

# Enumeration of briefly presented items by the chimpanzee (*Pan troglodytes*) and humans (*Homo sapiens*)

MASAKI TOMONAGA and TETSURO MATSUZAWA  
*Kyoto University, Aichi, Japan*

In this study, we compared the performances on an enumeration task (numerical labeling task) of 1 chimpanzee (*Pan troglodytes*) and 4 humans. In this task, two types of trials, with different exposure durations of the sample that was to be enumerated, were used. In the unlimited-exposure trials, the sample remained on until the subject made a choice. In the brief-exposure trials, the sample was presented for 100 msec and then was masked. The results show clear differences between the different species. The main differences had to do with accuracy during the unlimited trials and response times during the brief trials. Detailed analyses of the pattern of response times for the chimpanzee and of looking-back behavior during the task suggests that the enumeration process underlying the subject's performance was not counting but estimation.

In humans, three different processes of enumeration are frequently reported (Dehaene, 1992; Folk, Egeth, & Kwak, 1988; Jensen, Reese, & Reese, 1950; Jevons, 1871; Kaufman, Lord, Reese, & Volkman, 1949; Klahr, 1973; Mandler & Shebo, 1982; Oyama, Kikuchi, & Ichihara, 1981). The first is called *subitizing* (Kaufman et al., 1949) and is defined by an accurate and fast response. Response times increase only slightly (approximately 25–30 msec/item) when the number of items increases from 1 to 4 or 5. Some researchers suggest that the subitizing is a kind of perceptual process (Miller, 1993). The second is called *counting*, defined by the monotonically increasing response time function when the number of items increases from 5 or 6 to 9 or 10. The third process is called *estimation*. When the number of items is more than 10 or so, humans show a ceiling effect for response times and a gradual decline in accuracy as the number increases.

There is also a large literature on animals' numerical ability (Boysen & Capaldi, 1993; Davis & Memmott, 1982; Davis & Perusse, 1988; Dehaene, 1992; Gallistel & Gelman, 1992; Suzuki & Kobayashi, 2000). In animals, various types of tasks are used to assess their numerical abil-

ity, such as simple discrimination for relative or absolute numerical judgments (Davis, 1984; Dooley & Gill, 1977; Hicks, 1956; Honig & Stewart, 1989; Thomas, Fowlkes, & Vickery, 1980; Watanabe, 1998), discrimination of the number of rewards by using a runway apparatus (Capaldi & Miller, 1988), and identity matching of the number (Woodruff & Premack, 1981; cf. Hayes & Nissen, 1971). Despite the recent progress in animal studies on numerical competence, it is still controversial to draw conclusions concerning the underlying processes of enumeration. Some researchers argue that animals only show subitizing, a perceptual, automatic, preattentive process (Davis & Perusse, 1988). Others have argued for more complex cognitive processes (Boysen & Berntson, 1989; Capaldi & Miller, 1988; Pepperberg, 1994).

Some researchers have noted a confusion in the use of terms for enumeration processes. Subitizing, counting, and estimation were originally derived from response time experiments with humans (Kaufman et al., 1949; Klahr, 1973). Miller (1993) suggests that modifying the original meaning of these processes to explain the various types of results from animal experiments is not acceptable. To use response times as data, it is necessary to teach "numerals" to animals, a simple, human-like, but difficult task (Biro & Matsuzawa, 2001a, 2001b; Boysen & Berntson, 1989; Fester & Hammer, 1966; Matsuzawa, 1985; Matsuzawa, Asano, Kubota, & Murofushi, 1986; Matsuzawa, Itakura, & Tomonaga, 1991; Murofushi, 1997; Pepperberg, 1987, 1994).

In such experiments, the symbolic matching-to-sample paradigm is frequently used. For example, Matsuzawa, Murofushi, and their colleagues trained a female chimpanzee to select an Arabic numeral corresponding to the number of real objects up to 6, presented as the sample in the matching-to-sample format (Matsuzawa, 1985; Mat-

---

This research was financially supported by the Grant-in-Aid for Scientific Research from the Ministry of Education, Science, Sports, and Culture, Japan (Grants 05206113, 05267225, 06260222, 06451017, 07102010, and 07252214). The authors thank Kiyoko Murofushi for her generous support for and valuable comments on this experiment. Thanks are also due Sumiharu Nagumo for his technical support, Kiyonori Kumazaki and Norihiko Maeda for their care of the chimpanzee, Fei Xu and Rafael Nunez for their comments on earlier versions of the manuscript, and Vanessa Hayes for her critical reading of the manuscript. Correspondence concerning this article should be sent to M. Tomonaga, Section of Language and Intelligence, Department of Behavioral Brain Sciences, Primate Research Institute, Kyoto University, Inoyama, Aichi 484-8506, Japan (e-mail: tomonaga@pri.kyoto-u.ac.jp).

suzawa et al., 1986). This training in numerical labeling was later continued by replacing the real objects with white dots presented on a computer monitor (Matsuzawa et al., 1991; Murofushi, 1997). The chimpanzee was successively introduced to larger numbers up to 10.

Analyses of response times and accuracy revealed similar patterns in humans and the chimpanzee. The chimpanzee showed quite accurate and fast responding from 1 to 4 or 5. This was considered to be the same process as subitizing in humans. From 5 or 6 to the second largest number, she showed monotonically increasing response times and decreasing accuracy as a function of number. This phase was regarded as analogous to the process of counting in humans. She also correctly matched the numeral to dots that varied in size.

The chimpanzee, however, also showed differences from humans. When a new (larger) number was introduced, she showed no positive transfer (such as savings of acquisition) to that number. She showed improved accuracy and faster response times on the largest number, relative to the second largest number at each training period. Murofushi (1997) argued that the chimpanzee might use analogue magnitude codes during the numerical labeling task but might not use the codes as humans use them. An apparent increase in response times from 5 to the second largest number provides time for a different process from counting. This possibility was supported by her actual behavior. She often looked back and forth at the item display. Murofushi considered this behavior as reflecting decision processes (cf. Wright, 1990, 1992), but not counting processes.

If this chimpanzee does not count the items as humans do, other experimental manipulations might clarify the differences. In the present experiment, we tested this chimpanzee on the *mental counting* task (Kaufman et al., 1949; Mandler & Shebo, 1982). When Mandler and Shebo presented an array containing several items to human subjects for very brief exposure times (such as 200 msec), they found the three different processes of enumeration described earlier: subitizing from 1 to 4, counting from 4 to 6 or more, and estimation for the larger numbers. The human subjects showed mental counting when the representation of the array could be held in consciousness, as by imagery. When the array was terminated by the subject's response, on the other hand, the estimation process disappeared.

In the present experiment, the chimpanzee and 4 human subjects were tested under identical experimental settings in two conditions. In the unlimited-exposure condition, the array was terminated by the subject's response to the numerals, whereas in the brief-exposure condition, the array was terminated after approximately 100 msec and was masked by white-meshed patterns. If the chimpanzee processes numbers in the same manner as humans, both species should show similar patterns.

## METHOD

### Subjects

The chimpanzee (*Pan troglodytes*) subject was a female adult, named Ai. She was approximately 14 years old at the onset of the

present experiment. When she was 5 years old, she was initially trained on a numerical labeling task by using Arabic numerals (1 to 6) and real-life objects (Matsuzawa, 1985; Matsuzawa et al., 1986). The objects were later replaced by white dots appearing on a CRT display that also included the numbers 7, 8, 9, and 10 (Matsuzawa et al., 1991; Murofushi, 1997). Concurrently with the present experiment, she was also trained on sequential responding with Arabic numerals (Tomonaga, Itakura, & Matsuzawa, 1993; Tomonaga & Matsuzawa, 2000; see also Biro & Matsuzawa, 1999; Kawai & Matsuzawa, 2000). She lived in an outdoor enclosure with five other adult chimpanzees and maintained her free-feeding body weight during the present experiment.

For human subjects, 4 graduate students (1 female and 3 males; 27 years old, on average) participated in the present experiment.

### Apparatus

Ai was tested in an experimental room (2.7 m wide  $\times$  2.1 m high  $\times$  1.5 m deep) adjacent to the outdoor enclosure. Two 14-in. CRT displays were installed 15 cm apart in this room. The left CRT (item display), controlled by a personal computer (NEC, PC-9801F2), was used for the presentation of arrays of white dots (2  $\times$  2). The right CRT (numeral display) was used for the presentation of Arabic numerals (4  $\times$  4). These numerals were colored white. An optical touch panel (Carroll Touch International, UL-94V-0) was attached to the screen of this CRT. A response was defined as a touch to the screen of the CRT. This CRT was controlled by another computer (NEC, PC-98XA) linked with the first PC. Two 2  $\times$  2 keys were installed below the left and right sides of the left CRT. The left key could be illuminated and used for the presentation of the array (start key). The right key was used for presentation of the numerals. A food tray was installed below the right CRT. A universal feeder (Davis Scientific Instruments, UF-100) delivered small pieces of food (apples, pineapples, peanuts, almonds, raisins, etc.) into the food tray. The experiments with the humans were also conducted in this room.

