Representation of the Numerosities 1–9 by Rhesus Macaques (*Macaca mulatta*)

Elizabeth M. Brannon and Herbert S. Terrace
Columbia University

Three rhesus monkeys (*Macaca mulatta*) were trained to respond to exemplars of 1, 2, 3, and 4 in an ascending, descending, or a nonmonotonic numerical order (1→2→3→4, 4→3→2→1, 3→1→4→2). The monkeys were then tested on their ability to order pairs of the novel numerosities 5–9. In Experiment 1, all 3 monkeys ordered novel exemplars of the numerosities 1–4 in ascending or descending order. The attempt to train a nonmonotonic order (3→1→4→2) failed. In Experiment 2A, the 2 monkeys who learned the ascending numerical rule ordered pairs of the novel numerosities 5–9 on unreinforced trials. The monkey who learned the descending numerical rule failed to extrapolate the descending rule to new numerosities. In Experiment 2B all 3 monkeys ordered novel exemplars of pairs of the numerosities 5–9. Accuracy and latency of responding revealed distance and magnitude effects analogous to previous findings with human participants (R. S. Moyer & T. K. Landaeur, 1967). Collectively these studies show that monkeys represent the numerosities 1–9 on at least an ordinal scale.

It is often assumed that numerical thought is beyond the reach of animals because all human numerical systems use arbitrary symbols, (e.g., Hurford, 1987). That assumption can be questioned on both theoretical and empirical grounds (Carey, 1998; Dehaene & Changeux, 1993; Meck & Church, 1983). Recent experiments on species as diverse as pigeons, rats, raccoons, ferrets, dolphins, and monkeys provide evidence that animals have numerical abilities that do not require knowledge of symbols (see Boysen & Capaldi, 1993; Davis & Perusse, 1988; Dehaene, 1997; Dehaene, Dehaene-Lambertz, & Cohen, 1998; Gallistel & Gelman, 1992, for reviews). Indeed some laboratories have reported evidence that chimpanzees and a parrot can learn to pair arbitrary symbols with particular quantities (Boysen & Berntson, 1989; Matsuzawa, 1985; Mirofushi, 1997; Pepperberg, 1987). Other experiments suggest that primates can

sum collections of objects and symbols (Boysen & Berntson, 1989) and track additions and deletions of objects (Hauser, MacNeilage, & Ware, 1996).

The paradigms used to test the numerical abilities of animals are as varied as the species tested. Animals have been trained to respond differentially to a variety of numerically defined stimuli: the number of tones (e.g., Meck & Church, 1983; Roberts & Mitchell, 1994), the number of abstract visual stimuli (e.g., Davis, 1984; Emmerton, Lohmann, & Niemann, 1997; Hicks, 1956; Honig & Stewart, 1989; Thomas, Fowlkes, & Vickery, 1980), the number of responses to a manipulandum (Fetterman, 1993; Mechner, 1958; Platt & Johnson, 1971; Rilling, 1967), and the number of reinforcers (Olthof, Iden, & Roberts, 1997; Washburn & Rumbaugh, 1991).

The experiments described in this article address three basic issues that have been the focus of the growing literature on the numerical abilities of animals: the degree to which animals can discriminate number when nonnumerical cues are eliminated, the salience of number as a stimulus dimension, and how animals represent number.

Elizabeth M. Brannon, Department of Psychology, Columbia University; Herbert S. Terrace, Departments of Psychology and Psychiatry, Columbia University.

Correspondence concerning this article should be addressed to Elizabeth M. Brannon or Herbert S. Terrace, Department of Psychology, Columbia University, 1190 Amsterdam Avenue, New York, New York 10027. Electronic mail may be sent to liz@psych.columbia.edu or terrace@columbia.edu.

Controlling for Nonnumerical Factors

To conclude that an animal can discriminate number, one must rule out control by other stimulus dimensions (e.g., area, time, or hedonic value). This has rarely been done in experiments on the numerical ability of animals. For example, when stimuli are presented simultaneously, surface area or density covaries with numerosity (e.g., Honig & Stewart, 1989), and when stimuli are presented successively at a constant rate, time covaries with numerosity (e.g., Rilling, 1967). When food is used as a discriminitive stimulus, hedonic value covaries with numerosity (Washburn & Rumbaugh, 1991).

The prodigious memory animals have for particular stimuli raises another important methodological issue. In some circumstances, animals solve complex problems by memorizing a large number of stimuli rather than by learning a "simple" cognitive rule. For example, Farrar (1967) trained chimpanzees on a matching-to-sample task in which the sample was presented with three distractors. The chimpanzees were trained to a criterion of 90% and were then tested in the absence of the sample. Surprisingly, performance was not significantly disrupted. Thus, instead of learning an identity matching rule, the chimpanzees had learned 17 conditional discriminations. Similarly, Greene (1983) showed that pigeons could memorize hundreds of distinct photographs on a pseudoconcept formation task. The only way to rule out memory of a specific set of stimuli on a conceptual task is to present first-trial data to show that a subject's performance is not disrupted by the introduction of novel stimuli.

The Salience of Numerical Stimuli

Capaldi and Miller (1988), Gallistel and Gelman (1992), and Dehaene (1997) argued that animals perceive number naturally and that they make numerical computations in daily foraging decisions. By contrast, Davis and his colleagues claimed that animals rely on number only as a strategy of "last resort," that is, when there is no alternative solution available for the task at hand (Davis & Memmott, 1982; Davis & Perusse, 1988).

An influential experiment by Meck and Church (1983) provided evidence against the last-resort hypothesis by training rats to discriminate numerosity when it was confounded with time. Subjects were trained on a successive discrimination problem in which they were given a choice of two levers. They were rewarded for pressing Lever A on trials on which two tones were presented and for pressing Lever B on trials on which eight tones were presented. Both stimuli consisted of alternating tone-on and tone-off periods, each lasting 500 ms. Subjects could therefore learn the discrimination on the basis of stimulus duration (2 vs. 8 s), stimulus frequency (two vs. eight tones), or both stimulus dimensions.

After learning to discriminate the short two-tone stimulus from the long eight-tone stimulus, the rats in Meck and Church's (1983) experiment were given a generalization test to determine which feature(s) of the stimulus controlled their performance. Two types of test stimuli were used. For one type, the duration of all stimuli was held constant at 4 s, and the number of tones was varied from trial to trial (two, three, four, six, or eight tones). For the other, the number of tones was held constant at four, and the duration of the sequence was varied (2, 3, 4, 6, and 8 s).

The rats responded appropriately when the test stimulus matched the training stimulus in either duration or frequency. For example, they chose Lever A when the 4-s test stimulus contained two tones and Lever B when the 4-s test stimulus contained eight tones. The critical finding of the Meck and Church (1983) experiment was that the rats generalized equally on the basis of duration and number

even when the duration or the frequency of the test stimuli did not match the values of the training stimuli. For example, the likelihood that the rat would choose Lever A following a three-tone stimulus of 4-s duration was the same as the likelihood that it would choose Lever A following a 3-s stimulus consisting of four tones.

The results of Meck and Church's (1983) experiment would appear to contradict the last-resort hypothesis in that subjects did not require any special training to respond to test stimuli on the basis of their frequency. However, a failure to replicate the performance of Meck and Church's rats when other dimensions of the compound stimuli used during training were varied elucidates the difficulty of defeating the last-resort hypothesis. Like Meck and Church (1983), Breukelaar and Dalrymple-Alford (1998) trained rats on a discrimination task on which number and time were confounded. However, Breukelaar and Dalrymple-Alford obtained no evidence of control by number in a generalization test in which they varied the temporal pattern of the tones and held signal duration constant. As in Meck and Church's study, the likelihood of choosing a particular lever varied as a function of the duration of the signal when the number of tones was held constant. By contrast, the likelihood of choosing Levers A and B did not vary as the number of test stimuli was varied between two and eight.

Ordinal Versus Nominal Representation of Number

Once control by the numerical value of a stimulus has been demonstrated it is possible to ask, how does an animal represent number? Do animals simply form nominal categories for different quantities or do they also represent numerosity on an ordinal scale? For example, having learned to assign collections of four elements to Category A, collections of five elements to Category B and collections of two elements to Category C, does an animal know that Category B is greater than Category C? To answer that question subjects have to be tested on their ability to order novel numerosities after learning to discriminate two or more numerical stimuli.

One commonly used method for investigating stimulus control is to give subjects a generalization test with novel stimuli from the dimension in question (Terrace, 1966). That method is illustrated by the study described earlier by Meck and Church (1983). After rats were trained to respond to one lever following the presentation of two tones and another lever following the presentation of eight tones, they were tested on intermediate values (two, three, four, five, and six tones). Subjects responded to the test stimuli in a graded manner, suggesting that numerosity was the underlying dimension. However, as mentioned earlier, the performance of Meck and Church's (1983) subjects appears to have been influenced by the temporal rather than by the numerical properties of the discriminative stimuli (Breukelaar & Dalrymple-Alford, 1998). It should also be noted that all of the novel values in Meck and Church's experiment fell within the range defined by the training values. It is unclear whether an animal would respond appropriately to stimuli outside the training range.

Using a different paradigm for assessing ordinal knowledge, Thomas et al. (1980) trained monkeys to discriminate random dot patterns containing different numbers of dots on a simultaneous discrimination task. Pairs differing in the number of dots they contained were presented simultaneously. Subjects had to choose the stimulus containing the smaller numerosity. Training began on the pairs 1–2 and then shifted progressively to larger numerosities (2–3, 3–4, 4–5, 5–6, 6–7 and 7–8). Thomas et al. controlled for surface area by varying the size of the dots across trials. Subjects nevertheless discriminated patterns containing 7 and 8 dots suggesting that they may have learned to order stimuli with respect to numerosity. However, as Thomas et al. acknowledged, the design of their study did not allow them to rule out the possibility that their monkeys learned a series of pairwise associations based on nominal categories (e.g., choose 5 when paired with 6, and choose 7 when paired with 8). To demonstrate that monkeys represent number as an ordered dimension it is necessary to show that subjects could order novel numerical pairs the first time they are presented.

The issue of ordinal knowledge of numerosity and the last-resort hypothesis are the focus of the two experiments described in this article. A brief description of some of the findings of each experiment was presented in an article by Brannon and Terrace (1998). One experiment from that article showed that rhesus monkeys learned an ascending rule (1 → 2 → 3 → 4) for responding to the numerosities 1–4 on a task that could have been mastered by memorizing 35 sets of stimuli. A second experiment from that article showed that the monkeys, who learned the ascending rule for responding to the numerosities 1–4, spontaneously ordered pairs of the numerosities 5–9. In this article we describe new data showing that monkeys can also learn a descending rule (4 → 3 → 2 → 1) but not a nonmonotonic rule (3 → 1 → 4 → 2). We also describe psychophysical functions that show the influence of numerical distance and magnitude on latency and accuracy of responding.

Experiment 1

The task used in Experiment 1 differs from previous tasks used to assess animal numerical competence in that subjects were trained to respond in a fixed order to four numerical stimuli (1–4) and then tested with novel exemplars of those numerosities. The purpose of Experiment 1 was to compare a monkey's ability to respond to numerical stimuli in an ascending, descending, and a nonmonotonic order.

Method

Subjects

The subjects were three 3-year-old rhesus monkeys (*Macaca mulatta*), Rosencrantz, Benedict, and Macduff. They were housed in a rhesus colony room with 7 other monkeys who participated in similar experiments. Monkeys were housed in pairs in two adjoining cages (each measuring 0.74 deep × 0.61 wide × 0.89 high m). During experimental sessions an opaque divider separated the cages. Subjects were fed daily between 1300 and 1400 (Purina

monkey chow [Ralston Purina, Richmond, IN], fruit). Water was available ad libitum.

