

Analog Numerical Representations in Rhesus Monkeys: Evidence for Parallel Processing

Andreas Nieder and Earl K. Miller

Abstract

■ Monkeys have been introduced as model organisms to study neural correlates of numerical competence, but many of the behavioral characteristics of numerical judgments remain speculative. Thus, we analyzed the behavioral performance of two rhesus monkeys judging the numerosities 1 to 7 during a delayed match-to-sample task. The monkeys showed similar discrimination performance irrespective of the exact physical appearance of the stimuli, confirming that performance was based on numerical information. Performance declined smoothly with larger numerosities, and reached discrimination threshold at numerosity “4.” The nonverbal numerical representations in monkeys were based on analog magnitudes, object tracking process (“subitizing”) could not account for

the findings because the continuum of small and large numbers shows a clear Weber fraction signature. The lack of additional scanning eye movements with increasing set sizes, together with indistinguishable neuronal response latencies for neurons with different preferred numerosities, argues for parallel encoding of numerical information. The slight but significant increase in reaction time with increasing numerosities can be explained by task difficulty and consequently time-consuming decision processes. The behavioral results are compared to single-cell recordings from the prefrontal cortex in the same subjects. Models for numerosity discrimination that may account for these results are discussed. ■

INTRODUCTION

Evidence for numerical competence in animals has been obtained for many different species over the past decades. Mammals (reviewed by Boysen & Capaldi, 1993), birds (reviewed by Emmerton, 2001), and amphibians (Uller, Jaeger, Guidry, & Martin, 2003) can discriminate stimuli based on the number of items. These studies indicate that rudimentary numerical competence may be widely spread in the animal kingdom and support the hypothesis of phylogenetic precursor system(s) for higher, verbal-based numerical abilities in adult humans (Gallistel & Gelman, 1992). In fact, developmental psychologists showed that preverbal human infants at the age of several months are already able to distinguish sets of small numerosities and perform simple numerical computations (e.g., Wynn, Bloom, & Chiang, 2002; Xu & Spelke, 2000; Xu, 2003) (see also Feigenson, Carey, & Spelke, 2002, for a critical evaluation of older studies). Thus, a better understanding of nonverbal numerical abilities in animals can yield insight into the more advanced human numerical abilities.

Monkeys are excellent model organisms to study numerosity judgments. Several studies showed that macaque monkeys are endowed with considerable numerical competence. Brannon and Terrace (1998, 2000)

demonstrated that rhesus monkeys could distinguish between sets of visual elements on the basis of number alone. Even more, the monkeys were able to successfully represent the ordinal relations between the numbers 1 to 9. Elementary arithmetic abilities comparable to human babies (Wynn, 1992) have been reported for wild rhesus monkeys by Hauser, MacNeilage, and Ware (1996) and Hauser, Carey, and Hauser (2000); the monkeys were able to detect simple additive and subtractive changes in the number of objects. Because of the comparatively well-understood neural structures of the primate brain and the relative similarity between the monkey and human brain, macaques constitute an ideal model organism to investigate the neural substrates and mechanisms underlying numerical competence (Nieder, Freedman, & Miller, 2002; Nieder & Miller, 2003; Ninkura, Mushiake, & Tanji, 2003, 2004; Sawamura, Shima, & Tanji, 2002).

To date, it is widely assumed that there are basically two nonverbal systems for representing numerosity in animals, human infants, and adults (for a review, see Carey, 2001). One is an object tracking system, which yields relatively discrete representations (also termed “subitizing” in an older stipulation). This system keeps track of a small number of items by assigning markers (or pointers) to individual elements. Because each individual object has its own symbol (“file”), such representations are called object-file representations. These symbols (or

“tokens”) are derived from object segregation processes that take place in parallel in a preattentive stage of vision (Pylyshyn, 2001; Treisman, 1992). Based on the limited number of markers, this system can only represent up to four items (set-size limit); it fails for larger numbers. For larger numbers, an analog magnitude system (“estimation system”) is applied that has no immediate upper limit, but does become systematically less precise with increasing numbers. Thus, the hallmark of analog magnitude representations is that they obey Weber’s Law.

In a previous publication, we demonstrated that monkeys engaged in a delayed match-to-numerosity task formed analog numerical representations (Nieder & Miller, 2003). Both the behavioral and neural data obeyed Weber’s Law (the just noticeable difference between two numerosities becomes larger in proportion to increasing numerical values). This clear Weber fraction signature is not compatible with nonnumerical “object-file” representations, which would show a strict limit of quantities that can be discriminated.

Knowing the representational system the monkeys in our experiments used, it became possible to explore numerical processing in more detail. One fundamental question is whether monkeys employ a serial counting-like process when confronted with a number of simultaneously presented items. If so, we would assume that they serially scan the elements of visual displays when discriminating their number (Brannon & Terrace, 2000). Alternatively, the monkeys could use a parallel enumeration process that does not depend on serial scanning eye movements. So far, the issue of serial scanning has not been addressed, because prior studies have not measured eye movements.

A second important issue is related to the ubiquitous finding that reaction times (RTs) increase with a growth of the set size, already for small numerosities. According to Gallistel and Gelman (2000), this phenomenon argues for a serial counting process of some kind. An alternative explanation has been provided by Pylyshyn (2003), who assumes that judging the numerosity of a set of items may involve two different processing stages: an individuation stage that may derive object entities of a (visual) scene in parallel, and a (serial) enumeration stage in order to judge the numerosity of individuated items. Finally, it can be argued that the observed increase in RTs while judging numerosities is not a consequence of a numerical coding stage per se, but rather related to task difficulty and detection processes.

To address these issues, we trained two rhesus monkeys to perform a delayed match-to-numerosity task (Nieder et al., 2002) and employed behavioral methods that afforded us tight control and close monitoring of the animals’ behavior. By examining discrimination performance, eye movements, and RT data, we addressed questions about the parallel versus serial nature of numerical judgments.

RESULTS

Monkeys viewed a sequence of two displays separated by a memory delay and were required to judge whether the displays contained the same small number of items (1–7) (Figure 1A). To ensure that monkeys solved the task by judging number per se rather than simply memorizing sequences of visual patterns or paying attention to low-level visual features that correlate with number, we employed two types of stimulus manipulations. We randomly varied the position of the items over 24 locations centered around the monkey’s center of gaze as well as randomly varied the items between five different sizes. We also used eight sets of stimuli that, across them, controlled for changes in the total area of the items, total circumference, density, and exact appearance (Figure 1B) (see Nieder et al., 2002).

