Compressed Scaling of Abstract Numerosity Representations in Adult Humans and Monkeys

Katharina Merten and Andreas Nieder

Abstract

■ There is general agreement that nonverbal animals and humans endowed with language possess an evolutionary precursor system for representing and comparing numerical values. However, whether nonverbal numerical representations in human and nonhuman primates are quantitatively similar and whether linear or logarithmic coding underlies such magnitude judgments in both species remain elusive. To resolve these issues, we tested the numerical discrimination performance of human subjects and two rhesus monkeys (*Macaca mulatta*) in an identical delayed match-to-numerosity task for a broad range of numerosities from 1 to 30. The results demonstrate a noisy nonverbal estimation system obeying Weber's Law in both species. With average Weber fractions in the range of 0.51 and 0.60, nonverbal numerosity discriminations in humans and monkeys showed similar precision. Moreover, the detailed analysis of the performance distributions exhibited nonlinearly compressed numerosity representations in both primate species. However, the difference between linear and logarithmic scaling was less pronounced in humans. This may indicate a gradual transformation of a logarithmic to linear magnitude scale in human adults as the result of a cultural transformation process during the course of mathematical education. ■

INTRODUCTION

Several lines of evidence suggest that abstract numerical competence is a sovereign faculty independent of language. Innumerate human adults (Gordon, 2004; Pica, Lemer, Izard, & Dehaene, 2004), numerate human adults prevented from counting verbally (Barth, Kanwisher, & Spelke, 2003; Whalen, Gallistel, & Gelman, 1999), preverbal infants (Feigenson, Dehaene, & Spelke, 2004), and nonverbal animals (Nieder, 2005; Brannon & Terrace, 1998, 2000) all can discriminate numerical quantity. Still, the nature of numerosity representations remains controversial.

Data exploiting spontaneous behavior in infants and animals have suggested the representation via an object tracking mechanism, which is suited to discriminate sets limited to four items only. This implicit representation of small numerosities assigns "markers" to particular items in a set and allows for fast, precise, and automatic discrimination performance (Feigenson et al., 2004; Hauser, Carey, & Hauser, 2000; Kahneman, Treisman, & Gibbs, 1992; Mandler & Shebo, 1982). However, more recent studies showed that 6-month-old infants are also able to discriminate much larger numerosities (e.g., 8 vs. 16) which, by definition, cannot be processed by the object tracking mechanism (Xu, 2003; Xu & Spelke, 2000).

Another more traditional view of the nonverbal numerosity coding is the analog magnitude representation.

This system allows approximate numerosity discrimination relying on estimation (e.g., Nieder & Miller, 2004a; Nieder, Freedman, & Miller, 2002; Whalen et al., 1999; Dehaene, 1992). Estimation of numerical quantity is characterized by the "numerical distance effect" (discrimination performance between quantities improves with increasing numerical distance) and the "numerical magnitude effect" (greater numerical distances between quantities are required to discriminate larger absolute magnitudes). The analog magnitude system allows for a continuous estimation of set sizes without an upper limit of numerosity encoding, but does become systematically less precise with increasing numbers. Thus, the hallmark of analog magnitude representations is that they obey Weber's Law (Weber, 1850).

All studies performed in behaviorally trained animals argue for an explicit representation of both small and large numerosities via this continuous analog magnitude system showing the Weber Law signature. Two main types of training protocols have been applied: the delayed match-to-sample task, in which equal numerosities need to be matched (Nieder & Merten, 2007; Cantlon & Brannon, 2006; Nieder et al., 2002), and a bisection task, in which a range of set sizes needs to be judged relative to two trained numerical "anchor values" in a forced-choice situation (Emmerton & Renner, 2006; Jordan & Brannon, 2006a; Roberts, 2005, 2006). Ratio-dependent analog magnitude representations have been found both for simultaneously (Nieder & Merten, 2007; Cantlon & Brannon, 2006; Nieder & Miller, 2003; Brannon &

University of Tuebingen, Germany

Terrace, 1998) and sequentially presented numerosities (Beran, 2007; Nieder, Diester, & Tudusciuc, 2006).