Apparatus

Experimental sessions, which lasted approximately 40 min, took place in each subject's home cage, typically between 1000 and 1300. A mobile cart, which housed a Microtouch touch-sensitive 15-in. video monitor and a Gerbrands pellet dispenser, was positioned in front of the subject's cage before each session. The guillotine door in front of the cage was raised after the cart was secured to provide the subject with unimpeded access to the monitor. Subjects were unrestrained during testing. The experimenter, who had no contact with monkeys during testing, could observe each session by remote video from an observation room located 10 m from the colony room. A Power Macintosh computer using PsyScope software (Cohen, MacWhinney, Flatt, & Provost, 1993) controlled experimental events and data collection. Reinforcers were 190-mg Noyes (Noyes, Lancaster, NH) pellets (banana, orange, or grape flavored).

Previous Training

We pretrained monkeys by the simultaneous chaining paradigm on a list-production task (Terrace, 1984) to provide them with experience in producing 4-item lists. List training began immediately after magazine training and shaping. For each list, all of the items were presented simultaneously and continuously on the touch-sensitive video monitor in configurations that varied randomly from trial to trial. We varied the configuration of the items to prevent subjects from relying on a fixed motor pattern of responses to execute the sequence. List items were digitized color photographs (3.5 × 3.5 cm) of human-made objects. The subject's task was to touch the photographs in a particular order to receive a reward (e.g., car → motorcycle → table → building). Each correct response was followed by brief (100 ms) auditory and visual feedback to indicate that the response had been detected by the monitor. Feedback for correct responses conveyed no information as to the subject's progress in the sequence. For example, in the sequence described earlier, a response to the car provided no information that the next item the subject should touch was the motorcycle (as opposed to the table or the building). Accordingly, subjects had to keep track of their place in the sequence after responding to each item.

Any error terminated the trial immediately and resulted in a 15-s time-out (TO) during which the screen of the video monitor was darkened. A food pellet was dispensed only after the subject touched all four stimuli in the correct order. The intertrial interval (ITI) varied from 5 to 13 s with a mean of 8 s. Sixty trials were given in each session.

Given the conservative assumption that a subject did not return to a previously selected item, the probability of responding correctly by chance to the first item of a 4-item sequence is .25; to the second item, .33; to the third item, .5; and to the fourth item, 1.0. Thus, the chance probability of responding correctly to all four stimuli was .25 × .33 × .5 × 1 = .04.[1] Subjects were trained on seven 3-item and eleven 4-item lists of photographs. A detailed description of subjects' performance on these lists can be found in Terrace, Son, and Brannon (2000).

---

[1] If we allowed for backward errors, chance performance would be lower (.25 × .33 × .33 × .33 = .009). Both calculations ignore repeat responses because they had no consequence (e.g., *ABBBBBCD*).
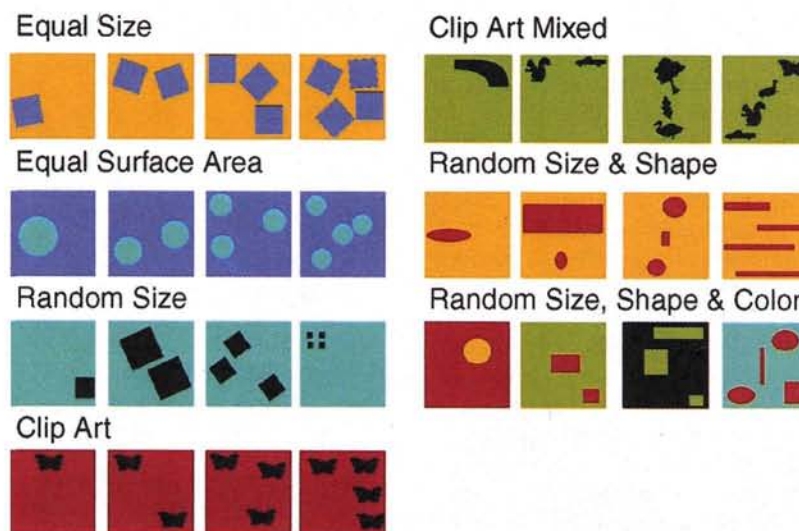
Figure 1. Exemplars of the seven different types of stimulus sets. All types were used with equal frequency in both 4-item training and 4-item testing. *Equal size:* Elements were of same size and shape. *Equal area:* Cumulative area of elements was equal. *Random size:* Element size varied randomly across stimuli. *Clip art:* Identical nongeometric elements selected from clip art software. *Clip art mixed:* Clip art elements of variable shape. *Random size and shape:* Elements within a stimulus were varied randomly in size and shape. *Random size, shape, and color:* Same as previous with background and foreground colors varied between stimuli.

Stimuli

Exemplars of numerosity were composed of stimuli that contained one, two, three or four abstract elements. Examples of the stimuli are shown in Figure 1. Elements, which were circles, ellipses, squares, rectangles, or complex clip art figures, were positioned randomly within each stimulus. The color of each element was red, blue, green, cyan, yellow, or black. The background color was selected from one of the five remaining colors. Clip art elements were gray or black. All of the 30 possible foreground–background color combinations were used with approximately equal frequency. The stimuli were generated by a graphics program.

A stimulus set contained one exemplar of each of the four numerosities. Seven types of stimulus sets were used to minimize the possibility of stimulus control by nonnumerical dimensions. These were equal size, equal surface area, random size, clip art, clip art mixed, random size and shape, random size, shape, and color. Examples of each type of stimulus set are shown in Figure 1.

For the equal size sets, the elements of all four stimuli were of identical size and shape. For the equal surface area sets, the sum of the elements' area within each stimulus was equal for each of the four stimuli within a set. Subjects could infer stimulus order on sets in which stimuli were of equal size by applying an increasing surface area rule. Similarly, they could infer stimulus order on sets in which the surface area of the elements were equal by applying a decreasing size rule. Those bases for inferring order were eliminated in the random size stimulus sets for which the size of the elements was selected at random.

Clip art stimulus sets contained stimuli with a single clip art shape (one leaf, two leaves, etc.). We used both "natural" objects (e.g., dogs, leaves) and human-made objects (e.g., heart shapes, cars). Mixed clip art sets were composed of stimuli that contained more than one shape (e.g., an exemplar of two might contain a dog and a cat). For the random size and shape sets, the size and shape of

the elements varied randomly. Similarly, for the random size, shape, and color sets, the size and shape of the elements varied randomly within each stimulus as did the background and foreground colors.

Stimuli, which were each 3.5 × 3.5 cm, were programmed to appear on the video monitor in one of 16 positions, each equidistant from one another. On each trial the stimuli were displayed in a novel configuration that was selected at random from 43,680 possible configurations. Sample configurations are shown in Figure 2.

Task and Procedure

The task used was identical to that described for pretraining except that we used exemplars of the numerosities 1–4 instead of photographic stimuli. Subjects were required to respond to each stimulus on the video monitor, in a specific order. A particular
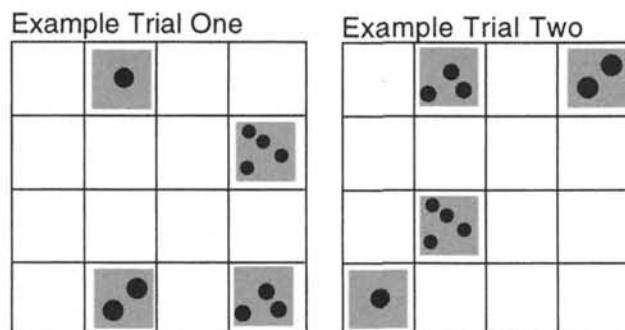
Figure 2. Example configurations. On each trial the four stimuli were displayed in a random spatial configuration. Sixteen spatial locations allowed for 43,680 possible configurations.
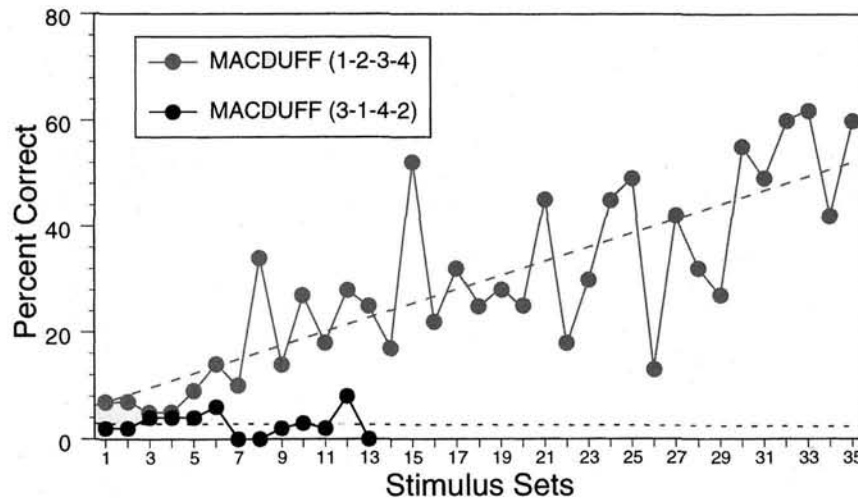
Figure 3. Comparison of 1 monkey's (Macduff) performance on 13 nonmonotonic stimulus sets $(3 \rightarrow 1 \rightarrow 4 \rightarrow 2)$ and 35 monotonic stimulus sets $(1 \rightarrow 2 \rightarrow 3 \rightarrow 4)$. Performance did not improve across the 13 nonmonotonic sets but quickly improved once he was switched to a monotonic rule. The dashed lines reflect the best-fit linear models.

numerosity was assigned to each ordinal position in the response sequence for each monkey. For Rosencrantz, the order was ascending $(1 \rightarrow 2 \rightarrow 3 \rightarrow 4)$; for Benedict, it was descending $(4 \rightarrow 3 \rightarrow 2 \rightarrow 1)$. Macduff was initially trained on a nonmonotonic order $(3 \rightarrow 1 \rightarrow 4 \rightarrow 2)$. After showing no evidence of learning the nonmonotonic rule during training on 13 stimulus sets, he was shifted to the ascending rule $(1 \rightarrow 2 \rightarrow 3 \rightarrow 4)$.

The experiment was carried out in three phases: training on individual lists, multilist training, and transfer. During the training phase, subjects were trained on 35 different stimulus sets. The purpose of the training phase was to provide subjects with a basis for learning the relationship between numerosity and ordinal position in the required response sequence (rather than to overtrain them on any given stimulus set). Each set was trained until subjects correctly ordered the stimuli on 20% of the trials in a single session (12/60 trials) or until they completed three sessions. New stimulus sets were always introduced at the start of a session.

The purpose of multilist training was to prepare subjects for the transfer sessions in which a *novel* stimulus set was presented on each trial. During multilist training, the number of stimulus sets used during each session was gradually increased. During the first five sessions, subjects were retrained on 15 of the original 35 stimulus sets. Three stimulus sets were presented during each session, in separate 20-trial blocks. The same 15 stimulus sets were subsequently presented in a random order within the same session. Training continued until the subjects responded correctly on 70% of the trials within a single session or until eight sessions of training on the randomly alternating stimulus sets were completed. During the final phase of multilist training, all 35 of the original stimulus sets were presented in a random order during each session. Training continued until subjects executed 60% of the trials correctly within a single session or until they completed six sessions.