### Procedure

Figure 1 shows the schematic diagrams of the trials, and Figure 2 shows the actual behaviors by the chimpanzee subject during the experimental sessions. Each trial began with the illumination of the left key and the white-meshed frame on the left CRT display. When the subject pressed this key three times, the array of dots was presented. The spatial arrangement of the dots was changed randomly from trial to trial. In the unlimited-exposure condition (Figures 1A and 2A), this array remained on until the subject responded to the numerals. In the brief-exposure condition (Figures 1B and 2B), on the other hand, this array was presented only for approximately 100 msec and then was masked by the white-meshed pattern. For the chimpanzee, the numerals (1 to 9) were presented on the right CRT display by pressing the right key, whereas they were automatically presented 750 msec after the onset of presentation of the array for the humans (see the Data Analyses section for details). The arrangement of the numerals also was changed randomly from trial to trial. The correct numeral appeared randomly and with equal likelihood at each predefined location from trial to trial. The subject was required to touch the numeral corresponding to the number of the dots. If the subject selected the correct numeral, a 1-sec chime was presented, and a food reward was given for the chimpanzee. If the subject made an error, a 0.5-sec buzzer was presented, and a 3-sec time out was added to the usual 1-sec intertrial interval. For some sessions, a correction procedure was in effect. When the subject made an error, the same trial with the same spatial configurations of dots and numerals was repeated until the subject made a correct response. For the brief-exposure trials, the array remained until the subject's response in the correction trials. These correction trials were not used for data analyses.

Ai was first given preliminary training on the numerical matching task after a blank period of several months. She was trained with

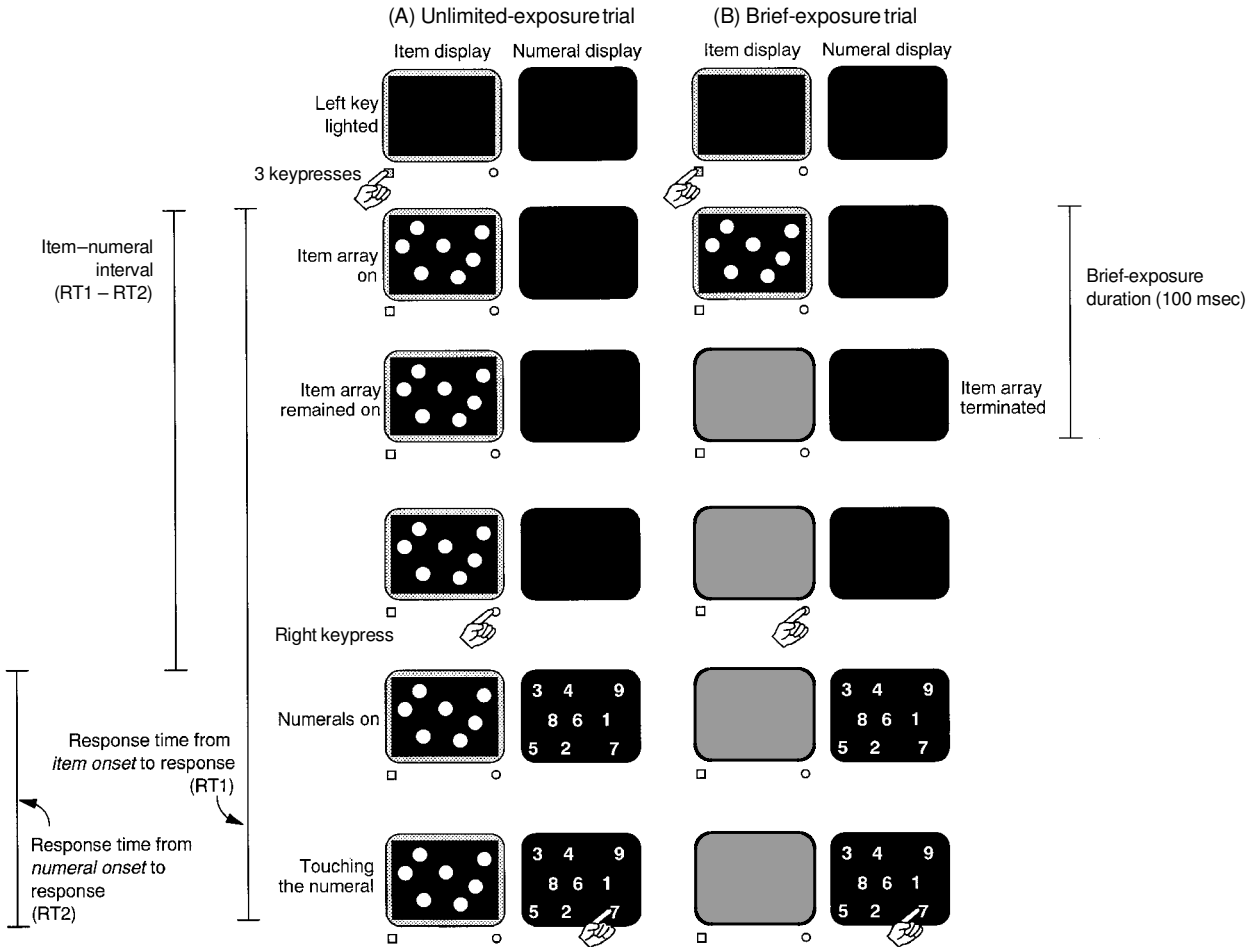


Figure 1. Schematic diagrams of trials for the unlimited-exposure (A) and the brief-exposure (B) conditions.

the unlimited-exposure condition for 7 sessions. Each session consisted of 120 trials. The array size was varied from 1 to 9. After she completed this initial training, she was then given 20 additional sessions for the exploration of the exposure-time parameter. The results of these sessions are not described in this paper (Tomonaga & Matsuzawa, 1992). Finally, Ai was given six 2-session blocks of testing in which the unlimited- and the brief-exposure conditions were randomly presented from trial to trial. Each session consisted of 144 trials. The array size was varied from 1 to 9.

Ai often looked back at the display of the array of dots by turning her head after the onset of the numeral display (Murofushi, 1997). This behavior was defined as a *looking-back* response. We did not measure the number of looking-back responses during the testing sessions. One week after the testing sessions, the subject was given a single 200-trial session for video recording, to measure the looking-back responses. Since slight eye movements without head movements could not be identified from the videotape, we defined the looking-back response here as the subject's turning her head to the item array, after she had looked at the numeral display, so that an imaginary vertical line drawn through the middle of the subject's head reached the center of the item display. The subject was allowed to make any number of looking-back responses in a trial. Looking-back responses were counted from the videotape for all the trials (243 trials), including error and correction trials. Only one experimenter checked the videotape, because the looking-back responses

were highly evident, so that it was not necessary to check the between-observer reliability.

The human subjects were also tested in the same experimental room under the same experimental setting, except for slight procedural changes (such as no food delivery and automatic presentation of the numerals). Each subject was initially given a single 40-trial practice session in which the two types of exposure time conditions were randomly presented. After a short break, a single 288-trial session was given to each subject.

**Data Analyses**

**Response time measurement.** In the present experiment, two types of response times were measured (Figure 1). RT1 was the time from the onset of the array to the response to the numeral. RT2 was the time from the onset of the numeral to the response. During the present experiment for the chimpanzee, the differences between RT1 and RT2 (item-numeral intervals) were quite stable across all conditions. Mean duration of these intervals averaged across all the 2,993 trials was 0.755 sec (*SD* = 0.158). The Pearson's correlation coefficient between RT1 and RT2, based on all the trials, was .969. We also calculated the mean item-numeral intervals and correlation coefficients between RT1 and RT2 separately for each numerosity. The mean item-numeral interval averaged across numerosities was 0.755 sec (range, 0.740–0.778), and the correlation coefficients between RT1 and RT2, calculated for each numerosity, ranged from

(A) Unlimited-exposure trial



Frame on



Stimulus array presented



Array remained on



Numerical array on

(B) Brief-exposure trial



Frame on



Stimulus array presented



Array masked



Numerical array on

**Figure 2.** The subject, chimpanzee Ai, performing the present experiment on unlimited-exposure (A) and brief-exposure (B) trials.

.786 to .984. These data suggest that the chimpanzee subject chained the responses to the left key (onset of RT1) and to the right key (onset of RT2) automatically as a result of extensive training under the present procedure. Because of this high stability, we used the re-

sponse times from the onset of the numerals to the response to the numeral (RT2) for subsequent data analyses in the present experiment.

As was described in the Procedure section, for human subjects, the second (right) key was inactivated but the item–numeral interval

was fixed to 0.75 sec on the basis of the chimpanzee data. We also used only RT2 for the analyses of the human data. Hereafter, we will use the term *response time* to refer to RT2.