Within the realm of the analog magnitude system, the scaling of the numerical representations remains disputed. According to the linear-coding hypothesis, internal behavioral performance distributions are symmetric and centered on each number on a linear scale (Brannon, Wusthoff, Gallistel, & Gibbon, 2001; Gallistel & Gelman, 2000; Gibbon & Church, 1981; Gibbon, 1977). On the other hand, within the framework of the nonlinear compression hypothesis, numerical representations are only symmetric distributions on a logarithmic-like compressed scale (Nieder & Merten, 2007; Nieder & Miller, 2003; Dehaene, 1992, 2001; Dehaene & Changeux, 1993; van Oeffelen & Vos, 1982). The critical variable to dissociate the two hypotheses is the symmetry of the internal representations. Therefore, a mathematical model is required to describe the behavioral performance distributions and to evaluate their symmetry for particular scaling schemes. To date, behavioral studies in monkeys either tested only small numerosities or used an experimental design not suitable to address the question of scaling quantitatively.

Moreover, it is not known whether nonverbal numerosity representations in humans and nonhuman primates are quantitatively similar. Basic quantity representations might undergo transformations under the guidance of precise symbolic representations in humans. If and how language has an impact also on the nonverbal scaling scheme is another debated issue. Interestingly, symbolic number representations in humans seem to experience a transformation of the scaling scheme with age and number proficiency (Siegler & Booth, 2004; Siegler & Opfer, 2003) which might, in turn, have an impact on nonverbal magnitude representations. This issue is not settled, mainly because identical behavioral protocols have rarely been applied in a comparative manner in human and animal species. The representation of large numerosities (up to 30) has been studied just once in rhesus monkeys and humans using an identical task (Cantlon & Brannon, 2006). Although the performance of both species clearly obeyed Weber's Law, the behavioral protocol applied in this study could not inform about the scaling scheme underlying numerical representations.

Behavioral data are necessary, but not sufficient, to fully elucidate the scaling of numerical representations because the behavioral outcome of an estimation task may be the result of different scaling schemes at different processing stages. However, recent single-cell recordings in numerosity-discriminating monkeys suggest a close match between behavioral and neuronal numerosity representations (Nieder & Merten, 2007; Nieder & Miller, 2004b; Nieder et al., 2002). Numerosity-selective neurons discharged maximally to a preferred quantity and showed a progressively declining activity with increasing numerical distance from the preferred numerosity. The resulting numerosity-tuning curves allow a direct comparison of neuronal tuning and behavioral tuning functions. Evidence that populations of numerosity-tuned neurons also exist in the human cortex came from a functional imaging study, in which the tuning behavior of neurons was read-out indirectly via an fMRI adaptation protocol (Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004).

In the current study, the numerical discrimination performance of rhesus monkeys and adult humans was tested for a broad range of numerosities (1–30) with the very same delayed matching-to-sample task. Under these circumstances, any differences in the humans' and monkeys' nonverbal discrimination performance can be attributed to representational differences, not to methodological incompatibilities. The detailed performance functions allow us to describe the data with the most suitable mathematical model, to quantitatively evaluate the distributions, to analyze the Weber fractions and to determine the best scaling scheme.

METHODS

Subjects

A total of 36 volunteers (20 men, 16 women), ages 22–29 years, participated in the human psychophysical experiments. In the monkey experiments, the behavioral performance of two adult male rhesus monkeys (*Macaca mulatta*) was investigated. Both monkeys, Monkey M and Monkey W, were socially housed with other monkeys and were treated in accordance with the guidelines for animal experimentation approved by the Regierungspräsidium Tübingen, Germany. From previous experiments, Monkey M was capable of discriminating numerosities 1 to 4 and Monkey W was familiar with discriminating numerosities 1 to 5.

Stimuli

Numerosity stimuli, consisting of multiple-dot patterns, were generated using custom-written MatLab software. For the standard stimuli, small black filled dots (diameter of 0.17°–0.28° visual angle) appeared on a gray background of a large circular area with a diameter of 7° visual angle. Each stimulus contained a defined set of dots that appeared at random locations within the background circle. The diameter of each dot was randomly varied within the given range.

To ensure that the number-discrimination task was solved by judging discrete quantity only, low-level visual features were controlled for with additional stimuli. Two sets of control stimuli were used (Figure 1A): area control (total area of all items in a display equated for all stimuli in a trial) and density control (same mean density of dot patterns for all stimuli in a trial). In order to prevent the subjects from memorizing the visual patterns of the displays, sample and test images that