The purpose of transfer sessions was to determine whether monkeys could respond in the correct sequence to stimuli from novel stimulus sets solely on the basis of numerosity. We presented 150 novel stimulus sets over the course of five successive sessions, 30 novel stimulus sets per session. During the first half of each session, 30 novel stimulus sets were presented in a random order. The same 30 stimulus sets were presented in a different random

order during the second half of the session, but we used only first-trial data to assess transfer performance. Sets from each of the seven stimulus categories were presented during each session with approximately equal frequency. During the transfer test, the consequences of correct and incorrect responses were the same as those that were in effect during the training sessions. Correct responses produced brief auditory and visual feedback, errors terminated a trial and a complete sequence of responses produced food reward.

Results and Discussion

All subjects trained on monotonic sequences (ascending or descending) became progressively more efficient at ordering exemplars of the numerosities 1–4. Although the attempt to train 1 monkey to learn a nonmonotonic sequence was unsuccessful, that subject subsequently mastered the ascending monotonic sequence. All subjects also responded as accurately during the transfer tests as they did on the last block of training sets. The absence of a decrement in accuracy during the transfer test, which consisted of 150 trial-unique stimulus sets, provides unequivocal evidence that subjects learned a numerical rule.

Training

Nonmonotonic sequences. Macduff, who was trained to respond to exemplars of numerosity in the nonmonotonic sequence $3 \rightarrow 1 \rightarrow 4 \rightarrow 2$ showed no evidence of learning. After he was switched to stimulus sets in which he was required to respond in an ascending order, Macduff's performance improved rapidly and was indistinguishable from that of the other subjects. Figure 3 shows Macduff's performance on the 13 stimulus sets trained by the nonmonotonic rule and on the 35 stimulus sets trained by the
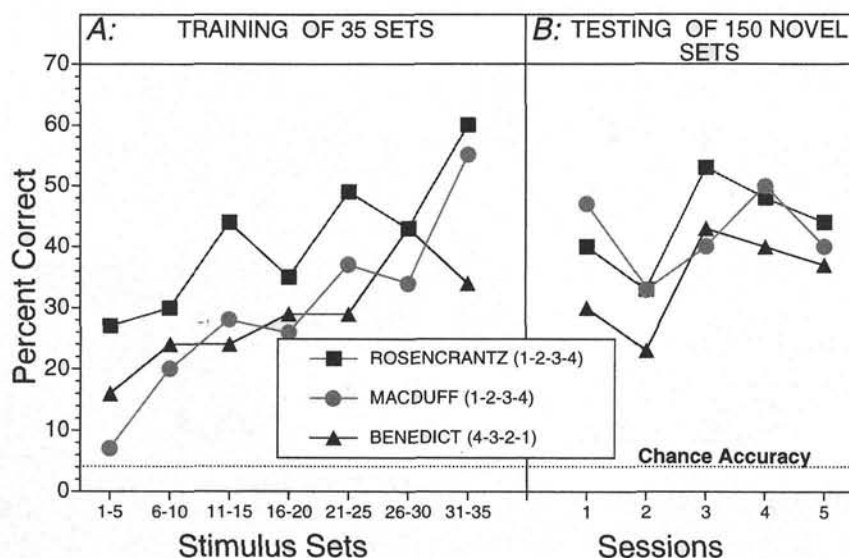
Figure 4. (A) Percentage of correctly completed trials during the first session for each of 35 training stimulus sets in blocks of five sessions. (B) Percentage of correctly completed trials on the 150 test sets.

ascending rule. His first-session accuracy did not exceed 5% on any of the nonmonotonic stimulus sets.[2]

That Macduff was unable to learn the nonmonotonic sequence 3 → 1 → 4 → 2 stands in marked contrast to the ease with which he acquired the monotonic sequence 1 → 2 → 3 → 4. His difficulty with the 3 → 1 → 4 → 2 sequence suggests that monkeys can perceive ordinal relations between numerical stimuli without specific training. It also raises the interesting question, could a monkey ever learn a 3 → 1 → 4 → 2 sequence (or other nonmonotonic sequences)? It would also be of interest to determine the relative difficulty of similar sequences for young children.

Monotonic sequences. Accuracy of responding on the 35 monotonic training sets increased progressively as subjects were exposed to new sets. The percentage of correct trials during the first session of training on a new list is shown in Figure 4A for each monkey. In each instance, accuracy was well above the level expected by chance, for Rosencrantz, $t(34) = 11.93$, $p < .01$; for Macduff, $t(34) = 8.80$, $p < .01$; for Benedict, $t(34) = 9.44$, $p < .01$. The slopes of each of the regression lines shown in Figure 4A were also significant (Rosencrantz, $r^2 = .25$, $p < .01$; Macduff, $r^2 = .63$, $p < .01$; Benedict, $r^2 = .26$, $p < .01$).

The increase in first-session accuracy shown in Figure 4A could be attributed to knowledge of a numerical rule, to growing expertise in learning arbitrary sequences (Harlow, 1949; Terrace et al., 2000), or to both factors. Because subjects were trained for at least 60 trials on each stimulus set, they could have developed expertise in associating some nonnumerical feature of each stimulus with its ordinal position, for example, the configuration of the elements within each stimulus. That explanation of the increase in first-session accuracy was ruled out by analyses of correct and partially correct sequences on the *first trial* of each new set.

The probability of responding correctly by chance to all four stimuli when presented with a new set of stimuli is .04.

The only way a subject could exceed that level of accuracy on the first trial of a new stimulus set was to use a numerical rule. Two of the 3 monkeys exceeded that level of accuracy on the first trial of training on new lists (Rosencrantz, 17%; Benedict, 20%). The conditional probabilities of responding correctly by chance to the first, second, third, and fourth numerosities are, respectively, .25, .33, .5, and 1. Three subjects exceeded the chance level of accuracy with their first response to A (Rosencrantz, 60%; Macduff, 46%; Benedict, 51%). Two subjects also exceeded the chance level of accuracy with their first response to B (Rosencrantz, 43%; Benedict, 78%). One subject exceeded the chance level of accuracy with his first response to C (Rosencrantz, 67%).

Multistimulus set training. Subjects' reliance on a numerical rule to order stimuli was also evident in their performance during sessions in which more than one stimulus set was presented. As can be seen in Table 1, the increase from 1 to 35 sets had no discernable effect on accuracy. Other than using a numerical rule, the only explanation of the high level of accuracy during training with multiple stimulus sets is that subjects remembered the order in which to respond to each stimulus of each of the 35 stimulus sets. This "brute memory" explanation is unlikely. A marked decrease in accuracy occurs when monkeys are trained to respond to more than one 4-item list of arbitrary items (photographs) during the same session (Swartz, Chen, & Terrace, 1991). That decrease is not surprising. Responding correctly to multiple lists of arbitrary items increases memory load. In contrast, responding correctly to multiple

---

[2] We occasionally allowed Macduff to exceed the three-session maximum for training on each set to see if he could master any nonmonotonic set. Although his performance improved slightly with extended training, his first session performance never improved.

Table 1
Percentage Correct on Blocks of Multiset Training Sessions

	Phase								
	Rosencrantz			Macduff			Benedict		
Session	1	2	3	1	2	3	1	2	3
1	62	55	50	38	62	43	52	38	53
2	50	63	57	53	52	45	22	37	40
3	78	68	63	53	57	50	53	43	33
4	52	63	60	50	57	37	38	49	35
5	62	75	57	51	65	28	40	46	52
6			50		64	40		45	38
7					53			45	
8					60			57	
Mean	61	65	56	49	59	41	41	45	42

Note. In Phase 1 of multilist training, 15 of the 35 training stimulus sets were retrained. Three sets were presented in each session in blocks of 20 trials each. In Phase 2, the same 15 sets were presented in a random order within each session. In Phase 3, all 35 training sets were presented in random order in each session.

lists of numerical items does not increase memory load if a numerical rule is used.

Testing

Each subject's performance on the transfer test provides unequivocal evidence that monkeys can detect the numerosities 1–4 and that they deduced the ascending or descending rule that was common to the training sets. Despite the abrupt shift from familiar to novel stimulus sets, accuracy of responding to the novel sets remained at the same level as it was during the last block of training. Accuracy of responding to the novel sets is shown in Figure 4B. A repeated measures analysis of variance (ANOVA) comparing the last five blocks of training to the five transfer sessions (Subject × Type of List) revealed no main effects, subject, $F(2, 11) = 3.69, p > .06$; training versus novel lists, $F(1, 2) = .62, p > .45$, and no interaction between those factors, $F(2, 11) = .84, p > .46$. Performance on partially correct transfer trials provided additional evidence that subjects detected the numerosities of the novel stimuli. Each monkey responded at greater-than-chance levels of accuracy at each transition in the sequence. Table 2 summarizes the values of the conditional probabilities at each transition on the 150 novel sets. One-sample t tests showed that the value of each conditional probability exceeded the value predicted by chance. The results of the t tests are summarized in Table 3.

Table 2
Percentage Correct at Each Transition on 150 Test Trials

Monkey	1 (4)	2 (3)	3 (4)	4 (1)
Rosencrantz	81.00	79.00	67.00	99.00
Macduff	69.00	69.00	76.00	96.00
Benedict	86.00	70.00	72.00	100.00
Mean	78.67	72.67	71.67	98.33
Chance	25.00	33.00	50.00	100.00

Note. Headings in parentheses are the values for Benedict, who was required to respond in descending order.

Table 3
t Tests Comparing the Percentage Correct at Each Transition to Chance Expectations

	A		B–A		C–AB	
Monkey	$t(4)$	$p<$	$t(4)$	$p<$	$t(4)$	$p<$
Rosencrantz	22.79	.01	24.35	.01	4.80	.01
Macduff	29.25	.01	15.09	.01	3.30	.03
Benedict	16.10	.01	13.12	.01	3.97	.02

Note. Percentage correct = 25%, 33%, and 50% for A, B–A, and C–AB, respectively. An average for each of the five transfer sessions was used for each monkey.

With but one exception, accuracy of responding to each of the seven stimulus classes exceeded the chance level of accuracy for all 3 monkeys. The exception was Benedict's performance on stimulus sets composed of heterogeneous clip art. Accuracy of responding on each of the seven stimulus classes and the rank orders of accuracy for each monkey are shown in Table 4. The heterogeneous classes were generally more difficult than the homogeneous classes. A two-way ANOVA (Subject × Stimulus type) revealed a significant main effect of stimulus type (homogeneity vs. heterogeneity), $F(1, 2) = 13.3, p < .01$, but no main effect of subject and no interaction between stimulus type and subject. The differential influence of stimulus class on the discriminability of numerosities is an interesting phenomenon that deserves further study. It should not, however, detract from the fact that subjects were able to order the numerosities 1–4, even those represented by the most difficult stimulus classes.

Experiment 2A

The results of Experiment 1 provide clear evidence that rhesus monkeys can order novel exemplars of the numerosities 1–4 in an ascending or a descending order. That ability suggests that monkeys can also perceive the ordinal values of the numerosities. It is possible, however, to explain the results of Experiment 1, without appeal to knowledge of ordinal numerical relations. The monkeys could have discriminated exemplars of the four numerosities as nominal categories (A, B, C, and D) and learned to order them as they might the stimuli of an arbitrary sequence (A → B → C → D). Experiments on sequence learning have shown that monkeys readily learn to execute arbitrary sequences composed of photographs of natural objects, for example, cat → rose → apple → beetle (Swartz et al., 1991) or geometric stimuli (D'Amato & Colombo, 1988). Indeed, prior to their training on numerical sequences, each of the subjects of Experiment 1 learned seven 3-item and eleven 4-item lists, each composed of novel photographs. It is therefore conceivable that Rosencrantz, Benedict, and Macduff learned to assign exemplars of each numerosity to a nominal category (e.g., A = stimuli with 3 elements, B = stimuli with 4 elements, C = stimuli with 2 elements and D = stimuli with 1 element) and then learned to respond to these categories in the arbitrary orders (D → C → A → B for "ascending" sequences and B → A → C → D for "de-

Table 4
Percentage Correct and Rank Orders for Each Stimulus Class in the Transfer Sessions of Experiment 1

Stimulus class	Rosencrantz		Benedict		Macduff	
	% Correct	Rank	% Correct	Rank	% Correct	Rank
Equal size	60	1	48	3	57	2
Equal surface area	56	2	66	1	59	1
Random size	50	3	31	4	23	6
Heterogeneous clip art	47	4	22	5	48	4
Homogeneous clip art	31	5	54	2	50	3
Heterogeneous random size and color	28	6	13	7	19	7
Heterogeneous random size	22	7	19	6	38	5

scending" sequences). Solving the task in this fashion would constitute evidence of a *nominal* numerical rule.