**Analyses of the single-case ( $N = 1$ ) data.** It is frequently argued that single-case ( $N = 1$ ) data are not suitable for parametric statistical tests (Barlow & Hersen, 1984), because of serial dependencies among the data points. To avoid misinterpretations of the data biased by serial dependency, we first calculated the autocorrelations (lag = 1) for the chimpanzee data across sessions or session blocks. During the preliminary training with the unlimited-exposure condition, there were no significant autocorrelations for either percentage of error or response time across sessions [percentage of error,  $r = .080$ ,  $t(4) = 0.160$ ; response time,  $r = .405$ ,  $t(4) = 0.885$ ; both  $ps > .05$ ]. During testing, there also were no autocorrelations across session blocks [percentage of error,  $r = -.018$ ,  $t(3) = 0.031$ ; response time,  $r = -.029$ ,  $t(3) = 0.051$ ; both  $ps > .05$ ]. On the basis of these preliminary analyses, the present results were tested with analyses of variance (ANOVAs), with sessions or session blocks as repeated measures (Barlow & Hersen, 1984). Multiple comparisons were also conducted, using Fisher's LSD tests, if the main effects or interactions were significant.

The other way to analyze single-case data is the use of distribution-free or nonparametric tests. One of the distribution-free methods for analyses of the data from factorial experimental designs is the randomization test (Edgington, 1987; Manly, 1997). We also conducted ANOVAs, based on random permutations of residuals (Manly, 1997), on the present response time data for the chimpanzee. The methods and results are shown in Appendix A. Furthermore, we additionally conducted nonparametric paired comparisons based on the Wilcoxon's signed rank tests. These results are shown in Appendix B. These results were generally the same as those of the parametric analyses (see the Results section).

The human data were analyzed with traditional ANOVAs.

## RESULTS

### Preliminary Training

Figure 3 shows the mean percentage of errors and mean correct response times for the chimpanzee, averaged across seven sessions of the preliminary training with the unlimited-exposure condition. Ai showed quite stable per-

formance during this training (89.0% correct, on average). The results replicated well the previous studies (Matsuzawa et al., 1991; Murofushi, 1997). Both percentages of errors and response times exhibited slight increases from 1 to 5, monotonic increases from 5 to 8, and decreases at 9. Results of a one-way ANOVA showed that main effect of the number of dots was significant for percentage of error [ $F(8,48) = 12.12$ ,  $p < .001$ ]. For response time, the main effect was also significant [ $F(8,48) = 26.93$ ,  $p < .001$ ]. Table 1A shows the results of multiple comparisons conducted on the percentage of error and response time data. See also Appendices A–B for the results of distribution-free tests conducted on the same data. ANOVAs based on random permutation yielded the same results. Wilcoxon's tests also yielded results similar to those of the LSD tests, except for the response times for small numerosities.

### Testing

Figure 4 shows the results of testing. The left panels show the results of Ai, and the right panels show those of the humans. The upper panels show the mean percentages of error, and the lower panels show the mean response times on correct trials. As is shown in this figure, there are distinct differences between the chimpanzee and the humans.

**Chimpanzee.** Ai showed no difference in percentage of error from 1 to 8 between the unlimited- and the brief-exposure conditions. However she showed severe impairment for the number 9 in the brief-exposure condition but improvement in the unlimited-exposure condition. She also showed no difference in response times between exposure conditions from 1 to 6. A two-way (exposure  $\times$  number) ANOVA was conducted on these data. Both main effects and the interaction were significant [exposure,  $F(1,5) = 98.18$ ; number,  $F(8,40) = 34.02$ ; interaction,  $F(8,40) = 10.81$ ; all  $ps < .001$ ]. Subsequent analyses of simple main effects (Kirk, 1995) indicated that the effect of exposure duration was significant only when the num-

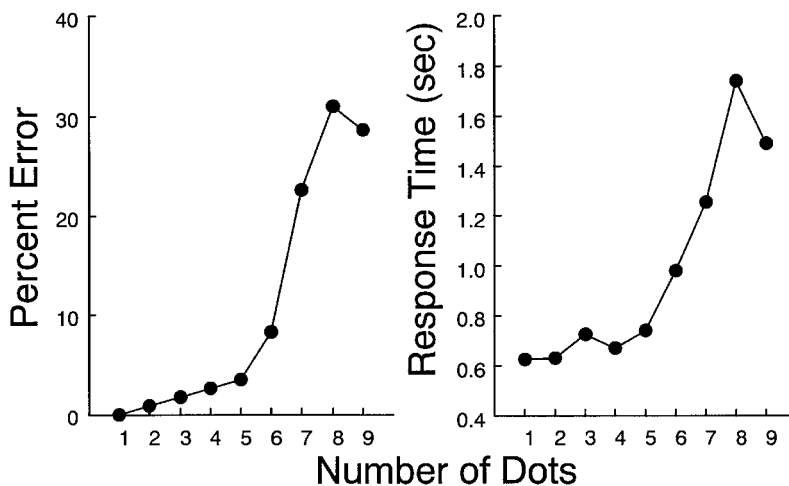


Figure 3. Mean percentages of errors and correct response times for Ai during preliminary training as a function of number of dots.

**Table 1**  
**Results of Multiple Comparisons (Fisher's LSD Tests) Between the Numbers of Dots for Each Condition**

A. Preliminary Training With Unlimited Exposure Condition for Ai								
Percent Error				Response Time				
	1	2	3	4	5	6	7	8
2	n.s.	—	—	—	—	—	—	—
3	n.s.	n.s.	—	—	—	—	—	—
4	n.s.	n.s.	n.s.	—	—	—	—	—
5	n.s.	n.s.	n.s.	n.s.	—	—	—	—
6	n.s.	n.s.	n.s.	n.s.	n.s.	—	—	—
7	***	***	***	***	***	**	—	—
8	***	***	***	***	***	***	n.s.	—
9	***	***	***	***	***	***	n.s.	n.s.

B. Test for Ai								
Percent Error/Unlimited				Response Time/Unlimited				
	1	2	3	4	5	6	7	8
2	n.s.	—	—	—	—	—	—	—
3	n.s.	n.s.	—	—	—	—	—	—
4	n.s.	n.s.	n.s.	—	—	—	—	—
5	n.s.	n.s.	n.s.	n.s.	—	—	—	—
6	***	***	**	***	*	—	—	—
7	***	***	***	***	**	n.s.	—	—
8	***	***	***	***	***	*	n.s.	—
9	*	*	n.s.	n.s.	n.s.	n.s.	*	***

Percent Error/Brief				Response Time/Brief				
	1	2	3	4	5	6	7	8
2	n.s.	—	—	—	—	—	—	—
3	n.s.	n.s.	—	—	—	—	—	—
4	n.s.	n.s.	n.s.	—	—	—	—	—
5	**	**	n.s.	n.s.	—	—	—	—
6	***	***	**	**	n.s.	—	—	—
7	***	***	**	***	n.s.	n.s.	—	—
8	***	***	**	***	*	n.s.	n.s.	—
9	***	***	***	***	***	***	***	***

Percent Error/Brief				Response Time/Combined <sup>a</sup>				
	1	2	3	4	5	6	7	8
2	n.s.	—	—	—	—	—	—	—
3	n.s.	n.s.	—	—	—	—	—	—
4	n.s.	n.s.	n.s.	—	—	—	—	—
5	n.s.	n.s.	n.s.	n.s.	—	—	—	—
6	**	**	**	n.s.	*	—	—	—
7	**	**	**	n.s.	*	n.s.	—	—
8	***	***	***	***	***	*	n.s.	—
9	***	***	***	***	***	***	***	n.s.

Note—n.s., not significant. <sup>a</sup>No significant exposure × number interaction (see text). \* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

ber 9 was presented [ $F(1,45) = 110.60, p < .001$ ]. The effect of number was significant for both exposure conditions [unlimited,  $F(8,80) = 9.20$ ; 100 msec,  $F(8,80) = 28.91$ ; all  $ps < .001$ ].

In contrast to errors, response times were stable at approximately 0.9 sec in the brief-exposure condition when the array size was larger than 6. A two-way (exposure × number) ANOVA showed significant main effects and a significant interaction [exposure,  $F(1,5) = 25.63, p < .01$ ; number,  $F(8,40) = 30.21, p < .001$ ; interaction,  $F(8,40) = 19.15, p < .001$ ]. The effect of exposure duration was significant when the array sizes were 7, 8, and 9 [7,  $F(1,45) = 17.54$ ; 8,  $F(1,45) = 151.67$ ; 9,  $F(1,45) = 11.60$ ; all  $ps < .001$ ].