Figure 1. Delayed match-to-numerosity protocol and example stimuli. (A) Example stimuli (standard, area control, and density control) for sample numerosity 14, which had to be discriminated from numerosities 6, 10, 18, and 22. (B) A trial started when the monkey grabbed the bar and fixated the fixation spot. A sample stimulus (here, 22 dots) was followed by a delay period. The test stimulus contained either the same number of items ("match") or a different quantity of dots (here, 35 dots) ("nonmatch"). Each nonmatch stimulus was followed by a match stimulus, with the same set size as the sample. When a match stimulus appeared, the monkey was required to release the lever to receive a reward.
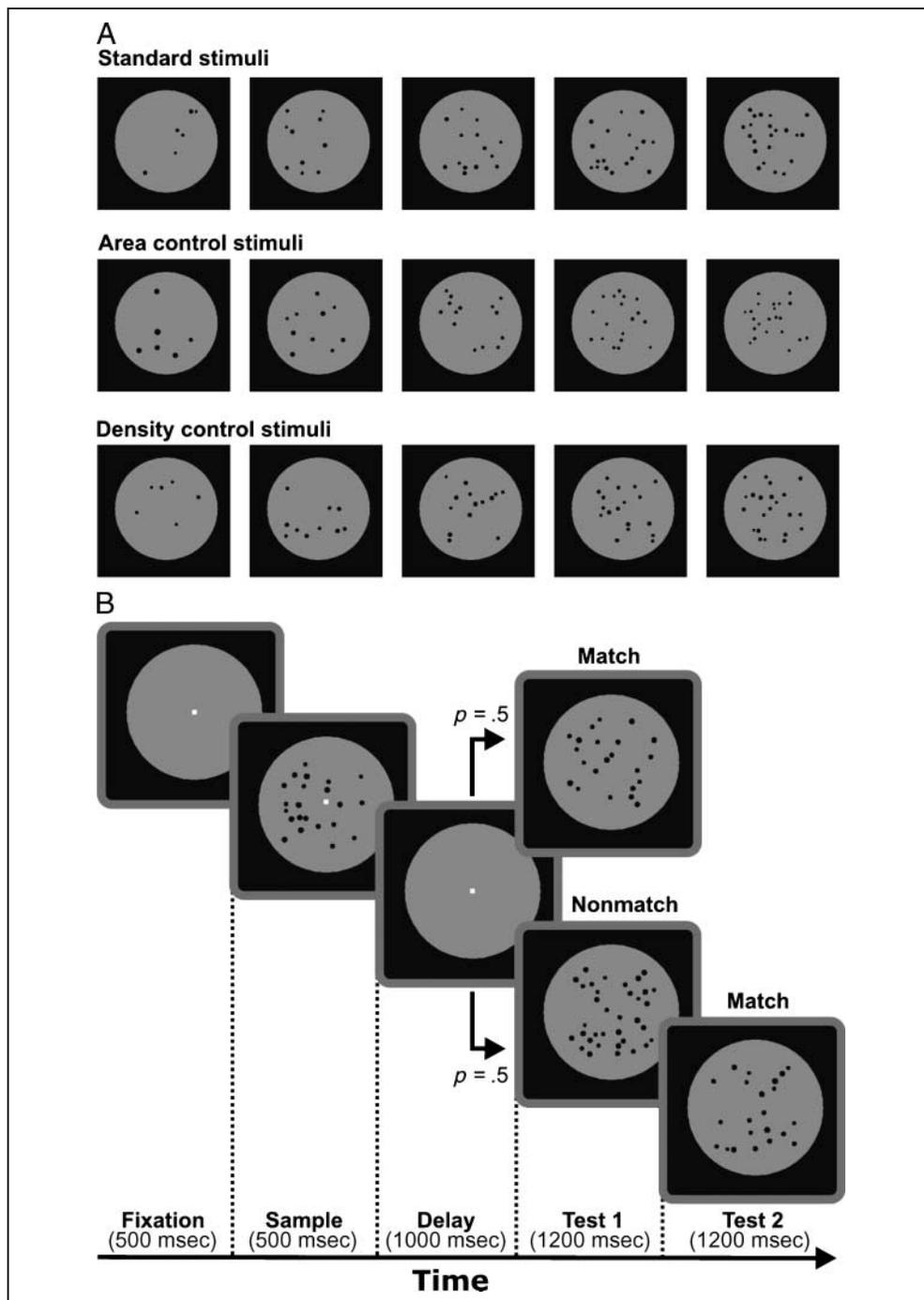
appeared in one trial were never identical and new sets of stimuli were generated for each testing session.

Apparatus and Behavioral Protocol

The stimuli were presented on a 15-in. TFT display and were viewed from a distance of 57 cm in a darkened room. Eye movements were monitored with an infrared eye tracking system, ISCAN, at a sample rate of 120 Hz. A PC running the CORTEX program (NIMH) was used for the experimental control and data acquisition.