The purpose of Experiment 2 was to provide evidence that the subjects who served in Experiment 1 had in fact acquired *ordinal* numerical rules (ascending or descending) for responding to numerically defined stimuli, rather than a specific rule for ordering nominally defined exemplars of the numerosities 1–4. This was done by testing Rosencrantz, Benedict, and Macduff on their ability to order the novel numerosities 5–9. If the numerosities 1–4 functioned as exemplars of nominal categories in Experiment 1, then subjects would be expected to respond at chance levels of accuracy to exemplars of the numerosities 5–9.

We tested Rosencrantz, Macduff, and Benedict on their ability to order the 36 pairs of numerosities that can be derived from the values 1–9. All 36 combinations of those numerosities were tested: 6 *familiar–familiar* pairs drawn from the familiar numerosities 1–4, 20 *familiar–novel* pairs composed of one familiar and one novel numerosity, and 10 *novel–novel* pairs drawn from the novel numerosities 5–9. No differential feedback was given on trials on which novel–familiar or novel–novel pairs were presented. Accurate performance on novel–novel trials cannot be attributed to *learning* ordinal relations between those numerosities. Accordingly, accurate responding on novel–novel trials would show that monkeys can extrapolate an ascending or descending ordinal rule outside the range of numerosities on which they were originally trained.

Method

Subjects and Apparatus

The subjects of Experiment 1 were the subjects of Experiment 2 (Rosencrantz, Benedict, and Macduff). The apparatus was also the same as that used in Experiment 1.

Stimuli

We constructed exemplars of the numerosities 1–9 as in Experiment 1 with the following modifications. Only circles, ellipses, squares, and rectangles were used as elements. Neither heterogeneous stimuli nor clip art shapes were used. The elements of each stimulus were homogeneous with respect to size, shape, and color. The same six colors used to construct stimuli in Experiment 1 were used in Experiment 2A. However, a few color combinations that seemed difficult to discriminate to the human eye were eliminated

(black and blue, cyan and yellow, green and yellow, green and cyan). To minimize nonnumerical differences, the shape of the elements, their color, and the background color were identical for each pair of stimuli. Elements were positioned randomly within each stimulus.

To prevent subjects from using surface area as a cue, the total area of the elements was smaller for the larger numerosity for half of the stimuli and larger for the other half. Examples of the stimuli used in Experiment 2A are shown in Figure 5. Novel exemplars of each of the 36 numerosity pairs were used on each trial during each of the 20 sessions of Experiment 2A. We used a total of 1,680 stimuli in Experiment 2A.

Task and Procedure

Two stimuli, each of a different numerosity, were presented on each trial. As in Experiment 1, the spatial configuration of the stimuli on the video monitor was varied randomly from trial to trial within a 4×4 matrix. The task was to touch the stimuli in the same ascending or descending numerical order that the subject learned in Experiment 1 (Rosencrantz and Macduff ascending; Benedict descending). The ITI was the same as that used in Experiment 1.

Differential reinforcement was provided only on trials on which *both* stimuli were familiar. For those six pairs (1–2, 1–3, 1–4, 2–3, 2–4, 3–4), a correct sequence produced a 190-mg pellet; an incorrect response, an 8-s TO. Neither food pellets nor TOs were provided on trials on which any of the novel numerosities were presented. Instead, brief visual and auditory feedback followed a response to *each* numerosity regardless of the order in which the monkey responded. Figure 6 provides a summary of the reinforcement contingencies for each of the 36 pairs.

Each session consisted of 90 trials. The relative frequency of trials on which each type of stimulus pair was presented was adjusted so that reinforcement would be available on two thirds of the trials. Each of the six familiar–familiar pairs (1–2, 1–3, 1–4, 2–3, 2–4, 3–4) was presented 10 times per session (60 trials). The other 30 numerosity combinations were presented only once per session (30 trials).

Results and Discussion

Both monkeys who learned the ascending rule in Experiment 1 (Rosencrantz and Macduff) correctly ordered approximately 75% of the pairs of the novel numerosities 5–9. Benedict, who learned the descending rule, performed at a chance level of accuracy on the novel–novel comparisons. Rosencrantz's and Macduff's performance shows that rhesus monkeys do not require explicit training to discriminate
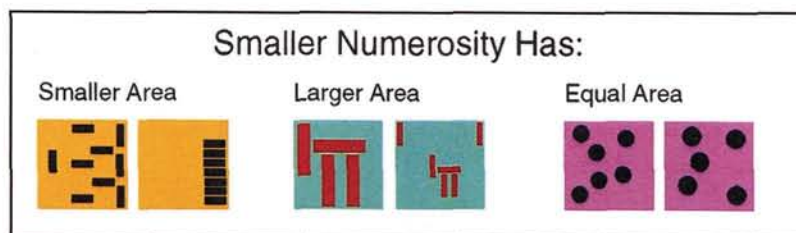
Figure 5. Examples of stimulus sets used in Experiment 2. In Experiment 2A the smaller numerosity had a larger cumulative surface area than the larger numerosity on 50% of all trials, and elements within each stimulus were identical in size, shape, and color. In Experiment 2B a third stimulus type was added; the smaller and larger numerosity had equal cumulative surface area.

novel numerosities and that they perceive ordinal relations between untrained numerosities.

Figure 7 shows each monkey's performance on trials with familiar–familiar, familiar–novel, and novel–novel pairs. Performance on novel–novel pairs was lower than performance on familiar–familiar pairs for all subjects, Rosencrantz, $t(19) = 7.76$, $p < .001$; Macduff, $t(19) = 4.26$, $p < .001$; and Benedict, $t(19) = 10.5$, $p < .001$. A two-way ANOVA (Subject × Condition) revealed a significant main effect of subject, $F(2, 4) = 58.08$, $p < .001$, a significant main effect of condition (familiar–familiar, familiar–novel, novel–novel), $F(2, 4) = 99.4$, $p < .001$, and a significant

interaction, $F(4, 171) = 3.9$, $p < .01$. The interaction reflects Rosencrantz's and Macduff's uniformly high performance on familiar–familiar and familiar–novel pairs and lower performance on novel–novel pairs, and Benedict's graded performance (highest levels on familiar–familiar pairs and lowest levels on novel–novel pairs). A Fisher's post hoc test revealed that the main effect of subject resulted from Benedict's poor performance, as compared with Rosencrantz and Macduff ($p < .01$). Fisher's post hoc tests also revealed that performance was significantly worse on novel–novel pairs than on familiar–novel and familiar–familiar pairs ($p < .01$).
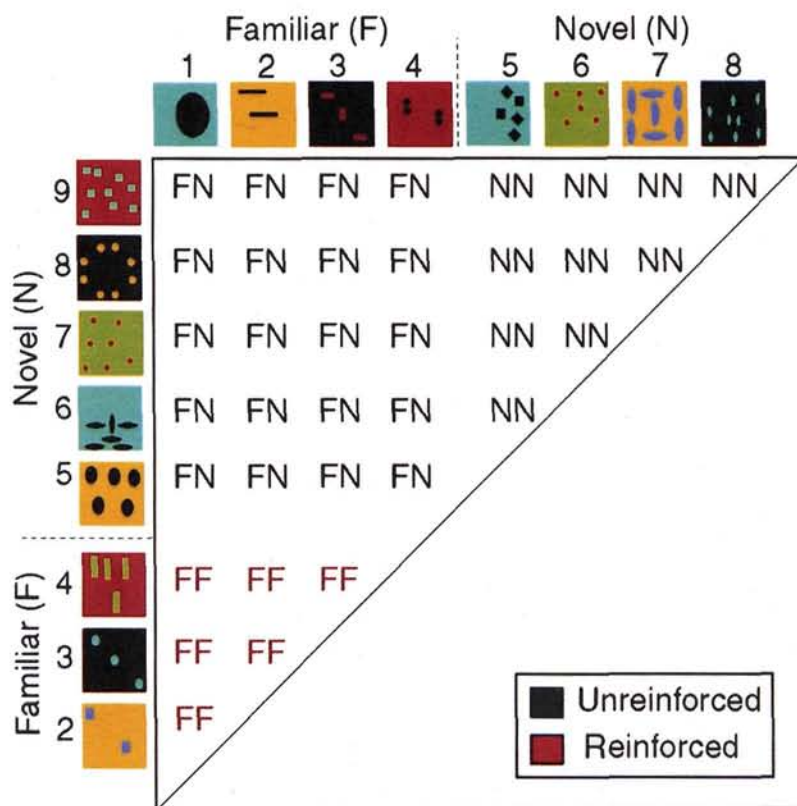

Figure 6. Diagram of the 36 pairs of the numerosities 1–9 used in Experiment 2A. These are segregated into three types that were defined with respect to the subjects' prior experience with the constituent numerosities: familiar–familiar (FF), familiar–novel (FN), novel–novel (NN). Only the red entries (FF) were reinforced.
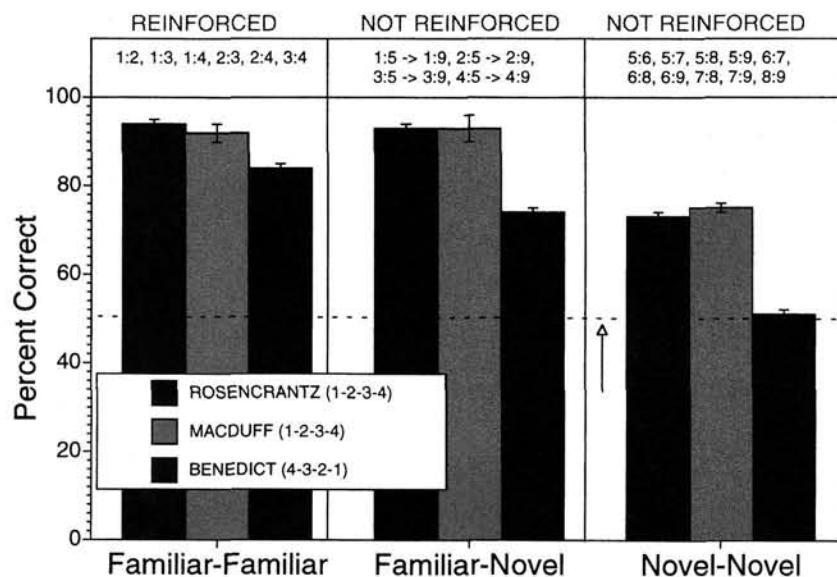
Figure 7. Performance on familiar–familiar, familiar–novel, and novel–novel numerosity pairs in Experiment 2A. Only familiar–familiar pairs were reinforced.