The effect of number was significant for both exposure conditions [unlimited,  $F(8,80) = 66.60$ ; 100 msec,  $F(8,80) = 8.37$ ; all  $ps < .001$ ]. Table 1B shows the results of multiple comparisons. See also Appendices A–B for the results of distribution-free tests. The results of ANOVAs based on random permutations were the same as those of parametric versions. The results of Wicoxon's tests were also similar to the LSD results, with few exceptions.

**Humans.** The humans showed quite different patterns of results from the chimpanzee. As can be seen in Figure 4, they made more errors during the brief-exposure condition (13.2%, averaged across numerosities) than during the unlimited condition (2.2%). For response

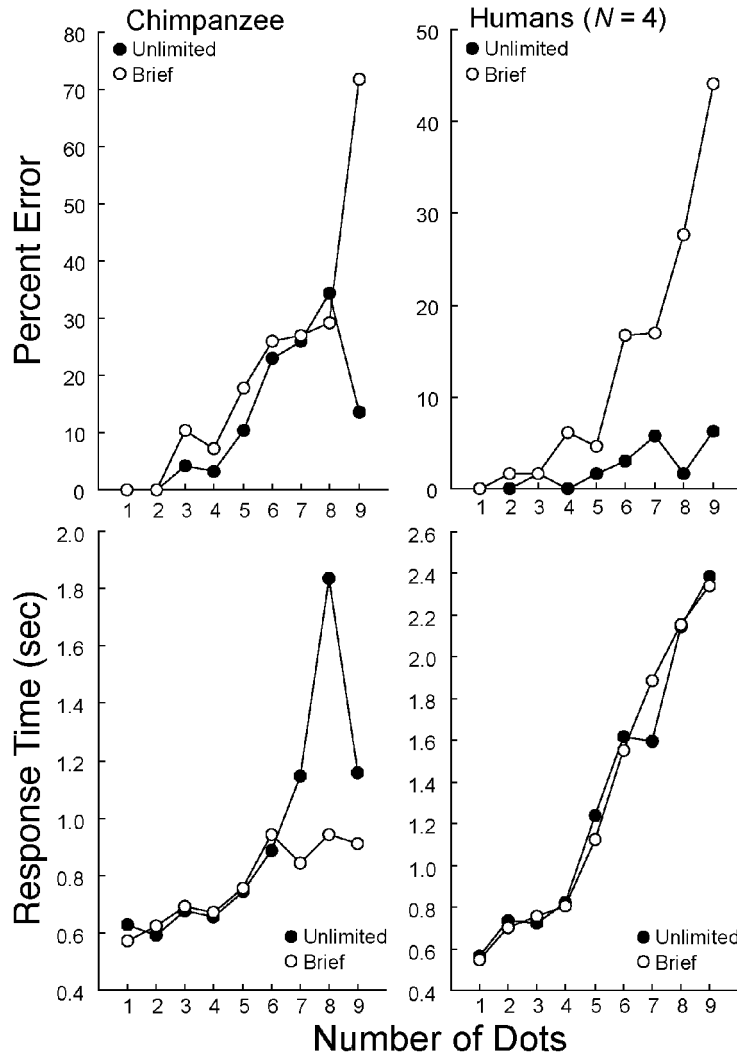


Figure 4. Mean percentages of errors and correct response times for Ai (left panels) and humans (right panels) during testing. Note that the vertical axes differ for the two species.

times, however, they showed no difference between exposure conditions. They showed quite fast response times from 1 to 4 and monotonic increases from 4 to 9, clearly suggesting a separation between subitizing and counting processes. Two-way ANOVAs, with subjects as repeated measures, revealed that all the main effects and the interaction were significant for percentage of error [exposure,  $F(1,3) = 50.05, p < .01$ ; number,  $F(8,24) = 9.12, p < .001$ ; interaction,  $F(8,24) = 6.68, p < .001$ ]. A simple main effect of exposure duration was significant when the array size was from 6 to 9 [all  $F_s(1,27) > 4.94, p_s < .05$ ]. The effect of number was significant only for the brief-exposure condition [ $F(8,48) = 17.11, p < .001$ ]. For response time, only the main effect of number was significant [ $F(8,24) = 37.78, p < .001$ ]. Table 1C shows the results of multiple comparisons.

**Response Time Distributions**

To examine the differences in response times between species and between exposure conditions further, we analyzed the distribution of response times for each numerosity across exposure conditions and species. The results are shown in Figure 5. To see the differences in response time distributions between numerosities and between exposure conditions, we conducted chi-square tests on the distribution data for adjacent numerosities in the same condition (e.g., “U1–2” in Figure 5) and on the data for the same numerosity in the different conditions (e.g., “U–B”). These results are also shown in the figure if they are statistically significant. Furthermore, we also calculated kurtosis and skewness for each response time distribution in order to quantify the shapes of the distributions. The results are shown in Figure 6. The kurtosis of the nor-

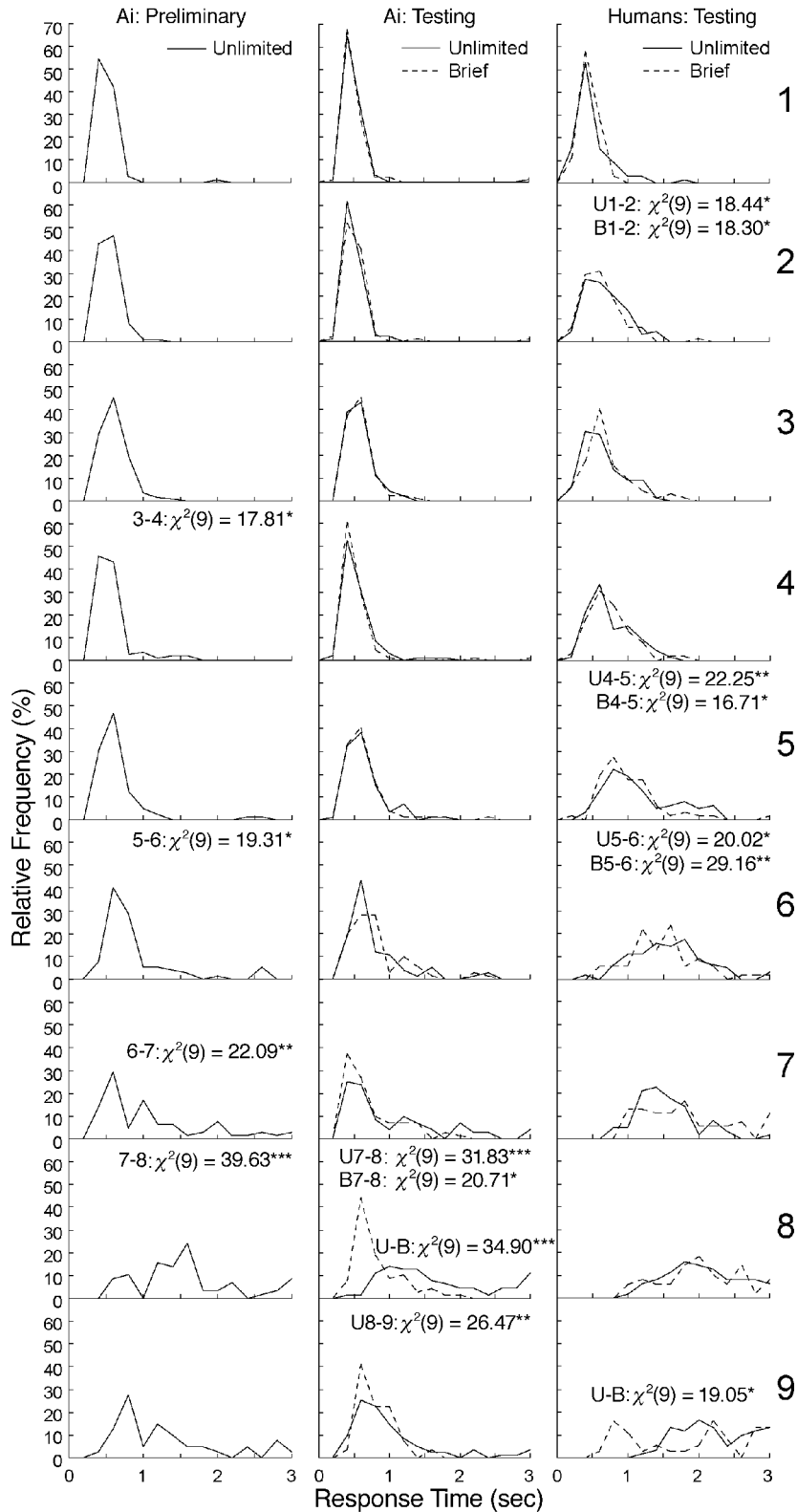


Figure 5. Response time distributions for preliminary training (left) and testing (middle) for Ai and for humans (right). Significant results of chi-square tests for the differences in distributions between adjacent numerosities or between exposure conditions with the same numerosity are also shown. The degree of freedom of each test is 9. \* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .



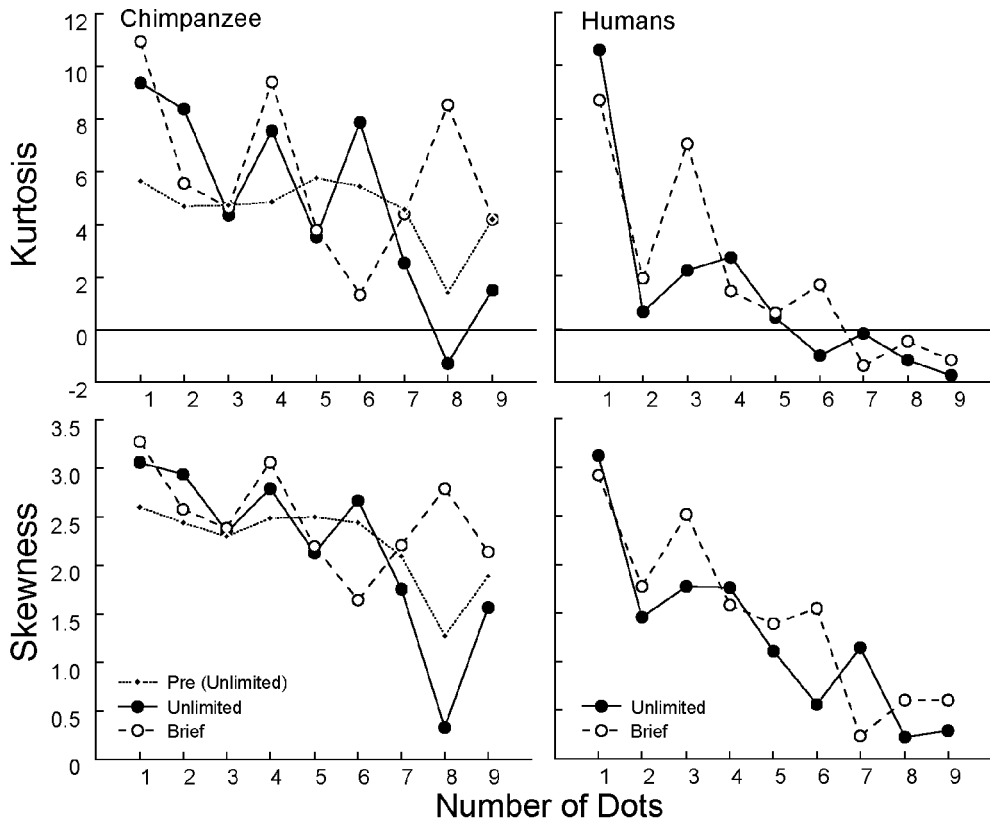


Figure 6. Kurtosis and skewness for each response time distribution shown in Figure 5 as a function of species, exposure condition, and numerosity.

mal distribution is 0. Distributions that are more leptokurtic than the normal distribution have positive kurtosis, whereas distributions that are more platykurtic have negative kurtosis. Skewness is a measure of the asymmetry of the data around the sample mean. If the skewness is negative, the data are spread out more to the left of the mean than to the right, whereas the skewness is positive when the data are spread out more to the right. Zero indicates a normal distribution.

For humans, the distribution became flatter in both conditions for the numerosities from 5 to 9. This pattern was also seen in the unlimited-exposure condition for the chimpanzee. In contrast, the peak shifted from 0.6 sec to larger when the number of dots increased from 4 to 6, which differed from the chimpanzee's results. Accordingly, skewness shifted from larger to smaller, indicating asymmetrical to symmetrical distributions (see the lower right panel of Figure 6). There were no differences between exposure conditions. Chi-square tests also verified these visual inspections.

For Ai, on the other hand, each of the distributions had a peak at around 0.6 sec for any number, which was verified by a stable and positive skewness (see the lower left panel of Figure 6). When the number increased to greater than 4, the relative frequency at the peak gradually de-

creased, and the distribution became flatter at the numbers 7 and 8 for the unlimited conditions (see also the decrease of kurtosis in the upper left panel of Figure 6). In the brief-exposure condition, however, the peak at 0.6 sec did not disappear until the number 9 (see also the lower left panel of Figure 6, showing a stable and positive skewness for the brief-exposure condition, and the high kurtosis in the upper left panel of Figure 6). A chi-square test also showed a significant difference in distribution between the unlimited- and the brief-exposure conditions at the number 8.

**Looking-Back Responses by the Chimpanzee**

Figure 7 shows the mean number of looking-back responses by the chimpanzee per each trial during the unlimited-exposure condition of the video-recorded session (filled square). Note that she made no looking-back responses in the brief-exposure trials. In the unlimited-exposure trials (141 trials in total), she made no looking-back responses in 73.1% of the trials (103 trials), one in 19.1% (27 trials), two in 7.1% (10), and three in 0.7% (1 trial). Furthermore, the mean number of looking-back responses per each trial was zero for the numbers from 1 to 3, monotonically increased for numbers from 4 to 8, and decreased for numbers from 8 to 9, showing the same pattern as percentage of error and response times.

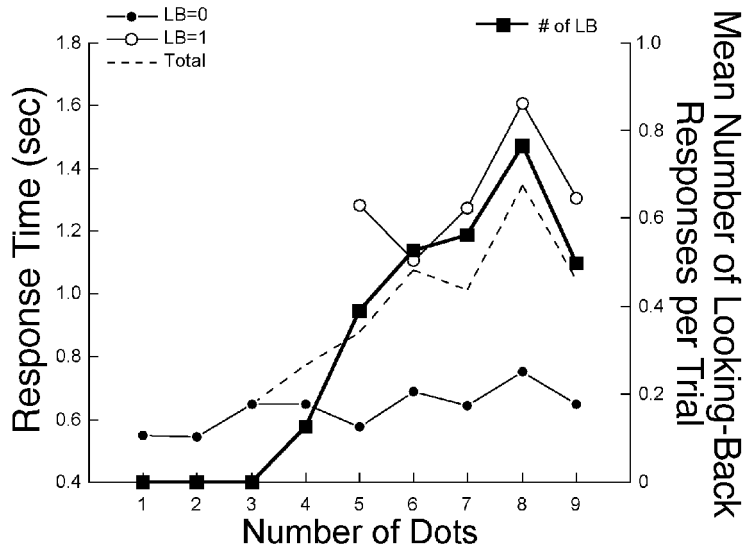


Figure 7. Response times (dashed lines and circles, left axis) and the number of looking-back responses (filled squares, right axis) for A1 during the single video-recorded session. Filled circle, response times when the subject made no looking-back responses; open circle, response times when the subject made one looking-back response.

### Relationship Between Looking-Back Responses and Response Times in the Chimpanzee

Figure 7 also shows the response times in the unlimited-exposure condition of the video-recorded session for the chimpanzee. Filled circles indicate the response times when she made no looking-back responses, whereas open circles indicate response times when she made a single looking-back response. There are no statistical differences in response times among numerosities both when she made no response (0.61 sec, averaged across 5–9;  $p = .155$ ; one-way ANOVA based on 5,000 random permutations, using the same procedure as that described in Appendix A; Edgington, 1987) and when she made one looking-back response (1.31 sec, averaged across 5–9;  $p = .203$ ; one-way ANOVA based on 5,000 random permutations). These results imply that the increase in response times shown in Figures 3 and 4 was due mainly to the increase in the number of looking-back responses (see the Discussion section for details).

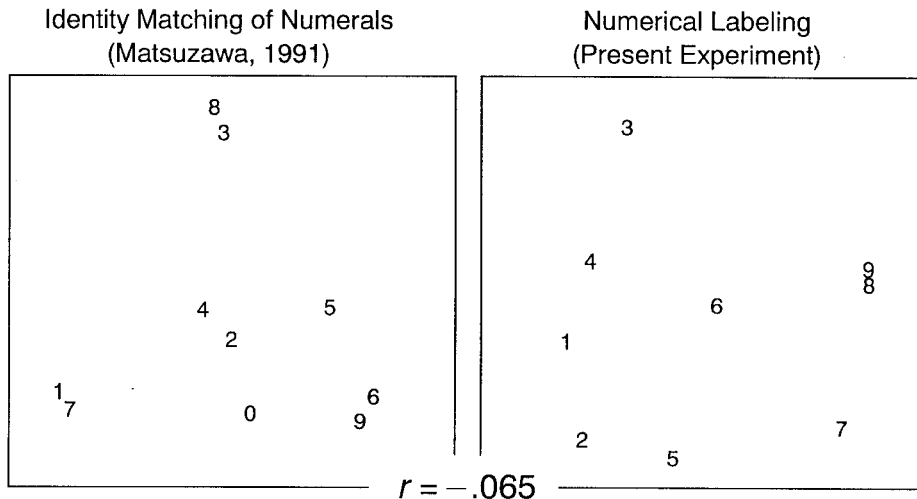
### Error Pattern Analyses for the Chimpanzee Data

In the present experiment, the spatial arrangement of the numeral display changed randomly from trial to trial. Thus, response time not only reflected the time for the enumeration processes, but also might be significantly biased by the time spent searching for the correct numeral in the numeral display. Since the correct numeral appeared randomly and with equal likelihood at each predefined location from trial to trial, the biases of manual movements among positions were counterbalanced. However, if the subject showed perceptual confusions among numerals,

the time spent searching might influence the overall response time differentially for each numerosity. Actually, at the initial phase of the introduction of the numerals, which had been more than 5 years before the present experiment, the subject had shown perceptual confusions among numerals during training in identity matching of numerals (Matsuzawa, 1991). Multidimensional scaling (MDS) is frequently used to provide a visual representation of the pattern of proximities (e.g., perceived similarities or distances) among a set of objects. MDS is a set of multivariate statistical methods for estimating the parameters for and assessing the fit of various spatial distance models for proximity data (Davison, 1992).