The experimental set-up was identical for both humans and monkeys except for two technical details. First, to maintain the viewing distance, monkeys sat in the primate chair and were head-fixed, whereas, for humans, an adjustable chin rest was used. Second, monkeys had to grasp a lever to start a trial and to release the lever to give a response and to receive fluid reward. Human subjects pressed a button to indicate responses.

The subjects performed a delayed match-to-sample task with numerosity displays as stimuli (Figure 1B). To initiate a trial, the subjects fixated the central fixation

spot for 500 msec. Each experimental trial started with a sample stimulus (500 msec) containing a particular numerosity of dots. After a delay of 1000 msec, a Test 1 stimulus was presented (1200 msec), which was either a match (containing the same number of dots as the sample display) or a nonmatch (containing more or fewer items than the sample). If the first test stimulus was a match, the subjects had to respond. Otherwise, if the first test was a nonmatch, the subjects had to withhold the response until a second test stimulus appeared. This second test stimulus (1200 msec) was always a match. Match and nonmatch trials appeared with equal probability. Whenever a response error occurred, a red screen was flashed before a new trial started after a timeout period of 1 to 3 sec.

The gaze was restricted to within 1.75° visual angle of the fixation spot during the sample presentation and the delay period. Monkeys' eye movements were monitored in all experimental sessions. Human subjects were instructed to fixate during these time intervals, and for 23 human participants, eye movements were monitored. There was no significant difference in discrimination performance between human subjects whose fixation was monitored and subjects whose eye movements were not tracked (Wilcoxon test, $p = .59$). If a fixation error occurred, the current trial was interrupted, a blue screen was flashed, and a timeout period of 1 to 3 sec was inserted. The specific experimental protocols are described in the following paragraphs.

Small Numerosity Protocol (Monkeys Only)

Prior to the experiment with large numerosities, monkeys' performance for small numerosities was determined using the same delayed match-to-sample task. Sample numerosities 1, 2, 3, and 4 were tested.

Transfer Trials Protocol (Monkeys Only)

Monkey W was engaged in the transfer trials experiment. The monkey was reinforced to discriminate numerosities 1 to 5 in baseline trials. Transfer trials showing novel numerosities 6, 7, and 8 were inserted in 15% of all trials among baseline trials. The monkey was randomly rewarded in 80% of the transfer trials, irrespective of performance so that it was not reinforced to respond "correctly."

Large Numerosity Protocol (Humans and Monkeys)

To assess subjects' performance to a broad range of quantities, numerosities 1, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, and 30 were presented as samples. For each sample stimulus, four match and four nonmatch stimuli were generated. Match stimuli were composed of the same number of items as the sample, but with different arrangements of dots and different dot sizes. The

numerosities of nonmatch stimuli (N_{NM}) contained larger or smaller numbers of items equidistant from the sample numerosity (N_S). The nonmatch numerosities were calculated using the factor $f = 0.3$ and 0.6 according to the equation:

$$N_{NM} = \text{round}(N_S(1 \pm f)). \tag{1}$$

Consequently, the numerosities of nonmatch stimuli were arranged symmetrically around the sample quantity. A total of 256 trials was presented in one block. Each experimental block contained a set of standard and a set of control stimuli (area or density). The monkeys completed four to eight blocks during one session. The type of control stimulus (area or density) changed from session to session. Each human participant completed two blocks (512 trials) in one experimental session and was randomly assigned to one of the control conditions. Experimental trials were randomized and balanced across all relevant features.

Precision Protocol (Monkeys Only)

To precisely map the shape of the performance curves, discrimination of sample numerosity 15 relative to all other numerosities from 1 to 29 (in steps of one) was examined. To ensure that the monkeys were paying attention to the sample numerosities, sample numerosity 15 was only shown in half of the trials of a session, whereas in the other 50% of the trials sample numerosities 1, 4, 8, 12, 18, 22, and 26 were presented. The numerosities of nonmatch stimuli were determined according to Equation 1. In this experiment, match and nonmatch trials were also equally likely to appear.

