questionnaire showed that 80% of the subjects checked "no" when asked whether they had counted. The remaining subjects indicated they counted "sometimes" (none checked "yes" or "most of the time"). Based on follow-up questioning of subjects who said they counted "sometimes," it was clear they did not intend the usual meaning of tagged counting in stable order. In fact, the subjects generally could not explain how they made their judgments.

DISCUSSION

The results of the present study using a numerousness discrimination method indicated that the upper limits of accurate numerousness judgments may be higher than previously reported for humans (Kaufman et al., 1949; Mandler & Shebo, 1982). The results here are consistent with results reported earlier for squirrel monkeys (Thomas et al., 1980; Terrell & Thomas, 1990), which suggests a numerical competence that may be widespread among primates and, perhaps, nonprimate animals. Higher percentages of the monkey subjects than the human subjects met criterion with numerousness arrays containing 6:7 and 7:8 items, but this is probably explained by the humans being limited to 200-ms viewing times and to 200 trials in which to reach criterion. On the other hand, humans here outperformed squirrel monkeys in that 2 of 20 human subjects, respectively, discriminated between 9:10 and 10:11 dots. One of two squirrel monkeys met criterion on 8:9 dots (Thomas et al., 1980), and it averaged about 75% correct on a 9:10 discrimination problem when his training was ended at 500 trials to avoid induction of "experimental neurosis" (Thomas et al., 1980). It seems reasonable that a sufficiently large sample of squirrel monkeys might yield a few monkeys able to discriminate between 9:10 and 10:11 as the few humans did, but the time, expense, and opportunity to obtain such data are likely to be prohibitive. It is also possible that with further and more prolonged testing, more than 10% of human subjects might reach criterion on the 9:10 and 10:11 tasks and, perhaps, even extend the range beyond 10:11. However, it also seems likely that the upper limit of the *typical* human's capacity to discriminate dot arrays is closer to 7:8.

The present investigation deliberately confounded the possibility of absolute and relative numerousness judgments because it was our intention to replicate as closely as possible the earlier monkey research (Thomas et al., 1980). Absolute and relative numerousness judgments were deliberately confounded in the monkey research because that study was done when little was known about the capacity of nonhuman animals for numerousness judgments (see literature reviewed by Thomas et al., 1980), and they wanted to maximize the likelihood of determining what that capacity might be for squirrel monkeys. Although the subjects in the present investigation were informed of the two numbers involved in a given task, it was also the case that they would know that one number represented more entities (higher number) and the other represented fewer entities (lower number). In fact, the response buttons were labeled "HIGH" and "LOW" for each subject, but even if we had changed the labels, that would not have prevented a subject from responding in terms of "high" and "low" rather than, for example, "7" and "8." However, the main point is that to meet criterion the subjects had to be able to recognize that the two successive numerousness sets represented discriminably different numbers of items. Therefore, for example, whether the subject conceptualized the task as "7" versus "8" or "low" versus "high," she or he had to be able to discriminate "sevenness" from "eightness" in order to reach criterion.

Comparing present results with those of Kaufman et al. (1949) and Mandler and Shebo (1982)

Kaufman et al. (1949), confirmed by Mandler and Shebo (1982), reported a systematic increase in RT as a function of increasing numbers of items in an array, and that result has been used to suggest that a serial, counting-like process may be involved in numerousness judgments by humans and nonhuman animals (Gallistel, 1988). Our finding of no difference in mean RT between successive numerousness displays and the nearly uniform mean RTs over the range of numerousness arrays used in the present study may appear to conflict with the findings of Kaufman et al. (1949) and Mandler and Shebo (1982). However, methodological differences between our study and theirs reasonably explain the seemingly conflicting results.

Probably the most relevant difference was that our subjects responded concurrently to two arrays, whereas subjects in the previous investigations responded concurrently to no fewer than 15 different numerousness arrays. Subjects in Kaufman et al.'s (1949) study were required to respond concurrently to 35 different arrays that ranged from 1 to 210 dots. Another important difference was that we controlled for cumulative area or differential reflected brightness cues by varying the sizes of the dots or polygons, whereas the previous studies used uniform sizes of discriminanda. Kaufman et al. used uniform dots and Mandler and

Shebo (1982) used uniform *X*s and *O*s in either separate or mixed arrays. Furthermore, as a control for density, Kaufman et al. clustered their entities in ways (see their Figure 3 and related discussion) that may have enhanced area and brightness cues, thereby adding them as cues that their subjects might have used. It is less clear from Mandler and Shebo's description, but it appears that they might have had a similar confound. However, we do not dispute the findings or conclusions of these previous studies, which were designed to investigate questions different from the ones investigated here.