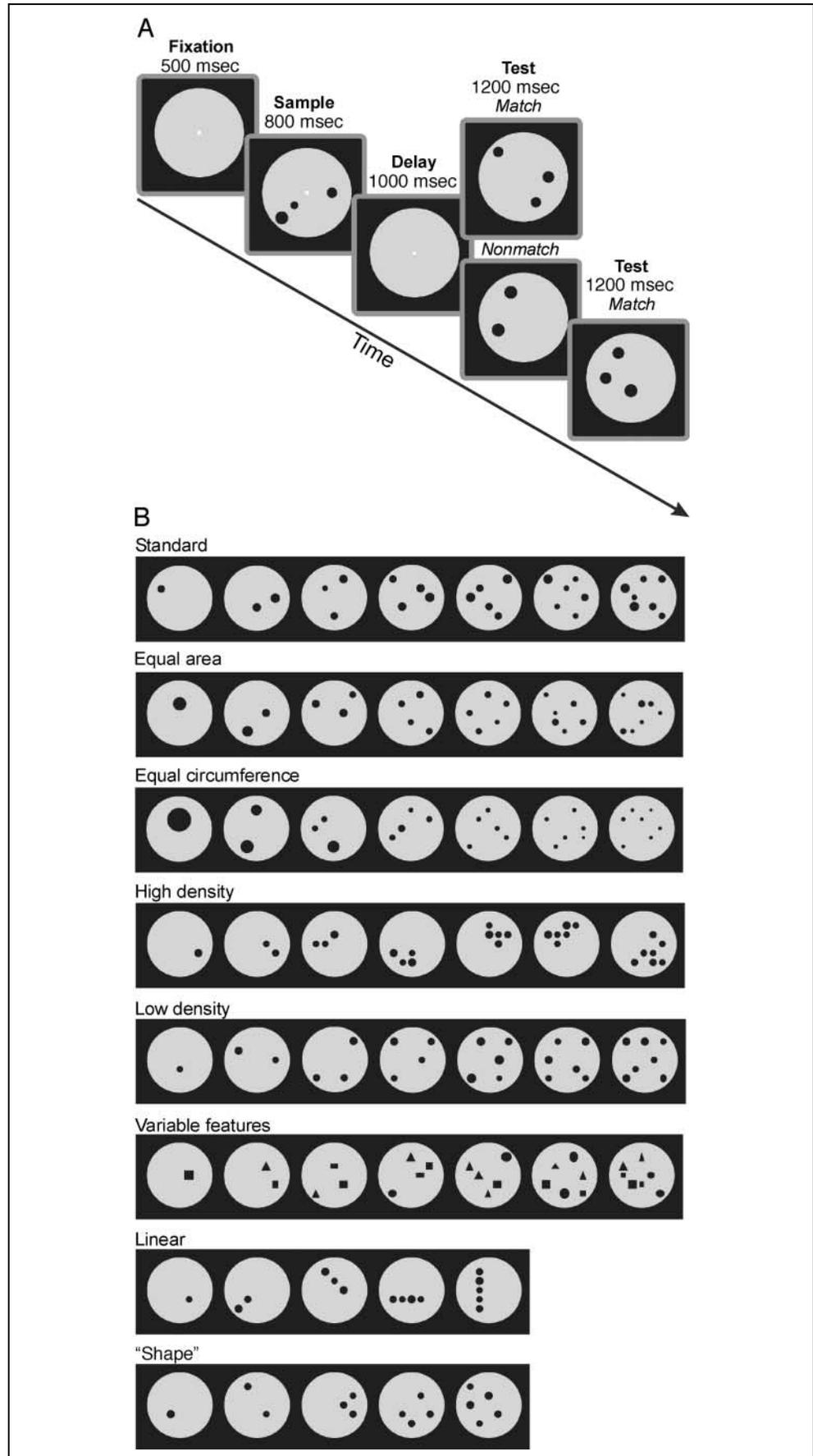
Performance for Numerosities 1 to 7

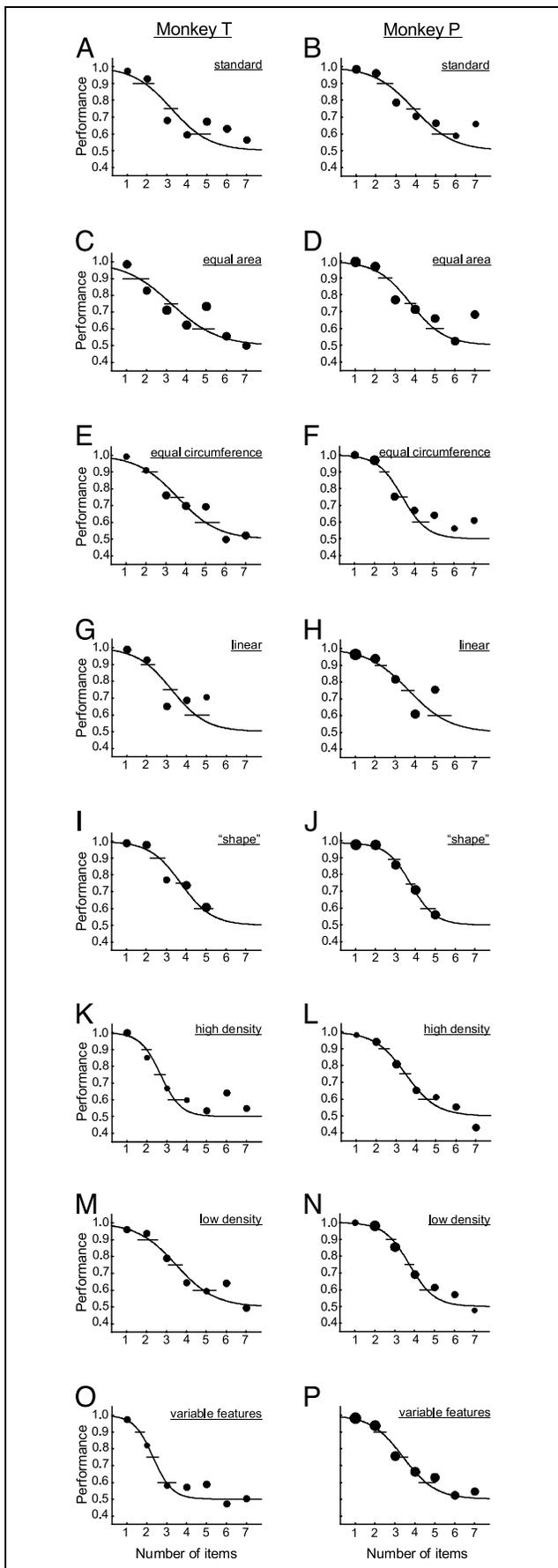
The performance data were fitted with a sigmoid function. Thresholds at performance levels of 90% (T90), 75% (T75), and 60% correct responses (T60) were derived from the fit to all of the eight different stimulus sets (Figure 2). The average performance curve of both monkeys for all conditions was a smoothly declining function that is well described by a sigmoid function [Figure 3; goodness-of-fit (r^2) = .986]. There was no sudden change in the performance for any of the tested seven numerosities, which additionally argues against strict set-size limitations and thus against the object-file model of the enumeration process.

Figure 4 displays the mean discrimination threshold values for both monkeys individually and their average (monkey T: T90 = 2.0, T75 = 3.2, T60 = 4.4; monkey P: T90 = 2.5, T75 = 3.6, T60 = 4.7). Each monkey performed between 108 and 314 trials per numerosity and stimulus set. According to a two-tailed binomial test with $n \geq 105$, a 60% correct response level is significant at $p < .05$. Thus, using 60% correct performance as criterion, the upper limit of discriminable visual quantities was between four and five items. The performance of monkey P was significantly superior compared with monkey T at the 90% and 75% performance level ($p < .05$, Wilcoxon test, two-tailed), but not at the 60% level ($p = .32$, Wilcoxon test, two-tailed).