Data Analysis

Data analysis was carried out with custom-written MatLab software. For the monkey data analysis, data were averaged across all sessions of a particular testing protocol. To reduce noise in the human numerosity discrimination performance, it was necessary to pool the data from all human participants. Otherwise, the nonmatch trials would have been sampled only twice per subject at each point because each of the two experimental blocks tested every nonmatch numerosity just once. Average performance accuracy (P_{avrg}) for each sample numerosity was calculated as follows:

$$P_{avrg} = 100 \cdot \frac{N_{Mcorr} + N_{NMcorr}}{N_{all}}, \tag{2}$$

where N_{Mcorr} is the number of correct responses in match trials (response after the first test stimulus),

N_{NMcorr} is the number of correctly recognized nonmatch trials (response after the second test stimulus), and N_{all} is the number of all presented test stimuli for the particular sample. Because match and nonmatch trials were equally likely to appear in the experiment, the chance level for correct responses was 50%.

Performance curves for each sample indicate the probability that displays in the test period were judged as containing the same number of items as the sample numerosity. The center data point of each performance curve indicates the correct performance in the match trials (where the first test display showed the same numerosity as had been cued in the sample period). The data points to the left and the right of the center indicate performance in the nonmatch trials (i.e., where the first test display showed a smaller or larger number of items); for the nonmatch numerosities, the percentage of errors for the respective nonmatch numerosity is plotted. For more distant nonmatch numerosities from the sample, fewer errors will be made, which demonstrates the numerical distance effect. Therefore, the averaged performance curves would have a peak function shape. In analogy to quantity-selective neurons (Nieder & Merten, 2007; Nieder et al., 2002), which discharge maximally for a preferred stimulus magnitude, give attenuated responses for not-preferred numerosities, and so, behave like band-pass filters, behavioral performance curves are also called filter functions.

The characteristics of the behavioral filter functions for the different sample numerosities (e.g., symmetry, curve width) could be determined by modeling the performance curves with defined peak functions and examining their parameters for the plots on linear or nonlinear compressed scales. To determine which peak function would model the behavioral filter functions best, a symmetric Gaussian distribution

$$F(x) = y_0 + a\left(e^{\frac{(x-x_C)^2}{2\sigma^2}}\right) \quad (3)$$

and a symmetric sigmoid function

$$F(x) = y_0 + 4a\left(\frac{1}{1+e^{-\frac{(x-x_C)}{s}}}\right)\left(\frac{1}{1+e^{\frac{(x-x_C)}{s}}}\right) \quad (4)$$

were fitted to the precise performance distribution curve for the sample numerosity 15 on a linear and a logarithmic scale. To acquire the best fit to the data, σ (standard deviation of the Gaussian distribution), s (steepness of the sigmoid function), and a (amplitude) were adjustable during the fitting procedure. The peak function's y-axis offset (y_0) was set to zero, and the center of the fitted distributions (x_C) was fixed at the function's sam-

ple value. Goodness-of-fit values (r^2) were calculated to evaluate the quality of the fits.

To address the question of the scaling scheme that results in symmetric behavioral tuning functions, data were plotted on a linear and three nonlinear compressed scales. To describe the nonlinear scaling of sensory impressions, Fechner proposed a logarithmic relationship between the sensation (S) and the physical magnitude of the stimulus (I) ($S = k \times \log(I)$) (Fechner, 1860). Stevens (1861) postulated that the sensation of a stimulus is a power function of the stimulus magnitude ($S = k \times I^n$). In the present study, we tested the representation of numerosities using a power function (Stevens' Law) with an exponent of $^1\!/_2$, a power function with exponent of $^1\!/_3$, and a logarithmic relationship (Fechner Law), in the order of increasing nonlinear compression. To evaluate the symmetry of the behavioral filter functions for different scales, the Gaussian distribution was fitted to all performance curves and r^2 statistics were used to evaluate the quality of the fitted curves. The more symmetrical the filter functions on a particular scale, the better the fit of the peak function and, therefore, the better that scale describes the data.

To investigate the numerical magnitude effect, standard deviations of the Gaussian fits (σ) describing the widths of the behavioral filter curves were plotted versus numerosity for different scales. Linear functions were fitted to the standard deviations of the performance curves for each scale. Here, low slopes of the linear fits to the data would indicate constant standard deviations across all numerosities.

The discrimination performance (just noticeable difference [JND] between two stimuli) was tested as a function of the absolute numerosity. Weber's Law states that the JND of a stimulus (ΔI) is proportional to the magnitude of the physical stimulus (I) for different sensory modalities. The Weber fraction (Wb) (the ratio of ΔI and I) should be a constant if numerosities, such as various sensory phenomena, are best represented on a nonlinear scale. In order to derive Weber fractions for the tested numerosities, data were plotted on a logarithmic scale. The Weber fraction for numerosities is

$$Wb = \frac{(n_{\text{JND}} - n)}{n}, \quad (5)$$

where n is the sample numerosity of a particular filter function and n_{JND} is the numerosity that was correctly discriminated from the sample in 50% of cases.