Performance as a Function of Numerosity

Subjects' performance on the 10 novel–novel pairs was idiosyncratic. This is not surprising given that there are only 20 data points for each monkey for each of the novel–novel numerosity pairs. Each subject's performance on each numerical comparison is summarized in Table 5. Rosencrantz and Macduff, who learned the ascending rule, exceeded the chance level of accuracy on all pairs except 6–8, 7–8, and 8–9.[3] In marked contrast, Benedict, the monkey who learned the descending rule, failed to exceed the chance level of accuracy on many of the 36 numerosity comparisons, including all but one of the novel–novel numerosity pairs. Anomalously, Benedict exceeded the chance level of accuracy on the 8–9 novel–novel numerosity pair.

Preliminary analyses suggested systematic effects of distance and magnitude on accuracy and latency. Because of the small number of observations, we reserve discussion of these effects until Experiment 2B, where we draw on a larger data set.

Benedict's failure to extrapolate the descending rule to novel numerosities should be interpreted in the context of his impaired performance relative to the other 2 monkeys on all pair types (familiar–familiar, familiar–novel, and novel–novel). His poor performance on the numerical comparison task could reflect individual difference or the asymmetry in the requirements of extrapolating the ascending and descending rules. For familiar–novel pairs, the ascending rule required an initial response to a familiar numerosity, whereas the descending rule required the subject to avoid a familiar numerosity and respond to a novel numerosity first. There was, however, no evidence of a bias for Benedict to respond first to the numerosity 4 on familiar–novel trials. Additional research is needed to clarify whether there are reliable differences in the manner in which monkeys learn ascending and descending sequences.

Stimulus Controls

The surface area of the elements had no effect on accuracy. When the smaller numerosity was depicted by a smaller surface area, accuracy of responding for Rosencrantz, Macduff, and Benedict was, respectively, 92%, 91%, and 79%; when it had a larger surface area, accuracy levels were, respectively, 91%, 89%, and 76%. A repeated measures ANOVA (Stimulus Type × Subject) showed no main effect of stimulus type, $F(1, 2) = 2.0$, $p = .16$, a main effect of subject, $F(2, 57) = 51.4$, $p < .01$, and no interaction between stimulus type and subject, $F(2, 57) = 0.28$, $p = .76$. A Fisher's post hoc test revealed that the main effect of subject was due to Benedict's inferior performance relative to the other 2 monkeys ($p < .01$).

Experiment 2B

Experiment 2A is the first experiment in which animals were tested on their ability to compare novel numerosities outside the original training values. Both subjects trained on an ascending series responded at above-chance levels of accuracy on novel–novel numerosity pairs, albeit at a lower level than they did on familiar–familiar and familiar–novel pairs. The subject trained on the descending series responded at a chance level of accuracy on the novel–novel numerosity pairs. Because there were only 200 trials on

---

[3] The finding that Rosencrantz and Macduff were both below chance on 6 versus 8 but not 6 versus 7 (a presumably harder discrimination) prompted us to investigate whether the 2 subjects were incorrect on the same exemplar pairs. We found no evidence that this difference was the result of particular stimulus pairs. There was no correlation between Rosencrantz's and Macduff's performance on specific pairs.

Table 5
Percentage Correct for Each Pairwise Numerical Comparison (Experiment 2A)

	1	2	3	4	5	6	7	8	9
					Rosencrantz				
1	—	94	97	99	100	100	89	95	100
2		—	93	95	95	100	100	90	100
3			—	83	89	95	100	100	95
4				—	89	80	90	100	85
5					—	75	75	95	79
6						—	70	60	70
7							—	65	83
8								—	42
9									—
					Macduff				
1	—	80	97	100	100	95	100	95	95
2		—	94	96	95	100	100	85	94
3			—	83	85	95	90	95	100
4				—	75	85	95	95	90
5					—	75	75	95	100
6						—	80	63	80
7							—	55	75
8								—	55
9									—
					Benedict				
1	—	87	90	96	95	85	85	90	85
2		—	80	85	80	75	80	75	75
3			—	66	50	75	70	75	60
4				—	65	60	65	40	50
5					—	35	40	45	40
6						—	60	50	65
7							—	40	50
8								—	85
9									—

Note. Data is based on 20 trials for each of the 36 numerical comparisons.

which novel–novel pairs were presented in Experiment 2A, it seemed prudent to see if accuracy levels would increase with additional training. In Experiment 2B, we tested subjects for an additional 40 sessions on the same 36 numerical comparisons (400 novel–novel trials).

An additional motivation for Experiment 2B was to compare animal and human representations of numerosity. When human subjects are asked to compare Arabic numerals or random dot patterns, they respond more accurately and more rapidly as the numerical distance between the members of a pair increases (Buckley & Gillman, 1974; Moyer & Landaeur, 1967, 1973; see Holyoak, 1978, for a review). Moreover, when numerical distance is held constant, accuracy decreases and latency increases as numerical magnitude increases. These findings have been interpreted as evidence that humans use analog numerical representations in numerical comparison tasks (e.g., Dehaene, 1997; Moyer & Landaeuer, 1967, 1973).[4] One of the goals of Experiment 2B was to determine whether similar accuracy and latency distance effects could be obtained from animals in numerical comparison tasks.

Method

The same subjects and apparatus used in Experiment 2A were used in Experiment 2B. The task was identical to that of Experiment 2A with two exceptions. In Experiment 2B all correct sequences were reinforced with food. Responses that did not conform to the required sequence produced a TO and ended the trial. The second difference was the relative frequencies of the familiar–familiar, familiar–novel and novel–novel pairs. Because reinforcement was available on every trial, the frequency of familiar–familiar pairs was not weighted differentially as in Experiment 2A. A unique pair of stimuli was used for each of the 36 numerosity comparisons during each of the 40 sessions. Each of the 36 numerosity pairs was presented twice during each session (72 trials). To ensure that accuracy data was not affected by learning of specific stimulus pairs, accuracy analyses are reported only for the first presentation of each stimulus pair. Latency analyses, however, were based on all correct responses.

The experiment consisted of two blocks of 20 sessions each. The same stimuli that were used in Experiment 2A were used in Block 1 of Experiment 2B (1,680 stimuli). The data obtained from Block 1 on trials on which the numerosities 5–9 were used, was equivalent to first-trial data because no reinforcement was provided in Experiment 2A on any of those trials. A new stimulus set was used in Block 2 (an additional 1,440 stimuli). The stimulus set used in Block 2 differed in two respects from that used in Block 1. The elements used to compose each stimulus were black, and one third of the stimuli contained pairs for which the smaller and larger numerosity had equal surface area. During Block 2, subjects were also required to touch a start signal (a white square on green screen) to initiate each trial. We did this to maximize attention to each pair at the start of a trial.

Results and Discussion

All subjects responded more accurately in Experiment 2B than in Experiment 2A. Most striking was the improvement in Benedict's performance. Benedict responded to all pairs of novel–novel numerosities at greater-than-chance levels of accuracy. This shows that his poor performance in Experiment 2A cannot be attributed to an inability to discriminate large numerosities. Experiment 2B also showed that the accuracy and the reaction time of the first response to each pair was strongly influenced by the numerical magnitude of the stimuli and the numerical distance between the stimuli in each pair.

Combining Accuracy and Latency Data From Blocks 1 and 2

Accuracy of responding to familiar–familiar, familiar–novel, and novel–novel pairs is shown in Figure 8. A repeated measures ANOVA (Block × Subject) revealed no main effect of either factor, block, $F(1, 2) = 1.44, p > .23$; subject, $F(2, 57) = 1.77, p > .18$, and no interaction between those factors, $F(2, 57) = 0.62, p = .54$. We therefore combined accuracy data from Blocks 1 and 2. Binomial tests on accuracy of responding to the 36 numeros-

[4] Magnitude and distance effects are also found with continuous dimensions such as line length (Welford, 1960).
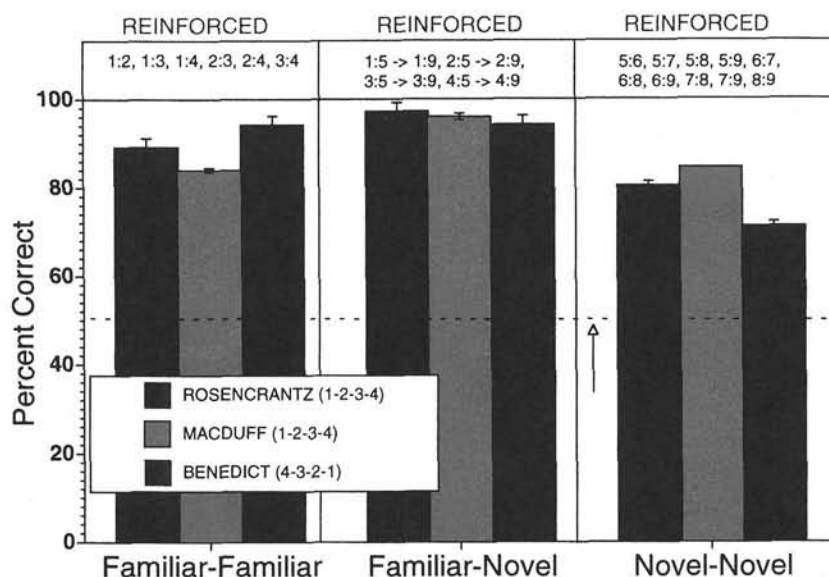
Figure 8. Performance on familiar–familiar, familiar–novel, and novel–novel numerosity pairs for 40 sessions (Experiment 2B). All trials were reinforced.

ity pairs showed that performance was greater than that expected by chance for each of the 3 monkeys on each of the numerical comparisons.

We also combined latency data from Blocks 1 and 2 because the trend within each block was the same. A repeated measures ANOVA (Numerical Distance × Block) on the latencies of the first response to each pair revealed main effects of both factors, distance, $F(7, 100) = 15.8, p < .0001$; block: $F(1, 100) = 22.6, p < .0001$, but no interaction between those factors, block × Distance, $F(7, 100) = 1.05, p = .4$. The main effect of block was due to an overall decrease in latency from Block 1 to Block 2. That decrease could have resulted from either or both of the

following factors: practice effects and the introduction of a start stimulus at the start of Block 2.

Accuracy and Latency Distance Effects

Accuracy and latency of responding to the 36 pairs of numerosities were influenced by the numerical difference between the members of each pair (distance effect). We obtained distance effects for all 3 subjects in Experiment 2B for both accuracy and latency. As shown in Figure 9, accuracy increased and latency decreased as the numerical distance between response alternatives increased. The slopes of the regression lines between distance and accuracy and
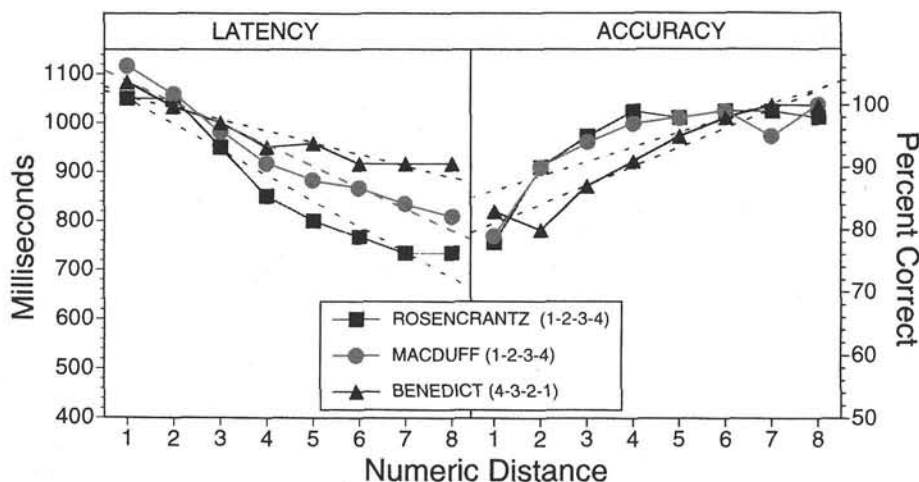

Figure 9. Accuracy and latency as a function of numerical distance for 40 sessions (Experiment 2B). Latency reflects correct responses. The dashed lines reflect the best-fit linear models.

between distance and latency were significant for each monkey, for accuracy, Rosencrantz, $r^2 = .46$, $p < .01$; Macduff, $r^2 = .56$, $p < .01$; Benedict, $r^2 = .33$, $p < .01$; for latency, Rosencrantz, $r^2 = .51$, $p < .01$; Macduff, $r^2 = .64$, $p < .01$; Benedict, $r^2 = .60$, $p < .01$.