To determine if certain stimulus sets (i.e., certain visual features) resulted in a systematic increase or decrease in discrimination thresholds, a Pearson correlation analysis for the monkeys’ thresholds at different performance levels was calculated. At a given threshold level, the threshold values of monkey T for different stimulus sets were plotted against the threshold values of the corresponding sets for monkey P. None of the correlations were significant ($p < .05$), and the correlation coefficients decreased for lower threshold levels (T90: $r = .65$; T75: $r = .50$; T60: $r = .17$). This finding

Figure 1. Stimulus protocol and example stimuli. (A) Stimulus protocol for the delayed match-to-sample task. A trial started when the monkey grabbed a bar. In the first 500 msec, the monkey only had to fixate a small fixation spot in the center of the display. A sample was displayed for 800 msec, followed by a 1000-msec delay period. The test stimuli contained either the same number of items (“match”), or one more or one less item (“nonmatch”) than the sample display. Matches and nonmatches appeared with equal probability. If a match appeared, the monkey had to release the lever to receive a reward. If a nonmatch appeared, the monkey had to wait for the second test stimulus (that was always a match) to get a reward for bar release. (B) Example images applied for the different stimulus protocols.





suggests that none of the stimulus protocols were especially easy (or difficult) for the two monkeys. Thus, it is unlikely that the monkeys used low-level visual features to solve the task. These results, along with our prior study (Nieder et al., 2002), confirm that the animals performed the task based on the quantity of items.

Reaction Times

The critical choice point in the trial was the presentation of the first test stimulus: The monkeys either released the lever if it was a match, or continued to hold it (for a nonmatch) across its disappearance to release the lever for the second test stimulus (match) (Figure 1A). The second matching test stimulus was used on “nonmatch” trials to ensure that the monkeys were paying attention and were engaged by the task on each and every trial. But, because the second match stimulus on nonmatch trials was predictable, we only analyzed behavioral RTs for lever release on trials in which the first test stimulus matched the sample (“match” trials).

Both animals showed significantly different RTs for different numbers of sample display items ($p < .0001$, Friedman test, Figure 5). Each data point was derived from 192 to 414 trials for monkey T (Figure 5A), and 275 to 671 trials for monkey P (Figure 5B). RTs increased significantly from “1” to “3” in both monkeys ($p < .01$, Mann–Whitney U test). For numerosities up to “4,” each additional item in the display resulted in a (average) latency increase of 85 msec for monkey T, and 49 msec for monkey P. Latencies for numerosity “4” were the maximum and inflection point in both functions. The RT median for numerosity “4” was 606 msec for monkey T and 472 msec for monkey P. For numerosity “5,” the latency tended to decrease for monkey T ($p < .06$, Mann–Whitney U test) and significantly declined for monkey P ($p < .02$, Mann–Whitney U test). Between “5” and “7,” the reaction latency had reached a plateau in both animals. Monkey P responded considerably faster than monkey T (for all numerosities $p < .001$, Mann–Whitney U test). On average, monkey P responded 26 msec faster to “1” (minimal difference) and 146 msec faster to “6” (maximal difference).

Figure 2. Performance data of monkey T (left column) and monkey P (right column) to stimuli with different set sizes. The stimulus protocol is indicated in the upper right corner of each panel. Numerosity ranged from one to five items for the “linear” and the “shape” protocol; all other protocol types contained displays with up to seven items. The size of the data points represents the relative number of trials per numerosity. By using a maximum-likelihood technique, a logistic psychometric function (solid line) was fitted to the data points. Discrimination thresholds at 90%, 75%, and 60% correct responses are marked by horizontal lines crossing the fit (line length represents confidence intervals).

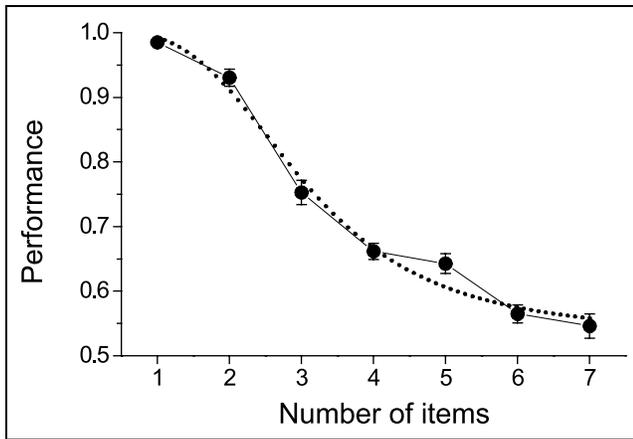


Figure 3. Averaged performance of both monkeys to all stimulus conditions. The dotted line represents the best sigmoid fit to the data (error bars \pm SEM).

Eye Movements

If the monkeys were enumerating the display in a serial fashion, this might be reflected in an increase in saccadic eye movements with an increase in display items. Alternatively, if the monkeys made more or less the same number of eye movements irrespective of the number of items, that would suggest a parallel encoding processing.

To guarantee a period of stable retinal images for electrophysiological recordings (those data are discussed elsewhere—see Nieder et al., 2002; Nieder & Miller, 2003), the monkeys maintained gaze on a central fixation target during the sample presentation and the memory delay. When the test stimulus was presented, however, they were allowed to freely scan it. Thus, we examined eye movements to this stimulus. The test display was shown for a maximum duration of 1200 msec, but disappeared if and when the monkey released the lever. The lever release occurred to a match and thus we limited the eye movement analysis to those trials only. This was to ensure that we only examined eye movements up to and not beyond the point at which the monkey finished making its numerical judgment. The only indication of this occurred on match trials (the lever release). Thus, we determined the number of eye movements from test-display onset to lever release (see Methods).

Eye movements during test stimulus presentation were always made towards individual display items. Figure 6 shows typical examples of eye movements made during the course of eight trials with different match numerosities. The top display shows the presented image together with eye traces (“white lines”), the diagrams below each image show the corresponding velocity profiles of the eye movements. In other words, the top panels give the spatial dimension of the eye movements, the lower panels show the temporal dimension.

The upper panels in Figure 6 show that the monkey was fixating at the center of the displays (zero coordinates) before test stimulus onset (i.e., during the

sample and delay epochs). Minor eye movements such as microsaccades resulted in a “clew” of eye traces in the center of the display for these periods. Once the test stimulus appeared, the monkey made rapid eye movements to one or two display items, indicated by lines leading towards them. The lower panels in Figure 6 illustrate the respective eye velocity displays. Differences in eye position between sampling frames are shown. An eye position data point was taken every 8.3 msec (120 Hz sampling). Only minor velocity changes were present during the first part of the trial from fixation onset to the onset of the test period (indicated by the dotted vertical line in the velocity panels). The rapid velocity changes during display of the match display reflect the monkey’s saccades.

The examples in Figure 6 illustrate the general finding. Even though the monkeys had 1200 msec to view the display, they made only a few saccades, and the number of saccades did not systematically increase with item numbers. Eye movements were analyzed separately for monkey T (360 to 860 trials per numerosity) and monkey P (631 to 1466 trials per numerosity). The distributions of the numbers of saccades were highly nonnormal for all numerosities and best described by a Poisson distribution ($p < .05$, one-sample Kolmogorov–Smirnov test). Monkey T made different numbers of saccades across the seven different numerosities ($p < .001$, Kruskal–Wallis test). The number of saccades was equal for numerosities 1 and 2, but significantly higher for numerosity 3 and up compared to 1 and 2 ($p < .001$, Mann–Whitney U test). However, no significant differences between numerosity 3 and all higher numerosities were observed, even though the range of performed saccades increased (see Figure 7A). The median for all distributions was only one saccade. For monkey P, no