For the analysis of reaction times (RT), only trials with correct responses to match trials were used. Response latencies of nonmatch trials were not included because the match stimulus in the second test was predictable and only used to ensure that subjects were paying attention.

RESULTS

Both monkeys and humans performed a delayed match-to-numerosity task that required them to assess the number of items in multiple-dot patterns in a sample period, retain that information in memory over a delay period, and respond to one of the two test stimuli that contained the same number of dots as were presented in the sample (Figure 1). The first part of the Results section deals with aspects specific to monkeys. In the second part, data from humans and monkeys are presented in comparison.

Monkeys' Abstract Numerical Quantity Concept

Small Numerosity Protocol

Both monkeys were previously involved in small numerosity discrimination experiments (Monkey M: 1 to 4; Monkey W: 1 to 5) for about a year. Prior to the introduction of large numerosities, they showed high discrimination performance of 81.7% (Monkey M, based on at least 252 trials per sample numerosity) and 87.4% (Monkey W; with a minimum of 575 trials per sample) for small numerosities 1 to 4 (Figure 2A and C). Both monkeys exhibited the characteristics of discrimination peak functions we had observed before, such as the

numerical distance and size effect (Nieder et al., 2006; Nieder & Miller, 2003, 2004a).

Transfer Trials Protocol

To test whether monkeys understand the concept of numerosity, and thus, also discriminate novel numerosities correctly, Monkey W was confronted with novel numerosities 6, 7, and 8 in transfer trials (see Methods). As can be seen in Figure 3A, the monkey continued to discriminate these novel numerosities; the performance curves for new numerosities showed the same characteristics and were comparable to the performance for the previously learned small set sizes.

Large Numerosity Protocol

To further demonstrate an abstract knowledge of the quantity concept, both monkeys were abruptly confronted (i.e., from one day to the other) with numerosities ranging up to 30. (Because the animals were rewarded for all correct responses, these trials are not true transfer trials). The results for the abrupt presentation of large numerosities are shown in Figure 2B and D for Monkeys M and W, respectively. Each performance curve consists on average of at least 251 standard and