The results of the MDS analysis, based on the errors during the initial identity-matching training, indicated that the subject had shown *perceptual* confusion between the numerals 1 and 7, 6 and 9, 3 and 8, and 2 and 4 (see Figure 8; Matsuzawa, 1991). After extensive training in numerical matching, however, these perceptual confusions were eliminated, and *numerical* confusions became stronger instead.

We also conducted an MDS analysis (MDSCAL; Davison, 1992) based on the confusion matrix of errors for all the sessions of preliminary training and testing in the present experiment for the chimpanzee. The resulting two-dimensional configuration is shown in the right panel of Figure 8. The stress value, a measure for goodness of fit between the spatial configuration and the actual data, was 0.117. Clearly, the subject showed numerical confusions (e.g., between 8 and 9), but not perceptual ones. On the basis of all the possible distances between numerals in the MDS spaces ( $n = 36$ ), we calculated the correlation coefficients and obtained no correlation between the two MDS



**Figure 8.** Two-dimensional spatial arrangements of multidimensional scaling analyses based on confusion matrices obtained from identity matching of numerals during the initial phase of numerical labeling training, approximately 5 years before the present experiment (Matsuzawa, 1991), and the present numerical labeling experiment by the chimpanzee.

results ( $r = -.065$ ). These results clearly suggest that perceptual-based bias of searching times was not the critical factor for the differences in response times among numerosities in the present experiment.

## DISCUSSION

The present experiment was conducted in order to clarify the enumeration processes in the chimpanzee, Ai, who had acquired matching between number of dots and corresponding Arabic numerals from 1 to 9. In a previous study, Murofushi (1997) concluded that Ai did not *count* numbers larger than 4 or 5 but estimated the numerosity of the array and used analogue magnitude codes to match the numeral to the numerosity. We tested her ability of enumeration by presenting the item array for brief exposure time (approximately 100 msec) and compared the results with those from humans who were tested in an identical experimental setting. The results indicated that the chimpanzee Ai did not show any sign of *counting* when the items were terminated after a very brief time.

The human subjects showed almost the same patterns of response times between the two exposure time conditions, whereas they showed an increase in errors when the exposure time was 100 msec. Dissociation between error rate and response times in humans might suggest that these two measures reflect different aspects of the enumeration processes. Humans might *count* items in the same way when the representation of items was maintained in their mind, as in the unlimited condition. However, decay of the representations or misencoding of the array might increase as a function of array size, resulting in an increase in the error rate. Our human subjects showed no evidence for estimation in the brief-exposure condition, unlike in the studies by Kaufman et al. (1949) and Mandler and

Shebo (1982), in which human subjects showed ceiling effects of response times when the number was larger than 6. The reason is unclear. It might be due to the difference in experimental conditions. Another possibility can be derived from the response time distributions shown in Figures 5 and 6. The distribution becomes flatter from 5 to 9 in both exposure conditions. This tendency seems stronger for the brief-exposure condition. The human subjects might utilize both estimation and mental counting strategies. This possibility should be further examined in the future.

For the chimpanzee Ai, accuracy did not differ between the brief- and the unlimited-exposure conditions from one to eight dots, but she showed severe impairment when nine dots were presented briefly. She also showed differences in response time between exposure conditions. Her response times were stable from 6 to 9 in the brief-exposure condition. She showed an improvement of performance for the largest number in the unlimited-exposure condition, but not in the brief-exposure condition.

Murofushi (1997) pointed out that the differential improvement in performance for the largest number implied relative magnitude estimation. This effect was sometimes observed in experiments with humans (e.g., Balakrishnan & Ashby, 1991; Simon & Vaishnavi, 1996). When human subjects know the largest number of items in a given experimental setting, they often make relative magnitude estimations when presented with a dense display. This might be the case also for the chimpanzee in the present experiment. Therefore, the improvement for the largest number *itself* could not be evidence against counting from 4 or 5 to the second largest number. It is quite possible that only the largest number might be processed differently.

During the video-recorded session, Ai increased her looking-back responses from 4 to 8. Therefore, the increase in response times in the unlimited condition might

be due to the increase in the number of looking-back responses. These responses might also be related to eye movements during enumeration. The relationship between counting and eye movements has been extensively investigated in humans (Noro, 1980, 1983; Schlingensiepen, Campbell, Legge, & Walker, 1986; Van Oeffelen & Vos, 1984a, 1984b). Among these studies, Van Oeffelen and Vos (1984b) recorded eye movements during enumeration in human children. When the number of dots was larger than five, children increased the number of fixations and saccades. When the array was arranged in two or three groups of dots, they scanned each group repeatedly. Van Oeffelen and Vos (1984b) suggested that such a repeated scanning was due mainly to interim verifications of results already established. The children were mentally loaded when storing partial results and processing new information at the same time. It may be plausible that Ai used a strategy similar to that of human children during the *counting* range, even though her response time was faster than that for humans during the counting range. Ai enumerated the dots but also verified the confidence of her results by looking back at the item array.

If looking-back responses function as interim verification of the enumeration process, does a greater number of looking-back responses result in an improvement in accuracy? The answer was no. In the analyses of looking-back responses shown in Figure 7, we combined the data from correct and error trials (see the Procedure section). When the correct ( $N = 115$ ) and error ( $N = 26$ ) trials were separated, the chimpanzee made one or more looking-back responses in 21.7% of the correct trials ( $N = 25$ ) and 50.0% of the error trials ( $N = 13$ ). A chi-square test revealed that the subject showed looking-back responses more often for error trials than for correct trials [ $\chi^2(1) = 8.6, p < .01$ ]. This additional analysis strongly implies that the looking-back response can be viewed as verifying confidence in the results of enumeration. That is, a greater number of looking-back responses might imply a low confidence in the results. In human experiments, the measure of confidence is frequently used (Mandler & Shebo, 1982). We propose that looking-back responses made by the chimpanzee are also a measure of confidence for enumeration.

How did Ai enumerate the numerosity, counting or estimation? The present results are favorable to the estimation account. Direct evidence is shown in Figure 7. When she made no looking-back responses, her response time was stable irrespective of the number of dots presented (0.61 sec, on average). This value was the same as the peak of response time distributions (see Figure 5). Furthermore, when she made a single looking-back response, there were also no statistical differences in response times among numerosities. These results again suggest that the gradual increase in response times from 4 to 8 may be due to the increase of the proportion of looking-back responses.

Inspections of response time distributions further supported this account. If the response time should reflect the time for counting, the peak of response time distribution would shift from smaller to larger values along numerosity. This is the case for humans. The chimpanzee's distri-

butions showed completely different patterns from humans, especially at larger numerosities. The peak was not shifted in the chimpanzee, unlike in humans. When the numbers were 7 and 8, the chimpanzee's response times had a more platykurtic and flatter distribution in the unlimited-exposure condition. These analyses are consistent with the results shown in Figure 7. Stable peaks (consistent positive values of skewness) indicate that the time for single estimation was unchanged across numerosities. During preliminary training and unlimited-exposure testing, Ai's response time reflected the number of looking-back responses, but not the time for counting. As was noted above, human children have shown frequent eye movements when a number was more than 5 (Van Oeffelen & Vos, 1984b). This was due to verification of the previous results of counting. For Ai, looking-back responses may reflect verification of the previous results of *estimation*.

In the brief-exposure condition, the peak at 0.6 sec did not disappear until the number 9. This strongly implies that she made only estimations without verification responses, but not "mental" counting, from 5 to 9 in the brief-exposure trials.

In sum, the present results further support the conclusion of Murofushi (1997) that the chimpanzee, who had been trained to match the numerals to the number of dots, did not use a one-by-one counting strategy but used an estimation of numerosity. From our results, however, it is still unclear what types of codes she used. Murofushi suggested that she used analogue magnitude codes. Further examination is necessary to address this question.