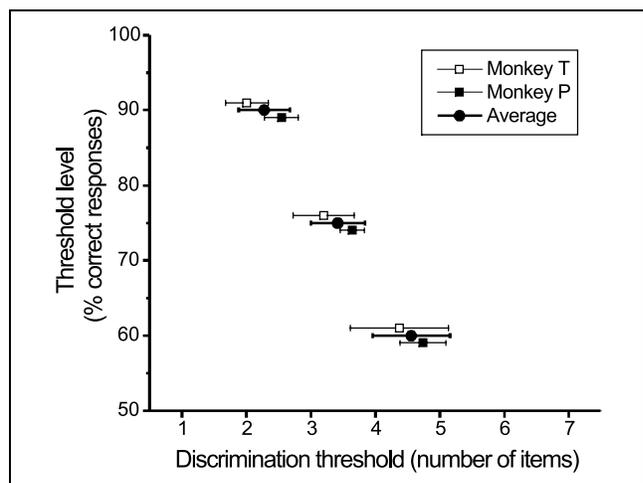
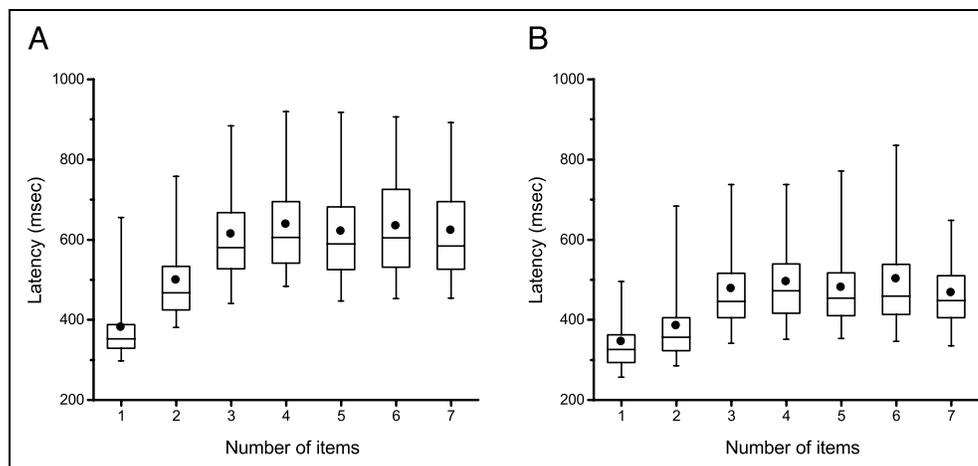


Figure 4. Mean discrimination thresholds of the two monkeys for threshold levels of 90%, 75%, and 60% correct responses. The means and the standard deviations from the means for all eight stimulus protocols are shown. (Data points are slightly shifted vertically to avoid overlapping.)

Figure 5. RTs to match stimuli for the seven tested numerosities. The box charts display values for monkey T (A) and monkey P (B). The bottom and top of the vertical line marks the 5th and the 95th percentile, respectively. The bottom and top of the box indicate the 25th and 75th percentile, respectively. The median line of the box marks the 50th percentile. The circle in the box marks the mean.



significant difference in the number of saccades was detected ($p = .13$, Kruskal–Wallis test) when numerosities 6 and 7 were excluded (which could not be discriminated above chance levels by the animals). For this monkey, the median for all numerosities was again one saccade (Figure 7B). Taken together, these results indicate that there was no systematic relationship between saccade number and the number of items, which would have been expected if the monkeys enumerated the display in a serial fashion.

DISCUSSION

In a previous article, we showed that the discrimination of two monkeys engaged in a delayed match-to-numerosity task obeyed Weber’s Law for both small and large numbers of items (Nieder & Miller, 2003). Based on the discrimination functions and the derived Weber fractions, no change for the discrimination of small and large numerosities that would have argued for a switch in representational systems (from object tracking to analog magnitudes) was observed. This suggests that the animals used an analog numerical system (estimation system) to judge quantities throughout the tested range of set sizes. Even though, RTs increased slightly and significantly up to numerosity 4, thus indicating some time-consuming aspects in judging higher numerical values. At the same time, the monkeys’ eye movements towards individual items remained constant across numerosities, suggesting that individual items were not scanned serially. In the following, the data derived from our rhesus monkeys will be discussed in the light of recent studies in monkeys, apes and humans.

RT Data and Object Tracking Mechanisms (Subitizing)

At a numerical distance of 1, the monkeys showed significant numerosity discrimination up to numerosity 4; from numerosity 5 on, performance was at chance

level. Thus, the numerical range that allows reliable conclusions (in terms of RT and eye movements) should be constrained to the range 1 to 4.

In our study, RT increased for numerosities from 1 to 4 by 67 msec per additional item for the two monkeys. These differences were small but significant. In human studies, even though the RT function for small numerosities tends to show only a mild increase, the differences in RT are usually significant (even though the older studies failed to perform proper statistics and plotted the data on compressed time scales). This can also be seen in a recent study with a chimpanzee. Appendix B of the study by Tomonaga and Matsuzawa (2002) displays the result of the appropriate nonparametric tests for the highly nonnormally distributed RT data of the ape; the RT differences are significant even for the small numerosities.

At first glance, the RT value of 67 msec in our study fits well with the general finding that enumerating sets of four or fewer items takes about 60 msec per item (called the “subitizing slope”). Subitizing was defined as an effortless, fast, and accurate process to judge a small number of items (Kaufman, Lord, Reese, & Volkman, 1949). Relatively fast RTs are traditionally taken as an indication for subitizing. Indeed, in humans, small numbers of items can be gathered with RTs of about 40–100 msec per item. The enumeration time, however, increases sharply by about 200–350 msec per item for numerosities beyond 5 (Simon, 1997; Trick & Phyllyshyn, 1993; Mandler & Shebo, 1982), which has been interpreted as a serial counting strategy for higher numbers. Subitizing has originally been explained as recognition of canonical patterns (Mandler & Shebo, 1982). According to this suggestion, we learn that collections of one to three objects fall into regular configurations: 1 is a singleton, 2 a line, and 3 a triangle. Recognition of these patterns would result in a flat response curve in terms of RT and error rate for the numbers 1 to 3. More recently, Trick and Phyllyshyn (1994) argued that subitizing exploits a limited-capacity parallel mechanism for item individu-

ation, the FINST mechanism. A limited number of object markers called FINSTs (FINGers of INSTantiation) are automatically attached to targets in the visual field for later processing (Trick & Pylyshyn, 1994). The FINST model proposes a “visual indexing mechanism” for “picking out, tracking and providing cognitive access to visual objects” (Pylyshyn, 2001).

What mechanism may cause a significant increase in numerosity judgment RTs for object tracking (subitizing)? In a strict sense, object tracking mechanisms (which include “subitizing” according to our understanding) are thought to work in parallel, and thus cannot explain even slight RT increases. Pylyshyn (2003), however, argues that subitizing involves two distinct subprocesses, an item individuation process as well as an enumeration process that judges the quantity of individuated items to

judge their numerosity. The latter enumeration process appears to be a serial process involving visiting each item (or index, respectively). In this framework, the individuation stage would be strictly parallel (see also Sagi & Julesz, 1984), but due to the subsequent enumeration stage, a slope of about 60 msec per item would result. However, if there are more than four or so items in the display, they cannot all be indexed and, thus, must be located one at a time (Pylyshyn, 2003). For larger numbers, another process must be postulated—perhaps estimating, perhaps segmentation of items into smaller groups, which then could be subitized again (as suggested by Mandler & Shebo, 1982). In the following, we will argue that such object tracking models cannot account for the monkeys’ RT data because monkeys used an analog magnitude system.