Figure 2. Behavioral tuning curves derived from the monkeys' performance to the broad range of numerosities. The functions reflect the probability that a monkey judged displays in the test period as containing the same number of items as the sample numerosity. The data point at the center of each colored function indicates the correct performance in the match trials for the sample numerosities (shown in the same color above each curve). The data points to the left and the right of the center reflect performance in the nonmatch trials (i.e., where the first test display showed a smaller or larger number of items), presented as the percentage of errors for the respective nonmatch numerosity. (A) Monkey M's and (C) Monkey W's performance curves for the task with small numerosities (1 to 4). (B) Monkey M's and (D) Monkey W's filter functions for the spontaneous transfer from the discrimination of small numerosities to the discrimination of large numerosities.
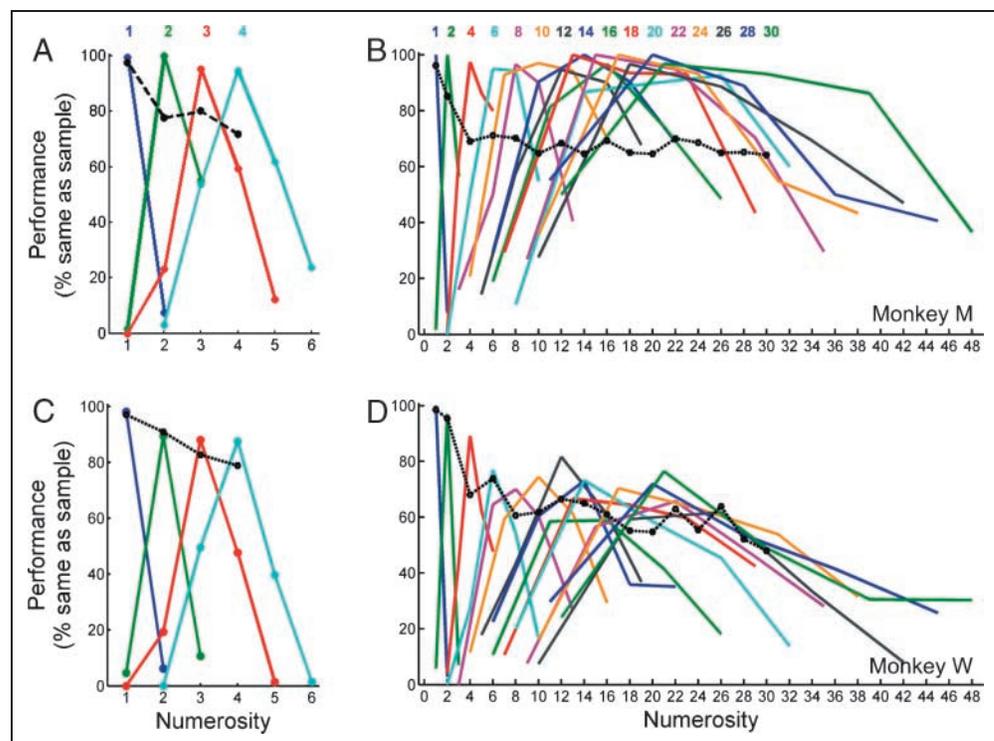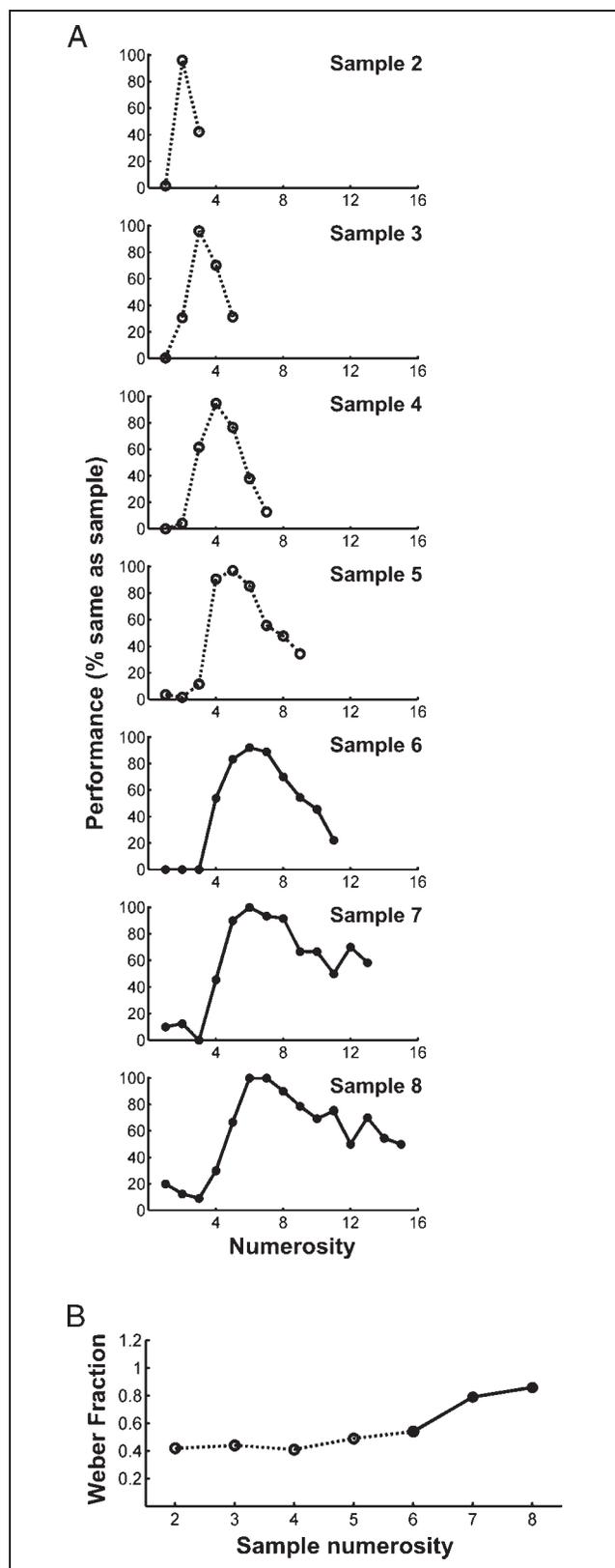
Figure 3. Monkey W's discrimination performance in the transfer trials experiment. (A) Discrimination filter functions for previously trained, small numerosities (dotted line) and new, larger numerosities (solid line). (B) Weber fractions for reinforced discriminations and transfer tests.

control trials. Nonmatch numerosities immediately adjacent to the sample numerosity (factor 0.3) were often judged erroneously as containing the same number of dots as the sample. At a larger numerical distance (factor 0.6), subjects made more correct rejections of nonmatch numerosities. Monkey M's average performance for each tested numerosity was significantly better than chance level (Binomial test, $p < .0001$). Similarly, for Monkey W, discrimination performance was better than chance level except for numerosity 20, where discrimination only approached significance (Binomial test, $p = .08$); For the highest numerosities 28 and 30, it did not differ from chance ($p > .4$). Notwithstanding, the characteristic numerical distance effect was present for the highest numerosities.