As shown in Figure 10A and 10B, we obtained an accuracy distance effect (gray functions) even when the value of the first numerosity was held constant. For example, for the monkeys who learned the ascending rule, accuracy was 80% for the 3–4 comparison but 98% for the 3–9 comparison. This effect would probably have been greater were it not for a ceiling of 100%.

Figures 10C and D (dashed gray functions) illustrate that latency distance effects were also apparent when the value of the first numerosity was held constant. For example, the average latency to the first response for the monkeys who learned the ascending rule was 950 ms for the 3–4 comparison and 800 ms for the 3–9 comparison. Similarly, Benedict (descending rule) took over 1,100 ms to order the 3–4 pair but only 900 ms to order the 3–9 pair.

Accuracy and Latency Magnitude Effects

Accuracy of responding was also influenced by the magnitude of the numerosities being compared, irrespective of numerical distance between the members of each pair. Consider, for example, the pairs 1–2, 2–3, 3–4, 4–5, 5–6, 6–7, 7–8, and 8–9. Although the numerical distance between the members of each pair is 1, Figures 10A and 10B (solid black functions) show that accuracy of responding generally decreased with increasing numerical magnitude for all 3 monkeys. The linear relationship between accuracy and magnitude was significant for Rosencrantz and Benedict at distances of 1, 2, and 3 ($p < .01$) but was not significant for Macduff.

The effect of numerical magnitude on latency of responding presents a more complicated picture. Figure 10C (solid black function) shows that the latencies of both monkeys who learned the ascending numerical rule increased with increasing magnitude when distance is held constant. However, Figure 10D (solid black function) shows that magnitude did not influence Benedict's latency of responding in the same manner. For example, Benedict took longer to correctly compare the numerosities 1 and 2 than to correctly compare the numerosities 8 and 9. However, as mentioned earlier his accuracy was similar to that of the other 2 monkeys on these pairs.

Stimulus Controls

The surface area of the elements of each stimulus had no systematic effect on accuracy. Indeed, the level of accuracy was at least 84% for each monkey under all three stimulus conditions (smaller numerosity had a smaller surface area; larger numerosity had a smaller surface area; surface area was equal for both numerosities). The relevant data are shown in Table 6, and the statistical analyses are summarized in Table 7.

General Discussion

We assessed the numerical ability of rhesus monkeys by two new paradigms: serial learning of numerical stimuli and numerical comparisons of novel numerosities. The performance of monkeys on the serial learning task provides clear evidence that monkeys do not rely on number as a "last-resort" strategy, as suggested by Davis and Perusse (1988). Instead of memorizing different sets of stimuli by rote (as predicted by a last-resort strategy), monkeys relied on the numerical attributes of stimuli to determine the correct order of responding on both ascending and descending sequences. The performance of monkeys on the numerical comparison task provides the strongest evidence to date that monkeys naturally represent numerosity on an ordinal dimension. The reaction time functions obtained on this task also revealed numerical magnitude and distance effects, the first such effects obtained from an animal.

Last-Resort Strategy

Opinion is divided about the interpretation of experiments purporting to show that animals have numerical ability. One view holds that animals naturally detect variations in numerosity; another, that they attend to numerical differences only as a last resort, that is, when all nonnumerical cues are eliminated. In support of the last-resort hypothesis, Davis and Perusse (1988) argued that animals avoid using numerosity when nonnumerical cues are available and that animals generally need a large number of training trials to produce evidence of numerical competence. By contrast, Gallistel and Gelman (1992) contended that animals attend to number routinely and automatically.

The debate over the salience of numerosity has focused on comparisons of numerosity and other stimulus dimensions, for example, time. In a seminal study, Meck and Church (1983) showed that rats encoded both time and numerosity during training on a bisection task on which the duration and the numerosity of the discriminative stimuli were confounded. They concluded that their data contradict the last-resort hypothesis. They also argued that rats use the same underlying mechanism to keep track of numerosity and time automatically.

The results of Meck and Church's (1983) experiment have been widely cited as strong support for the claim that animals are naturally attuned to the numerosity dimension (e.g., Gallistel & Gelman, 1992). However, a recent study by Breukelaar and Dalrymple-Alford (1998) provided support for the last-resort hypothesis. When time and number were confounded during training, time, but not number, exerted control of behavior during a generalization test. When time and number were trained separately, the rats' behavior was controlled by number; however, control by time was greater than control by number. Furthermore control by time was acquired more rapidly than control by number.

In the present study, numerosity was the only dimension along which subjects could order stimuli. Accordingly, the performance of our subjects is not directly relevant to the issue of the relative salience of number. Certain features of our results are, however, relevant to the issue of the
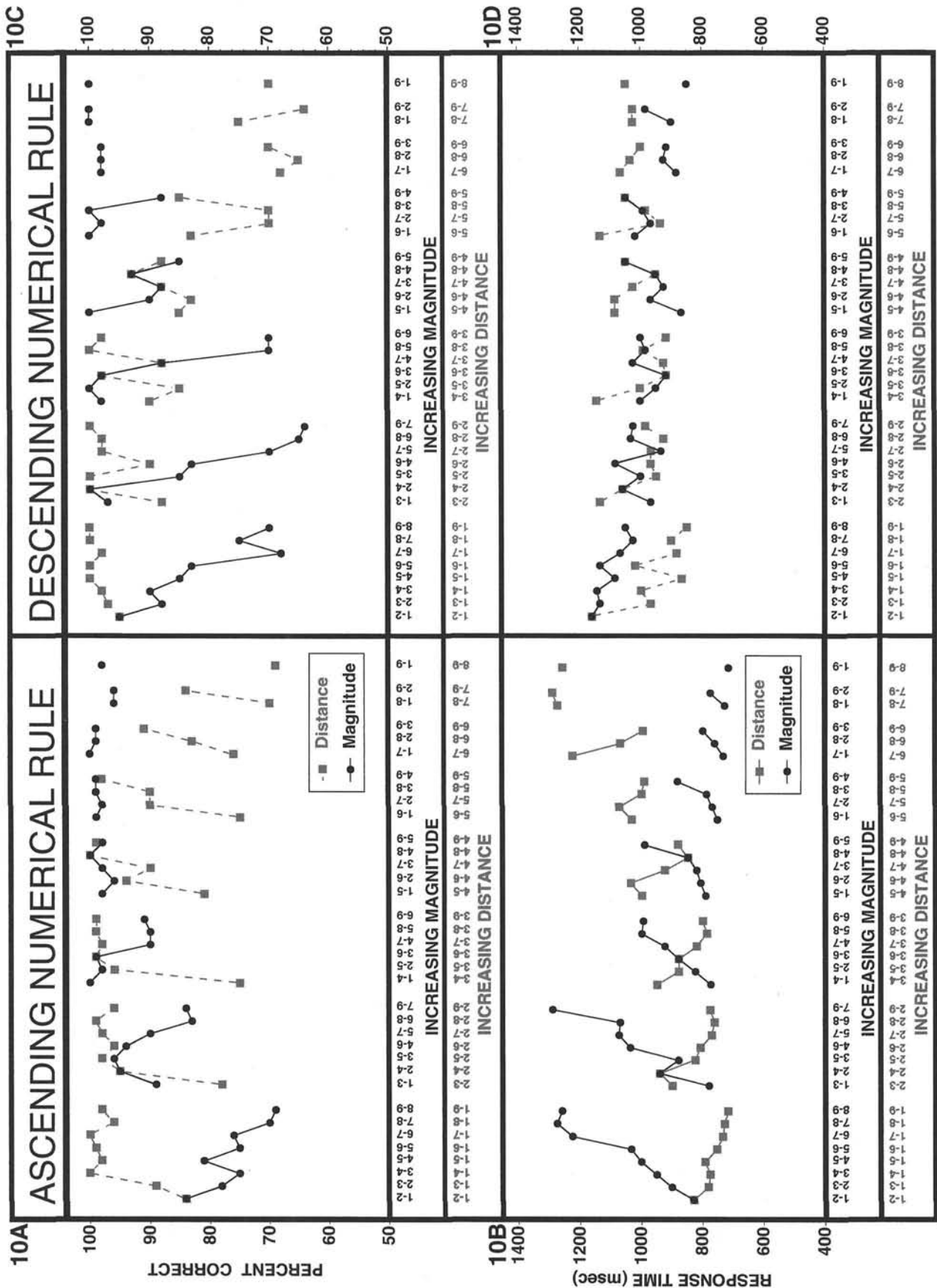
Table 6

Percentage Correct for Each Stimulus Type in Blocks 1 and 2 of Experiment 2B

Smaller numerosity has	Block 1			Block 2		
	Rosencrantz	Macduff	Benedict	Rosencrantz	Macduff	Benedict
Smaller area	84	86	90	90	86	85
Larger area	94	92	86	96	96	89
Equal area	NA	NA	NA	92	95	89

Note. Block 1 differed from Block 2 in that novel stimuli were used in Block 2, and a start signal was used in Block 2. Each block contained 20 sessions.

"naturalness" of number as a stimulus dimension. In Experiment 1, subjects could have memorized each of the 35 training lists and ignored the numerical dimension. Because each stimulus set was trained for at least 60 trials, the subjects could have learned the correct order in which to respond to those sets by trial and error. Indeed, these monkeys had previously learned to produce arbitrary lists of photographs in an experiment in which responding by trial and error and rote memorization were the only bases for learning the required sequence (Terrace et al., 2000).

The results of Experiment 1 provide two types of evidence that our monkeys used a numerical rule spontaneously. Rosencrantz and Benedict performed at above-chance levels of accuracy on the *first* trial of each of the 35 training sets. Such performance cannot be explained by appeal to memory of previously learned sets. There was also no difference in performance between training sessions where subjects could have memorized particular stimulus sets and transfer sessions during which memorization could not have been a factor. These two findings suggest that the monkeys took advantage of the opportunity to use a numerical rule rather than memorize 35 distinct stimulus sets.

Experiment 2 provided additional evidence that rhesus monkeys naturally attend to numerosity. Both monkeys who learned the ascending rule ordered the novel numerosities 5–9 spontaneously after training on the numerosities 1–4. This shows that monkeys do not need to be taught to discriminate and order a new range of numerosities after those skills were trained on a different range.

Stimulus Control by Numerosity

The complexity and variety of the training stimuli ruled out stimulus control by nonnumerical cues as an explanation

of our subjects' ability to discriminate exemplars of the numerosities 1–4. In Experiment 1, subjects responded with the same accuracy to stimulus sets for which surface area was held constant as they did to sets for which the size of the elements was held constant. Memorization of particular motor sequences as an explanation of correct sequences was ruled out by between-trial variation of the configurations of each of the 35 sets of training stimuli. Rote memorization of particular stimuli was ruled out by testing subjects with 150 trial-unique stimulus sets.