Miller's "magical number seven, plus or minus two"

Miller's (1956) magical number seven referred to "information processing channel capacity" as applied to the discriminability of unidimensional stimuli. Miller cited data to suggest that the limit of such information processing channel capacity was approximately 2.5 bits of information or about six discriminable stimuli. Miller himself discussed the relevance of Kaufman et al.'s (1949) results. Initially, Miller emphasized Kaufman et al.'s finding of the accuracy to six dots and the interpretation that "below seven the subjects were said to *subitize*" (Miller, 1956, p. 90). However, after further consideration Miller (1956, p. 90) concluded that "there seems to be a reasonable suspicion that it [the 6 to 7 "break" in the data] is nothing more than coincidence." Miller (1956, p. 90) also suggested their subjects' likely use of "20 or 30 distinguishable categories of numerousness," referring to Kaufman et al.'s 35 arrays ranging from 1 to 210 dots. This led Miller (1956, p. 90) to suggest that Kaufman et al. probably had "two dimensions of numerousness . . . area and density."

Miller (1956) did not address how area and density were the two dimensions of numerousness, but our consideration of Kaufman et al.'s (1949) dot arrays (see examples in their Figure 3) indicated that their dot arrays offered at least six potential cues in addition to number per se that varied systematically with the number of dots. First, Kaufman et al. used uniform-size dots, which results in a cumulative dot-area that increased systematically with the number of dots in an array; conversely, the cumulative area of the background devoid of dots decreased systematically. Second, Kaufman et al. clustered the dots to try to control for density. However, if one considered the cumulative area of the field that results from connecting the dots on the perimeter of a cluster, the resulting field represents another cumulative area cue that increased systematically with the number of dots; conversely, the ground on which this field appears decreased systematically as the number of dots increased. Finally, Kaufman et al. used white dots on a dark background, which means that the cumulative luminance of the dots increased systematically with the number of dots; conversely, the cumulative dark background decreased systematically with increasing numbers of dots. Although these six nonnumerousness cues are not independent, each could be used independently or cumulatively as a consistent correlate of numerousness. Therefore, Miller's assessment of 20–30 discriminable "numerousness categories" among Kaufman et al.'s discriminanda is reasonable.

In the present experiment, we controlled area and brightness cues by using dots of varying sizes. We used three dot sizes and randomly assigned them to loci on a 4 × 4 grid to construct each dot array. Although the randomization-of-size procedure should (and did) result in a slight but systematic increase in average cumulative area as a function of the number of dots, a subject could not reach the discrimination criterion between successive numerousness pairs by relying on cumulative area or cumulative luminance differences. Any series of trials involving successive number pairs included a significant percentage of the trials in which the fewer-dot-array had more cumulative area than the more-dot-array. Therefore, a subject could not use area or brightness cues and attain the criterion. The mean sizes of the polygons used in the present investigation did not differ significantly from each other for successive numerousness pairs, except that the mean area of the tetragons (39% of the field) was significantly larger than the mean areas of both the triangles (24%) and the pentagons (21%).

Using dots that varied in size and random assignment, using a 4 × 4 grid to define their cluster, should reduce density as a cue. However, what determines density and, therefore, what an appropriate control for density might be is not as apparent as it may seem. For example, which has greater density, seven dots or eight dots when seven dots have greater cumulative area? One could equate cumulative areas of the dots in the different numerousness arrays, but would such equalization eliminate density as a cue? Would seven-dot and eight-dot arrays that are equal in cumulative dot area and confined to the same-size field have equal density? Should field size be increased systematically with the number of dots? If so, would the variation in overall field size provide a extra-numerousness cue?

Raising such questions indicates a need for additional research regarding the importance, if any, of cues (including density) that might influence numerousness judgments other than or in addition to number of entities per se. Because a few subjects here performed better than the 2.5 bits of information predicted for unidimensional stimuli (Miller, 1956), possibly they used nonnumber cues to enhance their performances at the higher numbers. It may be recalled that the overall field size for the dots was not identifiable; nevertheless, it was constant, and a density-related cue might become more apparent to some subjects with the higher numbers of dots. On the other hand, Miller's "plus or minus two" denotes variability even among unidimensional stimuli, and for a few exceptional individuals that variability may be greater than two. The higher percentages of subjects who succeed when the numbers are closer to seven plus or minus two suggests the high salience of the number of entities as opposed to extra-number cues. In any case, these empirical questions can be addressed through further research.

There are other useful questions to investigate. For example, our finding that humans can judge numerousness accurately when the sizes of the dots and polygons are varied as a control for cumulative area and when the dots are distributed more randomly as compared to the clusters seen in Kaufman et al. (1950) indicates that it is feasible to compare performances using uniform-sized versus varied-size dots as well as using clustered versus more randomly distributed arrays to see whether such controls make any difference in the speed and accuracy of numerousness judgments. It would also be useful to investigate the capacities of human and nonhuman animals for absolute numerousness judgments using area and distribution controls, but under conditions in which the possibility of relative judgments has been eliminated, to see whether similar upper limits of accuracy will continue to be the case.