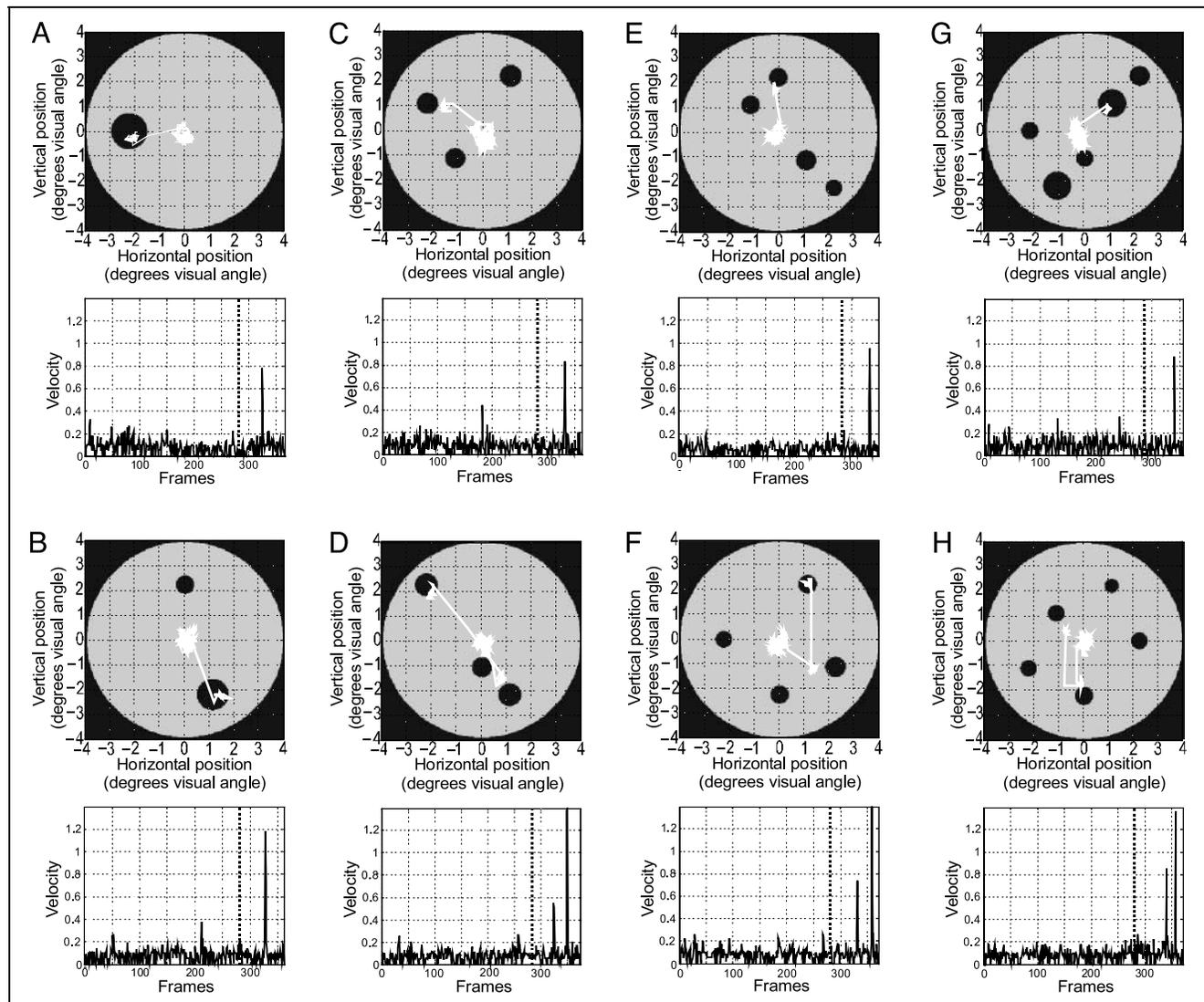
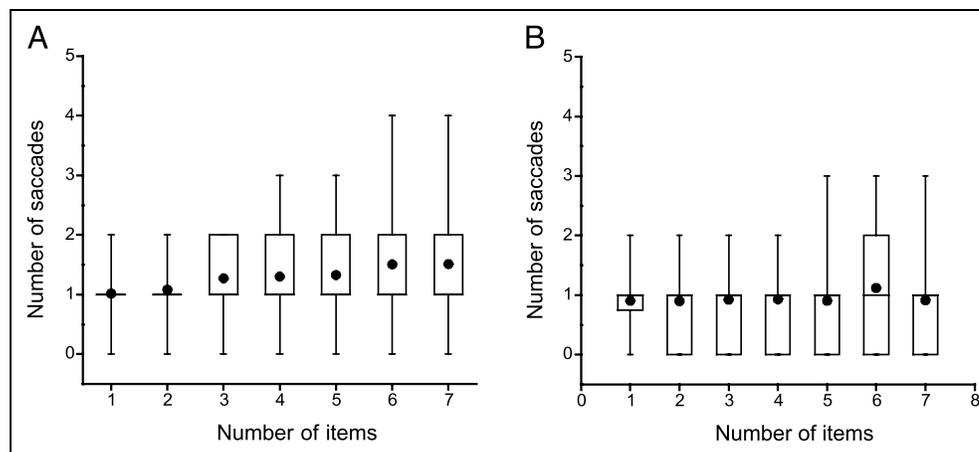


Figure 6. Typical examples of eye movements made by monkey T during the task. Examples for test displays with numerosity 1 (A), 2 (B), 3 (C, D), 4 (E, F), and 5 (G, H) are shown. The top row panels illustrate eye positions (“white line”) collapsed over time from fixation onset to bar response. The bottom panels corresponding to each match display show the velocity traces of the eye over the same time period (test onset is indicated by dotted vertical bars in the graph). Abrupt velocity changes indicate a saccade. Monkeys had to maintain fixation in the center of the displays (fixation target not shown) and were only allowed to move their eyes after the end of the delay period. Only match trials are shown.

Figure 7. Number of eye movements to match stimuli for the seven tested numerosities. The box charts display values for monkey T (A) and monkey P (B). The bottom and top of the vertical line mark the 5th and the 95th percentile, respectively. The bottom and top of the box indicate the 25th and the 75th percentile, respectively. The median line of the box marks the 50th percentile. The circle in the box marks the mean.



Evidence for Analog Magnitude Representations, and Not Object Tracking

Even though the RT values correspond well with the RT values discussed for subitizing phenomena, subitizing and its potential subprocesses cannot account for the monkeys' discrimination performance. First, subitizing by means of canonical pattern recognition (Mandler & Shebo, 1982) can be excluded, because the monkeys showed no differences in discrimination whether the items were linearly, shapelike, or in any other configuration (also see Nieder et al., 2002). Second, there was no sharp drop in performance beyond four items; the average performance curve from the monkeys shows a clear sigmoidal function and not the step function (at "4") expected for subitizing (Figure 3). Instead, performance declines smoothly through "4." And third, and most important, the clear Weber fraction signature of the monkeys' discrimination performance is not compatible with subitizing or object tracking. In a previous study where we tested numerosities up to 11 in a match-to-number task identical to this one (Nieder & Miller, 2003), we showed that variance of the monkeys' performance increased systematically with an increase in numerosity. These behavioral filter functions were best described on a logarithmic number scale. For all behavioral filter functions, a constant Weber fraction (0.35, on average) could be derived. In addition, the filter functions of single neurons in the prefrontal cortex of these monkeys mirrored the monkeys' performance in that they became progressively less selective with increasing preferred numerosity (Nieder & Miller, 2003). Thus, both the behavioral as well as the neural data obeyed Weber's Law. This clear Weber fraction signature argues for explicit numerical representations in monkeys according to an analog magnitude mechanism.