Stimulus control by nonnumerical cues was similarly ruled out in Experiments 2A and 2B as an explanation of performance on the numerical pairs task. In contrast to other studies (e.g., Honig & Stewart, 1989), the surface area and size of the elements of exemplars of novel numerosities were varied randomly. In addition, we used novel stimuli during each session to prevent subjects from memorizing stimulus order (e.g., Hicks, 1956). To eliminate the possibility of cueing by the experimenter, subjects were trained and tested by automated equipment in an isolated room. To ensure that numerosity was not confounded with hedonic value (e.g., Washburn & Rumbaugh, 1991), we held constant the amount of reinforcement provided on each trial.

Ascending and Descending Numerical Rules

In Experiment 1, the task of responding to numerical stimuli in an ascending or a descending order was equally difficult.[5] In Experiment 2A, however, the subject who learned to respond in a descending order to the numerosities 1–4 (Benedict) could not extrapolate that rule to the numerosities 5–9. In Experiment 2B, Benedict was able to discriminate novel exemplars of the numerosities 5–9. Any or all of the following factors could have contributed to Benedict's inability to discriminate novel exemplars of the numerosities 5–9 in Experiment 2A. It may have been more difficult to apply the descending than the ascending rule to novel numerosities. The ascending rule (1 → 2 → 3 → 4*****) may be easier than the descending rule (*****4 → 3 → 2 → 1) because of the salience of the numerosity 1. In the case of the ascending rule an initial

Figure 10 (opposite). (A) Accuracy for each of the 36 numerical pairs (Experiment 2B) averaged for the 2 monkeys who learned the ascending numerical rule. (B) Accuracy for each of the 36 numerical pairs (Experiment 2B) for Benedict who learned the descending numerical rule. (C) Latency for each of the 36 numerical pairs (Experiment 2B) averaged for the 2 monkeys who learned the ascending numerical rule. (D) Latency for each of the 36 numerical pairs (Experiment 2B) for Benedict who learned the descending numerical rule. In all four panels, the solid black functions reflect accuracy or latency as a function of increasing magnitude when distance is held constant. The dashed gray functions show accuracy or latency as a function of increasing distance when the magnitude of the first item is held constant.

---

[5] Note that Benedict's success at ordering novel exemplars of the numerosities 1, 2, 3, and 4 supports the claim that he either learned a specific numerical rule (1, then 2, then 3, then 4) or a more abstract descending rule. His failure on the unreinforced novel–novel numerosity pairs in Experiment 2A suggests that he may not have learned an abstract descending numerical rule.

Table 7
F *Values for ANOVAS (Subject × Stimulus Type)*

Variable	Block 1	Block 2
Main effect of subject	$F(2, 57) = 0.31$	$F(2, 55) = 12.12*$
Main effect of stimulus type	$F(1, 57) = 9.03*$	$F(2, 110) = 11.55*$
Interaction	$F(2, 57) = 7.81*$	$F(4, 110) = 1.43$

Note. Block 1 and Block 2 differ in that there were 3 stimulus sets in Block 2 (smaller numerosity had smaller, larger, or equal surface area).
$*p < .05$.

response to 1 was correct on any trial that contained that value. In contrast, 4 may be a much less salient value. It is also possible that Benedict's initial failure resulted from the particular values of the numerosities used during training and testing.

To determine the relative influence of the direction of the training rule (ascending or descending), it would be interesting to train monkeys on $4 \rightarrow 5 \rightarrow 6$ and $6 \rightarrow 5 \rightarrow 4$ sequences and then test them on all possible pairs of the numerosities 1–9. If the descending rule is more difficult to extrapolate, then one would predict poorer performance on the 1–2, 2–3, and 1–3 pairs following $6 \rightarrow 5 \rightarrow 4$ training than after $4 \rightarrow 5 \rightarrow 6$ training. Alternatively, if magnitude and distance are the major factors influencing the ease of discrimination, performance should be equivalent on the 1–2, 2–3, and 1–3 pairs after $6 \rightarrow 5 \rightarrow 4$ and $4 \rightarrow 5 \rightarrow 6$ training.

Nonmonotonic Versus Monotonic Rules

Monotonic and nonmonotonic sequences pose different representational problems. Both monotonic and nonmonotonic sequences can be represented as a series of specific numerosities. By contrast, monotonic rules can be represented directionally without regard to specific numerosities. A subject could learn to find the largest (or smallest) numerosity, than the next largest (or smallest), and so on. That strategy doesn't require the subject to remember the specific cardinal values of the numerosities.

A monkey who learned to execute a monotonic sequence by using a directional rule, rather than by memorizing the sequence as a series of particular numerosities, should be able to execute sequences that do not include specifically trained numerosities (see McGonigle & Chalmers, 1986, for a similar argument using the size dimension). That prediction was confirmed in Experiment 2 (but only in the case of the ascending numerical rule). Monkeys trained to order the numerosities 1–4 in an ascending sequence ordered pairs of the numerosities 5–9.

Macduff's failure to learn the $3 \rightarrow 1 \rightarrow 4 \rightarrow 2$ rule and his subsequent success on the $1 \rightarrow 2 \rightarrow 3 \rightarrow 4$ rule indicates that nonmonotonic rules are more difficult to learn than monotonic rules. This finding is consistent with the results of other studies of serial learning that have shown that children, monkeys, and rats learn monotonic sequences more easily than nonmonotonic sequences when the stimuli vary in size

or hedonic value (Hulse & Dorsky, 1977; McGonigle & Chalmers, 1986; Oshiba, 1997).

Macduff's inability to learn the nonmonotonic rule could have resulted from difficulty in making cardinal numerical judgments. For example, in the case of a nonmonotonic sequence, for example, $3 \rightarrow 1 \rightarrow 4 \rightarrow 2$, the subject must represent the cardinal value of each numerosity and its ordinal position (i.e., 3 comes first, 1 comes second, 4 comes third, and 2 is last). Alternatively, Macduff's difficulty with the nonmonotonic sequence may have resulted from interference from reversals in the valence of successive numerosities. To address those issues, it would be of interest to train monkeys on different kinds of nonmonotonic rules. For example, monkeys might be able to learn nonmonotonic rules with only one reversal, for example, $4 \rightarrow 3 \rightarrow 1 \rightarrow 2$ or $4 \rightarrow 1 \rightarrow 2 \rightarrow 3$.

Distance and Magnitude Effects Based on Accuracy and Latency

The distance and magnitude effects obtained in Experiment 2B were based on both accuracy and latency data. They are the first such effects obtained from an animal that can unequivocally be attributed to numerosity. Subjects responded more accurately and more rapidly as the numerical distance between the items in a test pair increased. When distance was held constant, increases in the magnitude of the items resulted in a decrease in accuracy and an increase in latency, in the case of the ascending rule. These robust distance and magnitude effects provide strong additional evidence that monkeys rely on ordinal representations of number when judging the relative numerosity of pairs of stimuli.

The magnitude and distance effects observed in our data are consistent with two hypotheses regarding the manner in which number is represented. One is that number is represented on a linear scale with scalar variability (i.e., the standard deviation grows proportionally to the mean; Gallistel & Gelman, 1992; Meck & Church, 1983); the other, that number is represented on a logarithmic scale with insignificant variability (e.g., Buckley & Gillman, 1974; Dehaene, 1992; Holyoak, 1978). Distinguishing between these hypotheses requires future research (e.g., see Gibbon & Church, 1981, for a comparison of linear and logarithmic representations of time).

In addition to a distance effect, the monkeys' performance in Experiment 2 was strongly influenced by the numerical magnitude of the stimuli. With distance held constant, performance generally decreased with increasing magnitude (e.g., accuracy was poorer on 6–7 comparisons than on 2–3). Magnitude had a consistent effect on accuracy for both monkeys who learned the ascending rule and for Benedict, who learned the descending rule. However, the effect of magnitude on latency differed for the ascending and descending rules. Latency increased as magnitude increased for both monkeys who learned the ascending rule. However, Benedict's response times were far less systematic. Superficially, Benedict's response times suggest the opposite pattern of faster responding to larger magnitudes (i.e., he was much faster at differentiating 8 and 9 compared with 1 and 2).

Further research is needed to clarify the basis of the different effects of magnitude on reaction time on ascending and descending sequences.

Magnitude and distance effects may contribute to the apparent generalization decrement in novel–novel pairs shown in Figure 7. Most of the adjacent novel–novel pairs have smaller Weber fractions than familiar–familiar pairs and have on average smaller numerical distances than the familiar–novel pairs. Thus, both distance and magnitude effects could contribute to the reduced performance on the pairs composed exclusively of the numerosities 5–9. If this hypothesis is correct, then the poorer performance on novel–novel pairs results from the particular values we chose rather than a generalization decrement. This hypothesis could be tested in the hypothetical experiment described previously in which monkeys would be trained on a 4 → 5 → 6 or a 6 → 5 → 4 rule and then tested on pairs composed of the novel numerosities 1–3 and 7–9. The generalization decrement hypothesis would predict a decrement in performance on pairs composed of the novel numerosities regardless of their magnitude. Alternatively, if performance were controlled by the Weber fraction, performance would be better on the novel–novel pairs composed of the numerosities 1–3 than on familiar pairs composed of larger numerosities (e.g., 4–6).

The similarity in the shape of the latency and accuracy functions for humans and rhesus monkeys suggests that the two species may be using similar processes to compare different numerosities (Buckley & Gillman, 1974). At the very least, these similarities indicate that symbolic training on numerosity is not a necessary condition for the distance effect. There remain, however, many details of the comparison process(es) used by each species that need to be clarified before one can conclude that they are truly analogous. One issue is the numerical proficiency of human participants and monkey subjects at the time of testing. Human participants are highly overtrained in judging the ordinal relations of Arabic numerals and visual displays, and this is reflected in their almost errorless performance.

Are Rhesus Monkeys Subitizing or Counting?

It is unclear how an animal translates the numerosity of a stimulus into a numerical representation. It has been proposed that humans use two distinct processes when enumerating: subitizing and counting (e.g., Kaufman, Lord, Reese, & Volkmann, 1949; Mandler & Shebo, 1982; Trick & Pylyshyn, 1994). In contrast to counting, subitizing has been characterized as a *parallel* process whereby the elements of a visual display are automatically translated into a numerical representation. Some investigators have also argued that animal numerical discriminations can be explained by subitizing (e.g., Davis & Perusse, 1988; Rumbaugh, Savage-Rumbaugh, & Hegel, 1987). However, subitizing has yet to be defined operationally, and there are no relevant reaction-time data from animals (for an exception, see Mirofushi, 1997). Moreover, there is no consensus that subitizing is a distinct process from counting (Balakrishnan & Ashby, 1992; Gallistel & Gelman, 1991).

The experiments presented in this article were not de-

signed to determine the validity of subitizing and counting as explanations of numerical discrimination by animals. There are, however, some features of our results that cannot be explained by subitizing. Each of our 3 subjects made accurate relative numerosity judgments of numerosities well beyond the range of the numerosities that humans are said to subitize. For example, in Experiment 2B all 3 monkeys reliably discriminated novel exemplars of the numerosities 8 versus 9. It seems unlikely that a process that cannot accommodate numerosities beyond 4 or 5 in humans can handle numerosities as large as 9 in monkeys. It is also unclear how subitizing could account for ordinal comparisons. If the numerical ability of animals were based entirely on subitizing, then ordinal relations between the subitized entities would have to be learned individually (e.g., A comes before B). Our subjects' ability to order novel numerosities spontaneously shows that monkeys rely on a general comparison process that does not require each pairwise ordered relation to be memorized.