Processes of numerical competence and implications for comparative cognition

In view of the fast and nearly uniform RTs, one might say that the range for subitizing had merely been extended. However, *subitizing* is a descriptive term that offers less as a potential explanation than, for example, a term and conceptualization such as *prototype matching.* Although we do not have direct evidence for a prototype matching process (nor do we know what such evidence might be), we believe that such a process can reasonably explain the results seen with monkeys (Terrell & Thomas, 1990; Thomas & Chase, 1980; Thomas et al., 1980) and the humans here. Regarding how such a process might work, Dehaene and Changeux's neuronal network model to explain numerousness judgments may be applicable. Dehaene and Changeux characterized their model as follows:

> Our simulations . . . demonstrate the feasibility of extracting approximate numerosity in parallel from a visual display, without serial counting. Our model therefore demonstrates how one may account for animals' and human infants' numerical abilities without assuming they can count. (1993, p. 401)

Although Dehaene and Changeux's (1993, p. 402) model was limited to five entities, they noted, "This limit is arbitrary and was chosen for computational convenience only."

How a prototype matching process might work, other than as proposed in the Dehaene and Changeux (1993) model, is unclear. However, the prototype matching process envisioned here is *not* like a judgment process based on canonical patterns (symmetrically spaced dots) such as those used by Mandler and Shebo (1982) during part of their investigation (see their Figure 9). Nor is the prototype matching process envisioned here comparable to the "perceptual norm" process described by Rumbaugh, Savage-Rumbaugh, and Hegel (1987), despite Rumbaugh and Washburn (1993, p. 95) having cited Rumbaugh et al.'s (1987) perceptual norm as "predat[ing] a related [process] . . . argued by Thomas and Lorden." Referring to Hicks's (1956) investigation of rhesus monkeys' use of a threeness concept, Rumbaugh et al. wrote,

> In reference to the kind of study conducted by Hicks, we suggest that a perceptual norm, rather than a number concept, has been acquired and that such a norm is based on the limited configurations inherent in the simultaneous presentation of three items or objects. . . . Their pattern is always that of a triangle unless they are aligned. . . . Thus we suggest that *particularly* for small numbers of things placed on a field, a perceptual norm, reflecting patterns of placement can serve as the basis for animals to learn what appears to be a number concept. (1987, p. 108)

While many trials with three entities, perhaps including some in the present study, might result in patterns contributing to a perceptual norm or to canonical patterns, the likelihood of such norms or patterns is less here because dots were unequal in size. Furthermore, the likelihood of having arrays in the present work or the earlier work with squirrel monkeys (Thomas et al., 1980) that contributed to a perceptual norm or formed canonical patterns should be much less with four entities and should be rare with five or more entities. Furthermore, that the mean RTs for three and four entities (as well as five and above) did not differ significantly with the dot arrays here suggests that similar rather than different judgment processes were being used. In hindsight, it might have been useful to question subjects more specifically in the posttest interview about their judgments of three and four entities in the training task as compared to their judgments with the higher-number pair to which they were assigned to see whether they perceived that they were using different processes.

Assuming that a noncounting process can reasonably explain the results reported for squirrel monkeys (Terrell & Thomas, 1990; Thomas et al., 1980) and explain most if not all of the human data reported here, there is an important implication of these results for investigations of counting by nonhuman animals. Even if tagging in one–one correspondence and stable order has been demonstrated (the two criteria that precede cardinality), it is still possible that the animal may have used a noncounting process to arrive at the cardinality attribute. Furthermore, it appears reasonable that a noncounting process might be used to determine the cardinality attribute of at least as many as eight or nine entities; recall that one of the monkeys in Thomas et al.'s study (1980) discriminated 8:9, which suggests that a definitive investigation of counting by animals may require counting beyond nine.

A noncounting process interpretation may seem more likely when entities are presented simultaneously; however, it is reasonable that, within limits, the numerousness of *successively presented* entities could be determined accurately without counting. To some extent research by Taubman (1950) supports this possibility, although one can readily suggest some necessary control procedures to clarify Taubman's results. Furthermore, it seems reasonable that a prototype matching interpretation or a noncounting serial model analogous to Dehaene and Changeux's (1993) numerosity detector, neuronal network model could be constructed to explain numerousness judgments of temporally presented entities.

Despite its appeal and wide use, parsimony is an ill-defined criterion for choosing among alternative explanations (Thomas, 1998). However, if it is agreed that prototype matching or numerosity detector, neuronal network processes are simpler than counting, then it is reasonable to prefer the simpler process explanation when an experiment confounds the possibility of counting and noncounting explanations. In any case, unless the noncounting explanation can be excluded by the evidence, a counting process explanation cannot be used unequivocally.