Based on our results, we favor the view that numerosities in monkey are represented by mental magnitudes. The same conclusion has been drawn in a study where monkeys learned to understand the ordinal relationships between numerosities (Brannon & Terrace,

1998, 2000). Also in this study, monkey showed a clear numerical distance and magnitude effect, consistent with a Weber fraction signature. Evidence for analog magnitude representations has been reported in other animals as well (Meck & Church, 1983; Mechner, 1958). Mechner (1958) trained rats to perform a certain number of lever presses (4, 8, 12, or 16 lever presses). The rats' performance became systematically more imprecise as the target numbers increased, resulting in progressive broadening of the lever-presses distributions. More recently, indications of a Weber fraction signature have also been observed in human infants (Xu & Spelke, 2000; Xu, 2003) and young children (Brannon & Van de Walle, 2001). Analog magnitude representations are also evident in humans. If humans are hindered from verbal counting during a numerosity judgment, the continuum of small and large numbers shows a clear Weber fraction signature (Cordes, Gelman, Gallistel, & Whalen, 2001).

Serial or Parallel Enumeration?

How should we interpret the slight increase of RTs within the analog magnitude system? It could be argued that this significant increase of RTs supports a serial, counting-like process. Several findings, however, argue against serial processing.

First, RTs for serial counting in humans are much slower. It usually takes an additional 200–350 msec per item (Simon, 1997; Trick & Phyllyshyn, 1993; Mandler & Shebo, 1982). It still may be argued that the monkeys are highly trained and could therefore show shorter RTs. However, a decrease of RT by a factor of 4 seems very unlikely.

An even stronger argument against serial processing is provided by the analysis of eye movements. Additional saccadic movements during the presentation of the test array would contribute to a better scanning of the array (Mandler & Shebo, 1982), even in parafoveal vision ($\pm 5^\circ$ of visual angle). However, neither monkey showed a systematic increase in the number of saccades with an

increase in the number of items; the median number of eye movements was one saccade for the different numerosities. This cannot be simply explained by a lack of time to perform several eye movements because the monkeys discriminated the numbers long before the maximum possible response time of 1200 msec was over. Indeed, the monkey could have made several saccades towards the items during the match interval before drawing a decision. Thus, they had had ample time to serially scan the images in order to derive the number of elements. But instead, they did not use the available time and responded much earlier (median RT 606 msec for monkey T and 472 msec for monkey P). That indicates that the time to judge numerosity did not matter (at least as long as a minimum time interval has transpired).

It is possible that covert shifts of attention independent of eye movements may account for increasing RTs. However, covert shifts of attention do not seem to have a prominent effect in humans. In the study by Mandler and Shebo (1982), the RT data for 200, 400, and 800 msec stimulus presentation durations resulted in comparable values (see their Figure 3B), only the error rate decreases with longer stimulus presentation times. Thus, the steeply rising RT slope in humans for numerosities higher than 4 is very likely not due to covert eye movements, because the slopes are almost identical for all three presentation times. Rather, the steep RT slope for higher numerosities may represent conditions where humans start to count internally (which can also be done after stimulus offset). In addition, only shifting eye position would bring the high acuity (foveal) portion of the retina onto a stimulus; covert shifts of attention would not suffice. So, it seems unlikely that covert shifts of attention would be used to acquire detailed visual information when there is sufficient time for eye movements to be made.

Finally, support for parallel processing of numerosity has also been found on the neural level. In the study by Nieder et al. (2002), neurons with preference for different numerosities were all activated on average 120 msec after sample onset, that is, there was no systematic increase in neural response latency with preferred number. The times when the neurons became selectively tuned to numerosity was also independent of a neuron's preferred numerosity. In addition, a recent fMRI study concludes that discrimination performance for small numerosities cannot be accounted for by a serial attention process, but rather, by a parallel process (Piazza, Giacomini, Bihan, & Dehaene, 2003).

Thus, both neural response latencies and the number of eye movements argue for parallel processing of simultaneously presented numerosities. (However, it should be pointed out that our finding does not exclude serial processing in other types of numerical tasks, e.g. when monkeys have to judge ordinal relationships as in the studies by Brannon & Terrace, 1998, 2000). But if it

is neither an object file representation with two different subprocesses (Pylyshyn, 2003), nor a serial counting-like process, how can the systematic increase in RTs be explained? We suspect that the mild increase in RT is due to a decision process rather than a numerosity encoding process. Numerical information may be extracted at the same time for all set sizes, but it gets more difficult to discriminate numerosities of equal numerical distance if their absolute values become larger ("numerical size effect"). Both the numerical size and distance effects are very obvious in rhesus monkey (Nieder & Miller, 2003; Brannon & Terrace, 2000). Interestingly, an increase in choice RTs can also be found for purely sensory discrimination tasks that obey Weber's Law and, thus, show a distance and size effect. For example, the time required to make a correct visual spatial frequency discrimination decreases with increasing spatial frequency difference (Greenlee & Breitmeyer, 1989). Similar RT effects can also be found in animals, in different modalities (Talwar & Gerstein, 1998). Therefore, the most parsimonious explanation is that a decision-making process alone is likely to account for the RT effects found in our monkeys. A similar conclusion has been drawn for the symbolic discrimination performance of humans; judgments of differences in magnitudes of numerals are thought to be the same as the processes involved in judgments of inequality for physical continua (Moyer & Landauer, 1967). Moyer and Landauer (1967) hypothesized that "the decision process . . . is one in which the displayed numerals are converted to analogue magnitudes, and a comparison is then made between these magnitudes in much the same way that comparisons are made between physical stimuli such as loudness or length of line."

Larger Numerosities (Beyond 4)

After the maximum of the RT function was reached at 4, the RT curves had reached a plateau. The plateau in the RF function coincided with the numerosity range where the monkeys responded at chance level because they were unable to discriminate these larger numerosities. In the present study, the animals had to discriminate set sizes with a numerical distance of 1, and they failed beyond 4 due to the limits of their discrimination capability characterized by the Weber fraction; if the numerical distance between pairs of numerosities was increased, the monkeys' performance recovered (see Nieder & Miller, 2003).

A plateau in the RT function can also be observed in a chimpanzee (albeit at a higher numerosity of 6) when the duration of the sample exposure is limited to 100 msec (Tomonaga & Matsuzawa, 2002), and this plateau corresponds also to a high percentage of errors in numerosity discrimination. Similarly, humans engaged in a numerosity judgment task show a saturated RT function (inflection around 6) concordant with high

error rates when sample exposure is temporally limited (Mandler & Shebo, 1982). Longer RTs together with fewer errors were observed for longer presentation times (800 msec) of the sample displays.

Why the rhesus monkeys in the current study showed a comparatively low inflection point remains to be investigated. It may be related to the fact that the monkeys had to maintain fixation during sample presentation. It may also indicate the capacity limits of working memory, which is thought to be about four items (Cowan, 2001; Luck & Vogel, 1997; McKone, 1995). Finally, it could indicate inferior numerosity discrimination capabilities in monkeys compared to apes and humans. But besides this threshold difference, the general form of the RT functions (increase of RT and then a plateau) are quite comparable.