Note that the present conclusion cannot be expanded to a general numerical ability of chimpanzees. It is quite plausible that training history, training procedure, task demand, and cognitive load easily biased the numerical processing (Murofushi, 1997). If we trained chimpanzees as human children are taught to count (cf. Gelman & Gallistel, 1978)—for example, if we trained ordinality first and then cardinality (cf. Biro & Matsuzawa, 1999, 2001a, 2001b; Kawai, 2001; Kawai & Matsuzawa, 2000, 2001; Rumbaugh & Washburn, 1993; Tomonaga & Matsuzawa, 2000; Tomonaga, Itakura, & Matsuzawa, 1993)—the one-by-one counting strategy might emerge (Beran, Rumbaugh, & Savage-Rumbaugh, 1998; Boysen, 1993; Rumbaugh, 1990; Rumbaugh, Hopkins, Washburn, & Savage-Rumbaugh, 1989; Rumbaugh & Washburn, 1993).

Recently, counting in humans has been considered to be a limited-capacity (or attentional) process (Pylyshyn, 1989, 1994; Trick & Pylyshyn, 1993, 1994a, 1994b). If the chimpanzee in the present experiment processes numerosity by using a strategy different from counting (probably by estimation), what are the implications for attention? Does she also use attentional processing when estimating a number greater than 5 (cf. Pepperberg, 1994)? Further examination is necessary to address these questions.

## REFERENCES

- BALAKRISHNAN, J. D., & ASHBY, F. G. (1991). Is subitizing a unique numerical ability? *Perception & Psychophysics*, *50*, 555-564.  
 BARLOW, D. H., & HERSEN, M. (1984). *Single case experimental designs*. New York: Pergamon.

- BERAN, M. J., RUMBAUGH, D. M., & SAVAGE-RUMBAUGH, E. S. (1998). Chimpanzee (*Pan troglodytes*) counting in a computerized testing paradigm. *Psychological Record*, **48**, 3-19.
- BIRO, D., & MATSUZAWA, T. (1999). Numerical ordering in a chimpanzee (*Pan troglodytes*): Planning, executing, and monitoring. *Journal of Comparative Psychology*, **113**, 178-185.
- BIRO, D., & MATSUZAWA, T. (2001a). Chimpanzee numerical competence: Cardinal and ordinal skills. In T. Matsuzawa (Ed.), *Primate origins of human cognition and behavior* (pp. 199-225). Tokyo: Springer-Verlag.
- BIRO, D., & MATSUZAWA, T. (2001b). Use of numerical symbols by the chimpanzee (*Pan troglodytes*): Cardinals, ordinals, and the introduction of zero. *Animal Cognition*, **4**, 193-199.
- BOYSEN, S. T. (1993). Counting in chimpanzees: Nonhuman principles and emergent properties of number. In S. T. Boysen & E. J. Capaldi (Eds.), *The development of numerical competence: Animal and human models* (pp. 39-59). Hillsdale, NJ: Erlbaum.
- BOYSEN, S. T., & BERNTSON, 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.
- CAPALDI, E. J., & MILLER, D. J. (1988). Counting in rats: Its functional significance and the independent cognitive processes that constitute it. *Journal of Experimental Psychology: Animal Behavior Processes*, **14**, 3-17.
- DAVIS, H. (1984). Discrimination of the number three by a raccoon (*Procyon lotor*). *Animal Learning & Behavior*, **12**, 409-413.
- DAVIS, H., & MEMMOTT, J. (1982). Counting behavior in animals: A critical evaluation. *Psychological Bulletin*, **92**, 547-571.
- DAVIS, H., & PERUSSE, R. (1988). Numerical competence in animals: Definitional issues, current evidence, and a new research agenda. *Behavioral & Brain Sciences*, **11**, 561-615.
- DAVISON, M. L. (1992). *Multidimensional scaling*. Malabar, FL: Krieger.
- DEHAENE, S. (1992). Varieties of numerical abilities. *Cognition*, **44**, 1-42.
- DOOLEY, G. B., & GILL, I. V. (1977). Acquisition and use of mathematical skills by a linguistic chimpanzee. In D. M. Rumbaugh (Ed.), *Language learning by a chimpanzee: The Lana project* (pp. 247-260). New York: Academic Press.
- EDGINGTON, E. S. (1987). *Randomization tests* (2nd ed.). New York: Dekker.
- FERSTER, C. B., & HAMMER, C. E., JR. (1966). Synthesizing the components of arithmetic behavior. In W. K. Honig (Ed.), *Operant behavior* (pp. 634-676). New York: Appleton.
- FOLK, C. L., EGETH, H. E., & KWAK, H.-W. (1988). Subitizing: Direct apprehension or serial processing? *Perception & Psychophysics*, **44**, 313-320.
- 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 understanding of number*. Cambridge, MA: Harvard University Press.
- HAYES, K. J., & NISSEN, C. H. (1971). Higher mental functions of a home-raised chimpanzee. In A. M. Schrier & F. Stollnitz (Eds.), *Behavior of nonhuman primates* (Vol. 4, pp. 59-115). New York: Academic Press.
- HICKS, L. H. (1956). An analysis of number-concept formation in the rhesus monkey. *Journal of Comparative & Physiological Psychology*, **49**, 212-218.
- HONIG, W. K., & STEWART, K. E. (1989). Discrimination of relative numerosity by pigeons. *Animal Learning & Behavior*, **17**, 134-146.
- JENSEN, E. M., REESE, E. P., & REESE, T. W. (1950). The subitizing and counting of visually presented fields of dots. *Journal of Psychology*, **30**, 363-392.
- JEVONS, W. S. (1871). The power of numerical discrimination. *Nature*, **3**, 281-282.
- KAUFMAN, E. L., LORD, M. W., REESE, T. W., & VOLKMAN, J. (1949). The discrimination of visual number. *American Journal of Psychology*, **62**, 498-525.
- KAWAI, N. (2001). Ordering and planning in sequential responding to Arabic numerals by a chimpanzee (*Pan troglodytes*). *Psychologia*, **44**, 60-69.
- KAWAI, N., & MATSUZAWA, T. (2000). Numerical memory span in a chimpanzee. *Nature*, **403**, 39-40.
- KAWAI, N., & MATSUZAWA, T. (2001). Reproductive memory processes in chimpanzees: Homologous approaches to research on human working memory. In T. Matsuzawa (Ed.), *Primate origins of human cognition and behavior* (pp. 226-234). Tokyo: Springer-Verlag.
- KIRK, R. E. (1995). *Experimental design* (3rd ed.). Pacific Grove, CA: Brooks/Cole.
- KLAHR, D. (1973). Quantification processes. In W. G. Chase (Ed.), *Visual information processing* (pp. 3-34). New York: Academic Press.
- MANDLER, G., & SHEBO, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General*, **111**, 1-22.
- MANLY, B. F. J. (1997). *Randomization, bootstrap and Monte Carlo methods in biology* (2nd ed.). New York: Chapman & Hill.
- MATSUZAWA, T. (1985). Use of numbers by a chimpanzee. *Nature*, **315**, 57-59.
- MATSUZAWA, T. (1991). *The visual world of a chimpanzee*. Tokyo, Japan: University of Tokyo Press. (Japanese text)
- MATSUZAWA, T., ASANO, T., KUBOTA, K., & MUROFUSHI, K. (1986). Acquisition and generalization of numerical labeling by a chimpanzee. In D. M. Taub & F. A. King (Eds.), *Current perspectives in primate social dynamics* (pp. 416-430). New York: Van Nostrand Reinhold.
- MATSUZAWA, T., ITAKURA, S., & TOMONAGA, M. (1991). Use of numbers by a chimpanzee: A further study. In A. Ehara, T. Kimura, O. Takenaka, & M. Iwamoto (Eds.), *Primate today* (pp. 317-320). Amsterdam: Elsevier.
- MILLER, D. J. (1993). Do animals subitize? In S. T. Boysen & E. J. Capaldi (Eds.), *The development of numerical competence: Animal and human models* (pp. 149-169). Hillsdale, NJ: Erlbaum.
- MUROFUSHI, K. (1997). Numerical matching behavior by a chimpanzee (*Pan troglodytes*): Subitizing and analogue magnitude estimation. *Japanese Psychological Research*, **39**, 140-153.
- NORO, K. (1980). Determination of counting time in visual inspection. *Human Factors*, **22**, 43-55.
- NORO, K. (1983). A descriptive model of visual search. *Human Factors*, **25**, 93-101.
- OYAMA, T., KIKUCHI, T., & ICHIHARA, S. (1981). Span of attention, backward masking, and reaction time. *Perception & Psychophysics*, **29**, 106-112.
- PEPPERBERG, I. M. (1987). Evidence for conceptual quantitative abilities in the African gray parrot: Labeling of cardinal sets. *Ethology*, **75**, 37-61.
- PEPPERBERG, I. M. (1994). Numerical competence in an African gray parrot (*Psittacus erithacus*). *Journal of Comparative Psychology*, **108**, 36-44.
- PYLYSHYN, Z. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, **32**, 65-97.
- PYLYSHYN, Z. (1994). Some primitive mechanisms of spatial attention. *Cognition*, **50**, 363-384.
- RUMBAUGH, D. M. (1990). Comparative psychology and the great apes: Their competency in learning, language, and numbers. *Psychological Record*, **40**, 15-39.
- RUMBAUGH, D. M., HOPKINS, W. D., WASHBURN, D. A., & SAVAGE-RUMBAUGH, E. S. (1989). Lana chimpanzee learns to count by "Numath": A summary of a video-taped experimental report. *Psychological Record*, **39**, 459-470.
- 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.
- SCHLINGENSIEPEN, K. H., CAMPBELL, F. W., LEGGE, G. E., & WALKER, T. D. (1986). The importance of eye movements in the analysis of simple patterns. *Vision Research*, **26**, 1111-1117.
- SIMON, T. J., & VAISHNAVI, S. (1996). Subitizing and counting depend on different attentional mechanisms: Evidence from visual enumeration in afterimages. *Perception & Psychophysics*, **58**, 915-926.
- SUZUKI, K., & KOBAYASHI, T. (2000). Numerical competence in rats (*Rattus norvegicus*): Davis and Bradford (1986) extended. *Journal of Comparative Psychology*, **114**, 73-85.
- THOMAS, R. K., FOWLKES, D., & VICKERY, J. D. (1980). Conceptual numerosness judgements by squirrel monkeys. *American Journal of Psychology*, **93**, 247-257.