Although some researchers have argued that there is sufficient evidence that animals count (Breukelaar & Dalrymple-Alford, 1998; Capaldi & Miller, 1988; Meck & Church, 1983), others contend that animal numerical competence lacks important features of counting (Davis & Perusse, 1988). It is commonly agreed that counting involves three essential principles (Gelman & Gallistel, 1978). First, the one-to-one principle, which states that a unique cardinal tag is applied in 1:1 correspondence to each to-be-counted element. Second, the stable-order principle, which states that the cardinal tags must be applied in a consistent order. Third, the cardinal principle, which states that the last tag applied serves to represent the numerosity of the set. Although Gelman and Gallistel's definition does not exclude nonverbal organisms, finding evidence of these principles in nonhuman animals has proven extremely difficult. There are, however, reports in the literature on numerical competence in animals that support one or more of the Gelman and Gallistel counting principles (e.g., Boysen & Berntson, 1990; Capaldi & Miller, 1988; Meck & Church, 1983).

Meck and Church (1983) proposed a mode-control or "accumulator" model of animal counting. The model posits that each to-be-counted element causes a switch to be closed, which allows a pacemaker to increment an accumulator by a constant amount. The accumulation process outlined by Meck and Church's model is consistent with all of the counting principles described by Gelman and Gallistel (1978). It is important to note, however, that this model was developed as an explanation of a rat's ability to discriminate stimuli that were presented *sequentially* and that it cannot explain how an animal would discriminate stimuli that are presented *simultaneously*.[6]

---

[6] It is also interesting that the empirical latency values we obtained are inconsistent with the specific claim of the mode-control model that a switch is closed for 200 ms for each to-be-counted event (Meck & Church, 1983). For example, when the monkey is comparing two values, one possibility is that one accumulator is incremented x times and a second accumulator is incremented y times. The two accumulator values could then be compared to determine which contained a greater magnitude. The

Dehaene and Changeux's (1993) connectionist model of nonverbal numerical discriminations does provide for simultaneously presented stimuli. Dehaene and Changeux proposed that objects of varying size are normalized to produce size-independent activation that is highly correlated with numerosity. It is interesting that this model yields approximate numerical values and predicts both a distance and magnitude effect even though it doesn't follow Gelman and Gallistel's (1978) counting principles.

At the heart of the subitizing–counting debate is the question of whether a rhesus monkey's enumeration process is serial or parallel. Are each of the to-be-counted elements attended to sequentially or is the numerosity of the set apprehended in parallel? One way to address that question is to track eye movements as subjects make numerical judgments. Evidence of successive eye movements to each to-be-enumerated item would suggest that the enumeration process is serial as opposed to parallel and would lend support to a counting hypothesis.

Conclusion

Experiment 1 showed that rhesus monkeys can differentiate visual stimuli solely on the basis of numerosity on a task in which they were required to order the numerosities 1–4 in an ascending or a descending order. The results of Experiment 1 contradict the hypothesis that animals discriminate number only as a "last resort" in that knowledge of an ordinal rule was immediately apparent when subjects were tested with novel exemplars of the numerosities 1–4. Experiment 1 also showed that monotonic numerical rules were more easily learned than a nonmonotonic rule. This suggests that monkeys naturally attend to ordinal relations between numerosities. Experiment 2A showed that monkeys trained on an ascending numerical rule spontaneously infer the ordinal values of novel numerosities when numerosity is varied from 5 to 9. Experiment 2B provided the first demonstration of numerical distance and magnitude effects for accuracy and latency of responding in an animal. Taken together the results of these experiments provide compelling evidence that number is a meaningful dimension for rhesus monkeys.

Our results demonstrate that neither language nor numerical symbols are necessary for discriminating and ordering visual stimuli on the basis of their numerosity. The ease with which our monkeys learned to discriminate and order numerical stimuli suggests that the difficulty of the tasks we used could be increased considerably before we would reach the limits of a monkey's numerical ability. For example, it is of considerable interest to determine whether monkeys can discriminate values larger than 9, whether their cardinal numerical skills are as advanced as their ordinal skills, whether their numerical representations are abstract in the

sense that they are not bound to a particular modality, and the extent to which monkeys can learn to represent number symbolically. Experiments designed to address these issues should help to define how a monkey processes numerical information and help delineate the precursors of human mathematical ability.

References

Balakrishnan, J. D., & Ashby, F. G. (1992). Subitizing: Magical numbers or mere superstition. *Psychological Research, 54,* 80–90.

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

Boysen, S. T., & Berntson, G. G. (1990). Development of numerical skills in the chimpanzee. In S. T. Parker & K. R. Gibson (Eds.), *"Language" and intelligence in monkeys and apes.* Cambridge, England: Cambridge University Press.

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

Brannon, E. M., & Terrace, H. S. (1998, October 23). Ordering of the numerosities 1–9 by monkeys. *Science, 282,* 746–749.

Breukelaar, J. W. C., & Dalrymple-Alford, J. C. (1998). Timing ability and numerical competence in rats. *Journal of Experimental Psychology: Animal Behavior Processes, 24,* 84–97.

Buckley, P. B., & Gillman, C. B. (1974). Comparisons of digit and dot patterns. *Journal of Experimental Psychology, 103,* 1131–1136.

Capaldi, E. J., & Miller, D. J. (1988). Counting in rats: Its functional significance and the independent cognitive processes which comprise it. *Journal of Experimental Psychology: Animal Behavior Processes, 14,* 3–17.

Carey, S. (1998, October 23). Knowledge of number: Its evolution and ontology. *Science, 282,* 641–642.

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior, Research, Methods, Instruments, and Computers, 25,* 257–271.

D'Amato, M. R., & Colombo, M. (1988). Representation of serial order in monkeys (*Cebus apella*). *Journal of Experimental Psychology: Animal Behavior Processes, 14,* 131–139.

Davis, H. (1984). Discrimination of the number three by a raccoon (*Procyon lotor*). *Animal Learning and Behavior, 12,* 409–413.

Davis, H., & Memmott, J. (1982). Counting behavior in animals: A critical evaluation. *Psychological Bulletin, 92,* 547–571.

Davis, H., & Perusse, R. (1988). Numerical competence: From backwater to mainstream of comparative psychology. *Behavioral Brain Sciences, 11,* 602–615.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44,* 1–42.

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

Dehaene, S., & Changeux, J. (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 representation of numbers in the animal and human brain. *Nature Neuroscience, 21,* 355–361.

Emmerton, J., Lohmann, A., & Niemann, J. (1997). Pigeons' serial ordering of numerosity with visual arrays. *Animal Learning and Behavior, 25,* 234–244.

latency to make such a judgment should then be at least as great as 200 ms times the greater numerosity. The empirical values we obtained do not match such a prediction (e.g., 6 vs. 9 = 997 ms, almost half that of the predicted minimum of 1,800). It is, however, possible that the 200-ms value is not constant across species.

Farrar, D. (1967). Picture memory in the chimpanzee. *Perceptual and Motor Skills, 25,* 305–315.

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

Gallistel, C. R., & Gelman, R. (1991). Subitizing: The preverbal counting process. In W. Kessen, A. Ortony, & F. Craik (Eds.), *Memories, thoughts and emotions: Essays in honor of George Mandler* (pp. 65–81). Hillsdale, NJ: Erlbaum.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44,* 43–74.

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

Gibbon, J., & Church, R. M. (1981). Time left: Linear versus logarithmic subjective time. *Journal of Experimental Psychology: Animal Behavior Processes, 7,* 87–108.

Greene, S. L. (1983). Feature memorization in pigeon concept formation. In M. L. Commons, R. J. Herrnstein, & A. L. Wagner (Eds.), *Discrimination processes* (pp. 209–231). Cambridge, MA: Ballinger.

Harlow, H. F. (1949). The formation of learning sets. *Psychological Review, 56,* 51–65.

Hauser, M. D., MacNeilage, P., & Ware, M. (1996). Numerical representations in primates. *Proceedings of the National Academy of Science, 93,* 1514–1517.

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

Holyoak, K. (1978). Comparative judgments with numerical reference points. *Cognitive Psychology, 10,* 203–243.

Honig, W. K., & Stewart, K. E. (1989). Discrimination of relative numerosity by pigeons. *Animal Learning and Behavior, 17,* 134–146.

Hulse, S. H., & Dorsky, N. P. (1977). Structural complexity as a determinant of serial pattern learning. *Learning and Motivation, 8,* 488–506.

Hurford, J. R. (1987). *Language and number.* Oxford, England: Basil Blackwell.

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

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

Matsuzawa, T. (1985). Use of numbers by a chimpanzee. *Nature, 315,* 57–59.

McGonigle, B. O., & Chalmers, M. (1986). Representations and strategies during inference. In T. Myers, K. Brown, & B. O. McGonigle (Eds.), *Reasoning and discourse processes* (pp. 141–164). New York: Academic Press.

Mechner, F. (1958). Probability relations within response sequences under ratio reinforcement. *Journal of the Experimental Analysis of Behavior, 1,* 109–122.

Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes, 9,* 320–334.

Mirofushi, K. (1997). Numerical matching behavior by a chimpanzee (*Pan troglodytes*): Subitizing and analogue magnitude estimation. *Japanese Psychological Research, 39,* 140–153.

Moyer, R. S., & Landaeur, T. K. (1967). Time required for judgments of numerical inequality. *Nature, 215,* 1519–1520.

Moyer, R. S., & Landaeur, T. K. (1973). Determinants of reaction time for digit inequality judgments. *Bulletin of the Psychonomic Society, 1,* 167–168.

Olthof, A., Iden, C. M., & Roberts, W. A. (1997). Judgments of ordinality and summation of number symbols by squirrel monkeys (*Saimiri sciureus*). *Journal of Experimental Psychology: Animal Behavior Processes, 23,* 325–339.

Oshiba, N. (1997). Memorization of serial items by Japanese monkeys, a chimpanzee, and humans. *Japanese Psychological Research, 39,* 236–252.

Pepperberg, I. M. (1987). Evidence for conceptual quantitative abilities in an African grey parrot: Labeling of cardinal sets. *Ethology, 75,* 37–61.

Platt, J. R., & Johnson, D. M. (1971). Localization of position within a homogeneous behavior chain: Effects of error contingencies. *Learning and Motivation, 2,* 386–414.

Rilling, M. (1967). Number of responses as a stimulus in fixed interval and fixed ratio schedules. *Journal of Comparative and Physiological Psychology, 63,* 60–65.

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.

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.

Swartz, K. B., Chen, S., & Terrace, H. S. (1991). Serial learning by Rhesus Monkeys. I: Acquisition and retention of multiple four-item lists. *Journal of Experimental Psychology: Animal Behavior Processes, 17,* 396–410.

Terrace, H. S. (1966). Stimulus control. In W. G. Honig (Ed.), *Operant behavior areas of research and application* (pp. 291–344). Englewood Cliffs, NJ: Prentice Hall.

Terrace, H. S. (1984). Simultaneous chaining: The problem it poses for traditional chaining theory. In M. L. Commons, R. J. Herrnstein, & A. R. Wagner (Eds.), *Quantitative analysis of behavior: Discrimination processes* (pp. 115–138). Cambridge, MA: Ballinger.

Terrace, H. S., Son, L. K., & Brannon, E. M. (2000). The development of serial expertise in rhesus monkeys. Manuscript in preparation.

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

Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited capacity preattentive stage in vision. *Psychological Review, 101,* 80–102.

Washburn, D., & Rumbaugh, D. M. (1991). Ordinal judgments of numerical symbols by macaques (*Macaca mulatta*). *Psychological Science, 2,* 190–193.

Welford, A. T. (1960). The measurement of sensory-motor performance: Survey and reappraisal of twelve years progress. *Ergonomics, 3,* 189–230.