Models for Analog Numerosity Judgements

Two major models have been proposed to explain the representation of numerical information by analog magnitude, an “accumulator model” (Meck & Church, 1983), and a “neural filtering model” (Dehaene & Changeux, 1993). According to the accumulator model, each item is encoded by an impulse of activation, which is added to an accumulator. The magnitude in the accumulator at the end of the count is then read into memory, forming a representation of the number of a set. Representations derived with the accumulator model obey Weber’s Law; they are noisy and the variability is proportional to the magnitude. Memory is thought to be the main source of noise rather than the processes of accumulation or comparison (Gibbon, 1992). An attraction of the accumulator model is its ability to encode sequential events, and support for it comes from the Weber fraction signature of numerical representations. However, two observations in our data do not fit this model. First, the accumulator model is, by definition, serial; the stimuli are “fed into” the accumulator one after another. Our data, however, suggest a parallel encoding of numerosity. Second, we found no evidence for memory to be the main source of variability (Nieder & Miller, 2003); the total variability of the neuronal filter functions was already present at the time of encoding (during viewing of the stimulus). Holding the information in short-term memory did not further increase the variability of the neuronal filter functions (Nieder & Miller, 2003).

Another model, the “neural filtering model” by Dehaene and Changeux (1993), consists of four main levels. Most relevant for the current discussion is the first level, the numerosity detection system. Each stimulus is coded as a local Gaussian distribution of activation by topographically organized input clusters (e.g., the retina). Next, items of different sizes are normalized to a size-independent code. At that stage, item size, which was initially coded by the number of active clusters on the

retina (“quantity code”), is now encoded by the position of active clusters on a location map (“position code”). Clusters in the location map project to every unit of succeeding “summation clusters,” whose thresholds increase with increasing number and pool the total activity of the location map. The summation clusters finally project to “numerosity clusters.” Numerosity clusters are equipped with appropriate central excitation and lateral inhibition so that each numerosity cluster responds only to a selected range of values of the total normalized activity (i.e., their preferred numerosity). The Dehaene and Changeux model is able to capture several critical aspects of our results. First, it provides approximate detection of the number of items, which is characteristic of an analog magnitude process. Second, the model derives numerosity in parallel, a feature we observed both in the monkeys’ behavior and in prefrontal neurons (Nieder et al., 2002; Nieder & Miller, 2003). Third, the model’s numerosity units become less selective with increasing center numerosities (i.e., the distributions become broader), that is, they obey Weber’s Law. The model nicely describes extraction of simultaneously presented numerosities, but it is not clear how serially, or even multimodally, presented items/events would activate the network. If sequential and cross-modal events could be implemented within the framework of such a filtering model, we suspect it may provide a full description of analog magnitude processes in general.

Summary

The nonverbal numerical representations in monkeys performing a delayed match-to-numerosity task are based on analog magnitudes. There is no need to assume a separate object tracking process (subitizing) for small numerosities because the continuum of small and large numbers shows a clear Weber fraction signature. The lack of increased scanning eye movements with increasing set sizes, together with indistinguishable neuronal response latencies for neurons with different preferred numerosities, argues for parallel encoding of numerical information. The slight but significant increase in RT with increasing numerosities can be explained by task difficulty and, consequently, time-consuming decision processes. Whether rhesus monkeys may be able to switch to other coding schemes (such as those that engage serial mechanisms) remains to be examined by future investigations.

METHODS

Subjects and Apparatus

The subjects were two adult male rhesus monkeys (*Macaca mulatta*) weighing 9.8 and 11.0 kg. Both monkeys were also used for electrophysiological recordings

(Nieder et al., 2002; Nieder & Miller, 2003). The monkeys earned their liquid consumption during the daily experimental sessions. Care and treatment of the monkeys were in accordance with the NIH and MIT guidelines for animal experimentation.

The animals were trained to sit in a monkey chair positioned at a viewing distance of 57 cm in front of a computer screen and inside a dark booth. A head used for fixing head position was affixed to the skull using standard surgical procedure and under general anesthesia. This enabled us to monitor eye movements with an infrared eye tracking system (ISCAN, Burlington, MA). A juice tube positioned at the monkey's mouth supplied drops of apple juice for reward. A personal computer running the CORTEX software controlled experimental events and behavioral data collection.

Delayed Match-to-Numerosity Task

A trial started when the monkey grasped a lever and fixated a central fixation target. A sample display containing one to seven items was shown for 800 msec. After that, there was a memory delay (1000 msec) in which only the gray background circle without items was shown. Next, a test display was presented for up to 1200 msec. It was either a match (it contained the same number of dots as the sample display) or a nonmatch. The nonmatch stimuli for intermediate samples (1 to 6) were one number up and down (probability = .25). Nonmatch for "one" was only "two"; for "seven" the nonmatch was only "six." Match and nonmatch displays appeared pseudorandomized and with equal probability ($p = .5$). If the test display was a match, monkeys released the lever before it disappeared to receive a juice reward. If the test display was a nonmatch, the monkeys held the lever until the second test display, which was always a match, appeared. This also required a lever release to receive a reward. Trials were randomized and balanced across all relevant features. The chance level for this task protocol was 50% correct responses.

The two basic error types included fixation breaks (eye movements away from the fixation point) and false responses (i.e., bar releases) to nonmatch stimuli. Fixation breaks were not counted when calculating performance level; only incorrect numerical judgments were counted. Both errors terminated the trial immediately and resulted in a timeout (1.5 sec for breaking fixation and 4 sec for a false response, respectively). The inter-trial interval was 1.5 sec. Monkeys performed between 500 and 1000 correct trials per session (day). Monkeys had to keep their gaze within 1.25° of the fixation point during sample presentation and the memory delay. The eye movements during each trial were monitored with an infrared eye tracking system (ISCAN) at a temporal resolution of 120 Hz and stored to disk for off-line analysis.

Visual Stimuli

The items were black (diameter range 0.8° to 1.3° of visual angle) and were displayed on a gray circular background (diameter: 8° of visual angle). To prevent the monkeys from simply memorizing the visual patterns of the displays, each quantity was tested with 100 different images per session (by randomly varying the size and location of the items) and the sample and test displays that appeared on each trial were never identical.

Efforts were made to exclude the possibility that the monkeys could perform the discrimination by attending low-level visual features that happen to correlate with numerosity. The spatial arrangement of the items was randomized on a 5 by 5 matrix (except when controlled for linear or shapelike arrangement). The monkeys were trained with "standard" stimuli (Figure 1B), which comprised dots of different sizes. On average, the surface area, the circumference, and the density of the items increased with increasing numerosity for the "standard" stimuli. Therefore, controls were included with displays in which the total area ("equal area") or the total circumference ("equal circumference") was equated across different quantities. In addition, we controlled for dot density effects by applying "high-density" and "low-density" protocols. The dot density was determined by calculating the average distance between the dots. For the "high-density" stimuli, the dots had an average distance of $<1.4^\circ$ of visual angle (measured from the dots' center). For the "low-density" stimuli, the items were arranged with an average distance of $>2.5^\circ$ of visual angle. Moreover, we displayed different geometric objects (squares, bars, triangles, ovals, dots) of different sizes in the "variable features" protocol. Finally, the dots were linearly arranged in the "linear" protocol, or three dots were arranged as triangle, four dots as quadrangle, and five dots as pentagon in the "shape" protocol. There was a total of seven numerosities and all seven were used in each session (except for the "line" vs. "shape" stimuli) and all displays were newly generated for each session by pseudorandomly shuffling all relevant item features (e.g., position, size, identity).