TOMONAGA, M., ITAKURA, S., & MATSUZAWA, T. (1993). Teaching ordinal to a cardinal-trained chimpanzee. *Primate Research*, **9**, 67-77. (Japanese text with English summary)

TOMONAGA, M., & MATSUZAWA, T. (1992, September). *Processing of number by the chimpanzee: Effect of the brief exposure of items*. Paper presented at the 56th Annual Meeting of Japanese Psychological Association, Kyoto.

TOMONAGA, M., & MATSUZAWA, T. (2000). Sequential responding to Arabic numerals with wild cards by the chimpanzee (*Pan troglodytes*). *Animal Cognition*, **3**, 1-11.

TRICK, L. M., & PYLYSHYN, Z. W. (1993). What enumeration studies can show us about spatial attention: Evidence for limited capacity preattentive processing. *Journal of Experimental Psychology: Human Perception & Performance*, **19**, 331-351.

TRICK, L. M., & PYLYSHYN, Z. W. (1994a). Cueing and counting: Does the position of the attentional focus affect enumeration? *Visual Cognition*, **1**, 67-100.

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

VAN OEFFELEN, M. P., & VOS, P. G. (1984a). Enumeration of dots: An eye movement analysis. *Memory & Cognition*, **12**, 607-612.

VAN OEFFELEN, M. P., & VOS, P. G. (1984b). The young child's processing of dot patterns: A chronometric and eye movement analysis. *International Journal of Behavioral Development*, **7**, 53-66.

WATANABE, S. (1998). Discrimination of "four" and "two" by pigeons. *Psychological Record*, **48**, 383-391.

WOODRUFF, G., & PREMACK, D. (1981). Primitive mathematical concepts in the chimpanzee: Proportionality and numerosity. *Nature*, **293**, 568-570.

WRIGHT, A. A. (1990). Markov choice processes in simultaneous matching-to-sample at different levels of discriminability. *Animal Learning & Behavior*, **18**, 277-286.

WRIGHT, A. A. (1992). Learning mechanisms in matching to sample. *Journal of Experimental Psychology: Animal Behavior Processes*, **18**, 67-79.

**APPENDIX A**  
**Results of ANOVAs for the Chimpanzee Data Based on the Random Permutations**

The randomization test is a powerful technique for distribution-free tests (Edgington, 1987; Manly, 1997). Generally, the null hypothesis in the randomization test is stated as follows: "The measurement (or set of measurements) associated with each experimental unit is independent of the assignment of units to treatments" (Edgington, 1987, p. 39). This technique is also applicable to the data from factorial designs. A randomization-based ANOVA can be done with the following procedures (Manly, 1997, pp. 120-121).

1. Calculate the  $F$ -statistic for the observed data ( $F_1$ ) in the usual ANOVA.
2. Randomly allocate the observed data (or residuals, which are the differences between the observed data and the sample mean) to groups and calculate  $F_2$ , the  $F$ -statistic from an ANOVA on the randomly allocated data.
3. Repeat Step 2  $N$  times (a large enough number) to generate values  $F_2, F_3, \dots, F_N$ .
4. Declare  $F_1$  to be significantly large at a given  $100\alpha\%$  level (e.g., 5%) if it exceeds  $100(1-\alpha)\%$  of the values  $F_1, F_2, F_3, \dots, F_N$ .

In the analyses shown in Table A1, we also calculated the  $F$ -statistics for the simple main effects (Kirk, 1995) and conducted the same analyses as those for the other main factors and interactions. If the interaction is significant on the basis of the randomization test, the results of the simple main effects are also shown in the table. Each analysis is based on 5,000 random permutations.

**Table A1**

Factor	df	Percentage of Error		Response Time	
		F(Parametric)	Probability	F(Parametric)	Probability
Preliminary training					
Number	8,48	12.12	.000	26.93	.000
Testing					
Exposure	1,5	98.18	.005**	25.62	.005**
E at N1	1,45	0.00	.958	0.63	.433
E at N2	1,45	0.00	.957	0.20	.652
E at N3	1,45	1.27	.274	0.04	.825
E at N4	1,45	0.56	.464	0.04	.836
E at N5	1,45	1.73	.215	0.02	.887
E at N6	1,45	0.32	.584	0.63	.430
E at N7	1,45	0.04	.849	17.53	.001***
E at N8	1,45	0.88	.366	151.65	.000***
E at N9	1,45	110.60	.002**	11.60	.000***
Number	8,40	34.02	.000***	30.21	.000***
N at E1	8,80	10.21	.000***	46.77	.000***
N at E2	8,80	32.07	.000***	5.88	.000***
Exposure × number	8,40	10.81	.036*	19.14	.000***

Note—Probabilities are based on a randomization test. \* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

APPENDIX B

Results of Wilcoxon's Signed-Rank Tests Conducted for the Chimpanzee Data

We also conducted the Wilcoxon's signed-rank tests for the paired data to those from the chimpanzee. These are the nonparametric versions of the parametric multiple comparisons shown in Table 1. Only the significance levels are presented in Appendix B.

A. Preliminary Training With Unlimited Exposure Condition for Ai

Percent Error									Response Time									
	1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8	
2	NA								2	*								
3	NA	n.s.							3	*	*							
4	n.s.	NA	n.s.						4	n.s.	†	n.s.						
5	NA	NA	n.s.	n.s.					5	*	*	n.s.	†					
6	n.s.	n.s.	n.s.	n.s.	n.s.				6	*	*	*	*	†				
7	*	*	*	*	*	*			7	*	*	*	*	*	n.s.			
8	*	*	*	*	*	*	*		8	*	*	*	*	*	*	*	*	
9	*	*	*	*	†	†	n.s.	NA	9	*	*	*	*	*	*	*	n.s.	*

B. Test for Ai

Percent Error/Unlimited									Response Time/Unlimited								
	1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8
2	NA								2	n.s.							
3	NA	NA							3	n.s.	*						
4	n.s.	n.s.	n.s.						4	n.s.	*	n.s.					
5	†	†	n.s.	n.s.					5	*	*	†	+				
6	*	*	*	*	n.s.				6	*	*	*	*	n.s.			
7	*	*	*	*	n.s.	n.s.			7	*	*	*	*	*	n.s.		
8	*	*	*	*	+	n.s.	n.s.		8	*	*	*	*	*	*	*	*
9	*	*	n.s.	†	n.s.	n.s.	n.s.	†	9	*	*	*	*	*	†	n.s.	*

Percent Error/Brief									Response Time/Brief									
	1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8	
2	NA								2	n.s.								
3	†	†							3	*	n.s.							
4	*	*	n.s.						4	*	n.s.	n.s.						
5	*	*	†	n.s.					5	*	*	*	†					
6	*	*	n.s.	+	n.s.				6	*	*	*	*	*				
7	*	*	n.s.	*	n.s.	n.s.			7	*	*	n.s.	n.s.	n.s.	n.s.			
8	*	*	†	*	n.s.	n.s.	NA		8	*	*	*	*	*	n.s.	n.s.		
9	*	*	*	*	*	*	*	*	9	*	*	*	*	*	*	n.s.	n.s.	n.s.

Note—NA, test was not applicable because of many ties; n.s., not significant. \* $p < .05$ . † $p < .1$ .

(Manuscript received August 4, 2001;  
accepted for publication November 5, 2001.)