Notes

Dr. Phillips is a postdoctoral research fellow in the Department of Psychiatry and Behavioral Sciences, University of Oklahoma Health Sciences Center, Oklahoma City. Dr. Young is currently a research associate in the Department of Psychiatry at Vanderbilt University, Nashville, TN. We gratefully acknowledge the assistance of Mark S. Schmidt. Please address inquiries and requests to Roger K. Thomas, Department of Psychology, University of Georgia, Athens, GA 30602–3013 (e-mail: rkthomas@arches.uga.edu). Received for publication May 29, 1997; revision received October 15, 1997.

References

American Psychological Association. (1981). Ethical principles of psychologist (revised). *American Psychologist, 36,* 633–638.

Boysen, S. T. (1993). Counting in chimpanzees: Nonhuman principles and emergent properties of numbers. In S. T. Boysen & E. J. Capaldi (Eds.), *The development of numerical competence: Human and animal models* (pp. 39–59). Hillsdale, NJ: Erlbaum.

Boysen, S. T., & Berntsen, G. G. (1989). Numerical competence in a chimpanzee (*Pan troglodytes*). *Journal of Comparative Psychology, 103,* 23–31.

Boysen, S. T., & Capaldi, E. J. (Eds.). (1993). *The development of numerical competence: Animal and human models.* Hillsdale, NJ: Erlbaum.

Davis, H., & Pérusse, R. (1988). Numerical competence in animals: Definitional issues, current evidence, and a new research agenda. *Behavioral and Brain Sciences, 11,* 561–579.

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

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

Gallistel, C. R. (1988). Counting versus subitizing versus the sense of number. *Behavioral and Brain Sciences, 11,* 585–586.

Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number.* Cambridge, MA: Harvard University Press.

Hicks, L. H. (1956). An analysis of number-concept formation in the rhesus monkey. *Journal of Comparative and Physiological Psychology, 49,* 212–218.

Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *American Journal of Psychology, 62,* 498–525.

Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General, 111,* 1–22.

Meyer, M. E. (1976). *A statistical analysis of behavior.* Belmont, CA: Wadsworth.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity to process information. *Psychological Review, 63,* 81–97.

Rosch, E. H. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111–144). New York: Academic Press.

Rumbaugh, D. M., Hopkins, W. D., Washburn, D. A., & Savage-Rumbaugh, S. (1989). Lana chimpanzee learns to count by "Nu-math": A summary of a videotaped experimental report. *Psychological Record, 39,* 459–470.

Rumbaugh, D. M., Savage-Rumbaugh, S., & Hegel, M. T. (1987). Summation in the chimpanzee (*Pan troglodytes*). *Journal of Experimental Psychology: Animal Behavior Processes, 13,* 107–115.

Rumbaugh, D. M., & Washburn, D. A. (1993). Counting by chimpanzees and ordinality judgments by macaques in video-formatted tasks. In S. T. Boysen & E. J. Capaldi (Eds.), *The development of numerical competence: Animal and human models* (pp. 87–106). Hillsdale, NJ: Erlbaum.

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill.

Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1–49). New York: Wiley.

Taubman, R. E. (1950). Studies in judged number: I. The judgment of auditory number. *The Journal of General Psychology, 43,* 167–194.

Terrell, D. F., & Thomas, R. K. (1990). Number-related discrimination and summation by squirrel monkeys (*Saimiri sciureus sciureus* and *Saimiri boliviensus boliviensus*) on the basis of the number of sides of polygons. *Journal of Comparative Psychology, 104,* 238–247.

Thomas, R. K. (1992). Primates' conceptual use of number: Ecological perspectives and psychological processes. In T. Nishida, W. C. McGrew, P. Marler, M. Pickford, & F. B. M. de Waal (Eds.), *Topics in primatology: Vol. 1, Human Origins* (pp. 305–313). Tokyo: Tokyo University Press.

Thomas, R. K. (1998). Lloyd Morgan's canon. In G. Greenberg & M. Haraway (Eds.), *Comparative psychology: A handbook* (pp. 156–163). New York: Garland.

Thomas, R. K., & Chase, L. (1980). Relative numerousness judgments by squirrel monkeys. *Bulletin of the Psychonomic Society, 16,* 79–82.

Thomas, R. K., Fowlkes, D., & Vickery, J. D. (1980). Conceptual numerousness judgments by squirrel monkeys. *American Journal of Psychology, 93,* 247–257.

Thomas, R. K., & Lorden, R. B. (1993). Numerical competence in animals: A conservative view. In S. T. Boysen & E. J. Capaldi (Eds.), *The development of numerical competence: Animal and human models* (pp. 127–147). Hillsdale, NJ: Erlbaum.