Quantification of Psychometric Functions

We derived psychometric functions that described the relation between the number of items and the monkeys' ability to respond correctly to it. Since the delayed match-to-sample paradigm allows either a correct or an incorrect response per trial, a performance probability of .5 correct responses (or 50%) indicates chance level, and a probability of 1.0 represents perfect discrimination (100% correct). The dataset in our experiments is described by three vectors: x will denote the numerosity tested, n the number of trials per numerosity, and y the monkeys' correct responses.

To derive performance thresholds and other quantitative measures, a model psychometric function P that specifies the relationship between the probability of a correct response and the number of items x was fitted to the data using a maximum-likelihood technique (Wichmann & Hill, 2001a, 2001b; Harvey, 1986).

$$P(x) = \gamma + (1 - \lambda - \gamma)F(x; \alpha, \beta), \quad (1)$$

where the sigmoid function F is a two-parameter logistic function specified by α (the number of items at the halfway point) and β (the slope of the function). The lower bound of P is given by parameter γ (the probability of being correct by chance, “guess rate”) which was fixed at .5. The upper bound is given by $1 - \lambda$, where λ corresponds to the miss rate (“lapse rate”); it was initially set to 0.01. We used the program “psignifit” (formerly known as “psychofit”) by Jeremy Hill; it is available on <http://bootstrap-software.org>.

Determination of Eye Movements

The occurrence of saccades was determined by means of velocity changes of the eye traces. The threshold for eye movement to be defined as a saccade was determined by analyzing velocity values on a frame-by-frame basis (120 Hz temporal resolution) while the monkey fixated the fixation target (i.e., during the fixation, the sample, and the delay epochs). The 85th percentile of all velocity values for all correct trials per session for these epochs was taken as threshold value. The 85th percentile threshold allowed for microsaccades that could occur during fixation. Eye velocities exceeding this threshold value from the beginning of the test display to the bar release of the monkey were defined as saccades. Saccades were always made to individual items of the display (see examples in Figure 6).

Acknowledgments

A. N. was supported by a grant from the Deutsche Forschungsgemeinschaft and a long-term fellowship of the Human Frontier Science Program. This work was supported by an NIMH grant (1-R01-MH65252-01) and the RIKEN-MIT Neuroscience Research Center. We thank two anonymous reviewers for helpful suggestions.

Reprint requests should be sent to Andreas Nieder, Primate NeuroCognition Laboratory, Department of Cognitive Neurology, Hertie-Institute for Clinical Brain Research, University of Tuebingen, Hoppe-Seyler-Str. 3, 72076 Tuebingen, Germany, or via e-mail: andreas.nieder@uni-tuebingen.de.

REFERENCES

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

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

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

Brannon E. M., & Van de Walle, G.A. (2001). The development of ordinal numerical competence in young children. *Cognitive Psychology*, *43*, 53–81.

Carey, S. (2001). Cognitive foundations of arithmetic: Evolution and ontogenesis. *Mind and Language*, *16*, 37–55.

Cordes, S., Gelman, R., Gallistel, C. R., & Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychological Bulletin and Review*, *8*, 698–707.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *2*, 87–114.

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

Emmerton, J. (2001). Birds’ judgments of number and quantity. In R. G. Cook (Ed.), *Avian visual cognition*. [On-line]: <http://www.pigeon.psy.tufts.edu/avc/emmerton/>.

Feigenson, L., Carey, S., & Spelke, E. (2002). Infants’ discrimination of number vs. continuous extent. *Cognitive Psychology*, *44*, 33–66.

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

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

Gibbon, J. (1992). Ubiquity of scalar timing with a Poisson clock. *Journal of Mathematical Psychology*, *36*, 283–293.

Greenlee, M. W., & Breitmeyer, B. G. (1989). A choice reaction time analysis of spatial frequency discrimination. *Vision Research*, *29*, 1575–1586.

Harvey, L. O., Jr. (1986). Efficient estimation of sensory thresholds. *Behavior Research Methods, Instruments, and Computers*, *18*, 623–632.

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

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

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

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.

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

McKone, E. (1995). Short-term implicit memory for words and nonwords. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1108–1126.

Mechner, F. (1958). Probability relations within response sequences under ratio reinforcement. *Journal of the Experimental Analysis of Behaviour*, *1*, 109–121.

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.

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

- Nieder, A., Freedman, D. J., & Miller, E. K. (2002). Representation of the quantity of visual items in the primate prefrontal cortex. *Science*, *297*, 1708–1711.
- Nieder, A., & Miller, E. K. (2003). Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron*, *34*, 149–157.
- Ninokura, Y., Mushiake, H., Tanji, J. (2003). Representation of the temporal order of visual objects in the primate lateral prefrontal cortex. *Journal of Neurophysiology*, *89*, 2868–2873.
- Ninokura, Y., Mushiake, H., & Tanji, J. (2004). Integration of temporal order and object information in the monkey lateral prefrontal cortex. *Journal of Neurophysiology*, *91*, 555–560.
- Piazza, M., Giacomini, E., Le Bihan, D., & Dehaene, S. (2003). Single-trial classification of parallel pre-attentive and serial attentive processes using functional magnetic resonance imaging. *Proceedings of the Royal Society of London, B, Biological Sciences*, *270*, 1237–1245.
- Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition*, *80*, 127–158.
- Pylyshyn, Z. W. (2003). *Seeing and visualizing: It's not what you think*. Cambridge: MIT Press.
- Sagi, D., & Julesz, B. (1984). Detection versus discrimination of visual orientation. *Perception*, *13*, 619–628.
- Sawamura, H., Shima K., & Tanji, J. (2002). Numerical representation for action in the parietal cortex of the monkey. *Nature*, *415*, 918–922.
- Simon T. J. (1997). Reconceptualizing the origins of number knowledge: A “non-numerical” account. *Cognitive Development*, *12*, 349–372.
- Talwar, S. K., & Gerstein, G. L. (1998). Auditory frequency discrimination in the white rat. *Hearing Research*, *126*, 135–150.
- Tomonaga, M., & Matsuzawa, T. (2002). Enumeration of briefly presented items by the chimpanzee and humans. *Animal Learning and Behavior*, *30*, 143–157.
- Treisman, A. (1992). Perceiving and re-perceiving objects. *American Psychologist*, *47*, 862–875.
- 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 and Performance*, *19*, 331–351.
- Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently—A limited-capacity preattentive stage in vision. *Psychology Review* *101*, 80–102.
- Uller, C., Jaeger, R., Guidry, G., & Martin, C. (2003). Salamanders (*Plethodon cinereus*) go for more: Rudiments of number in an amphibian. *Animal Cognition*, *6*, 105–112.
- Wichmann, F. A., & Hill, N. J. (2001a). The psychometric function I: Fitting, sampling and goodness-of-fit. *Perception and Psychophysics*, *63*, 1293–1313.
- Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function: II. Bootstrap based confidence intervals and sampling. *Perception and Psychophysics*, *63*, 1314–1329.
- Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, *358*, 749–750.
- Wynn, K., Bloom, P., & Chiang, W. C. (2002). Enumeration of collective entities by 5-month-old infants. *Cognition*, *83*, B55–B62.
- Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*, B1–B11.
- Xu, R. (2003). Numerosity discrimination in infants: Evidence for two systems of representation. *Cognition*, *89*, B15–B25.