To investigate whether numerosity discrimination in monkeys improves with time, we analyzed the overall performance over several weeks of testing. For Monkey M and Monkey W, the dataset consisted of 55 and 45 days of testing, respectively, during each of which the monkeys performed, on average, 1200 (Monkey M) or 650 (Monkey W) trials. Figure 4 illustrates that the performance improved over time for both animals; a linear regression revealed significant correlations with positive slopes for both Monkey M ($r = .81$, slope $= 0.31, p < .001$, $n = 55$) and Monkey W ($r = .82$, slope $= 0.44, p < .001$, $n = 45$).

Precision Protocol

To find out which of the standard peak functions—the Gaussian function or the symmetric sigmoid function—was better suited to describe the behavioral data, both models were fitted to the detailed performance functions for sample numerosity 15 plotted on a linear and logarithmic scale. For both monkeys, these fits resulted in identical goodness-of-fit values for both peak functions (Monkey M: linear scale $r^2 = .80$, log scale $r^2 = .96$; Monkey W: linear scale $r^2 = .91$, log scale $r^2 = .97$). Therefore, for all further analyses, the Gaussian normal distribution was used to model the discrimination performance functions.

Characterization of Numerical Representations in Humans and Monkeys

The averaged performance curves derived from the large numerosity protocol revealed clear filter functions (see Methods) for all tested sample numerosities (Figure 5) for humans and monkeys. All subjects were relying on abstract quantity information rather than on lower-level visual features. For the monkeys, performance in sets of control stimuli was never worse than in standard trials. The analysis of average performance across sample numerosities 1 to 30 revealed no significant difference between standard (mean $\pm$ $SD = 69.3 \pm 3.3\%$), area control trials ($72.2 \pm 5.6\%$), and density control trials

Figure 4. Learning curves of the monkeys. Mean performance plotted as a function of days of training for (A) Monkey M and (B) Monkey W.
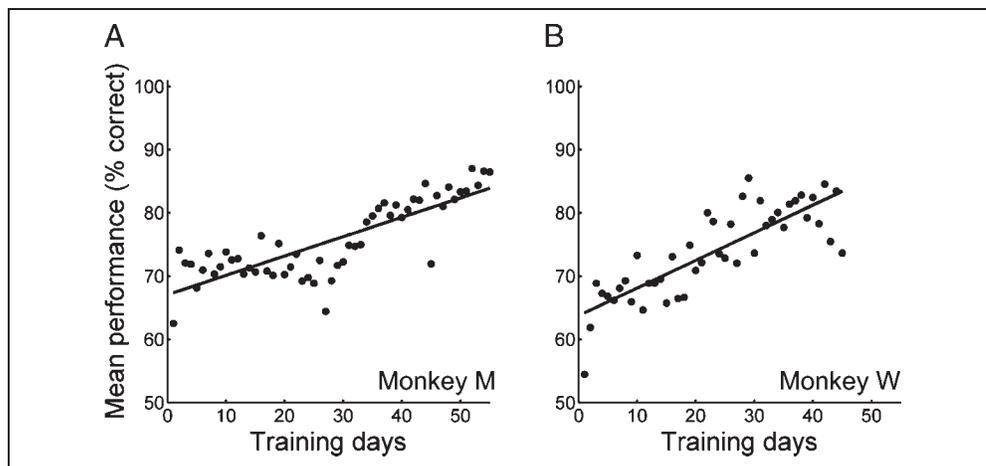
(67.3 ± 5.3%) in Monkey W (Friedman test, $p > .1$). Monkey M's averaged performance was not different for standard (72.2 ± 3.2%) and density (70.4 ± 3.5%) control trials (Wilcoxon test, $p > .2$), but he performed significantly better in area (78.2 ± 3.5%) control trials (Wilcoxon test, $p < .01$). Thus, only performance for standard trials was used for further analysis of the monkey data. Each filter function consisted of at least
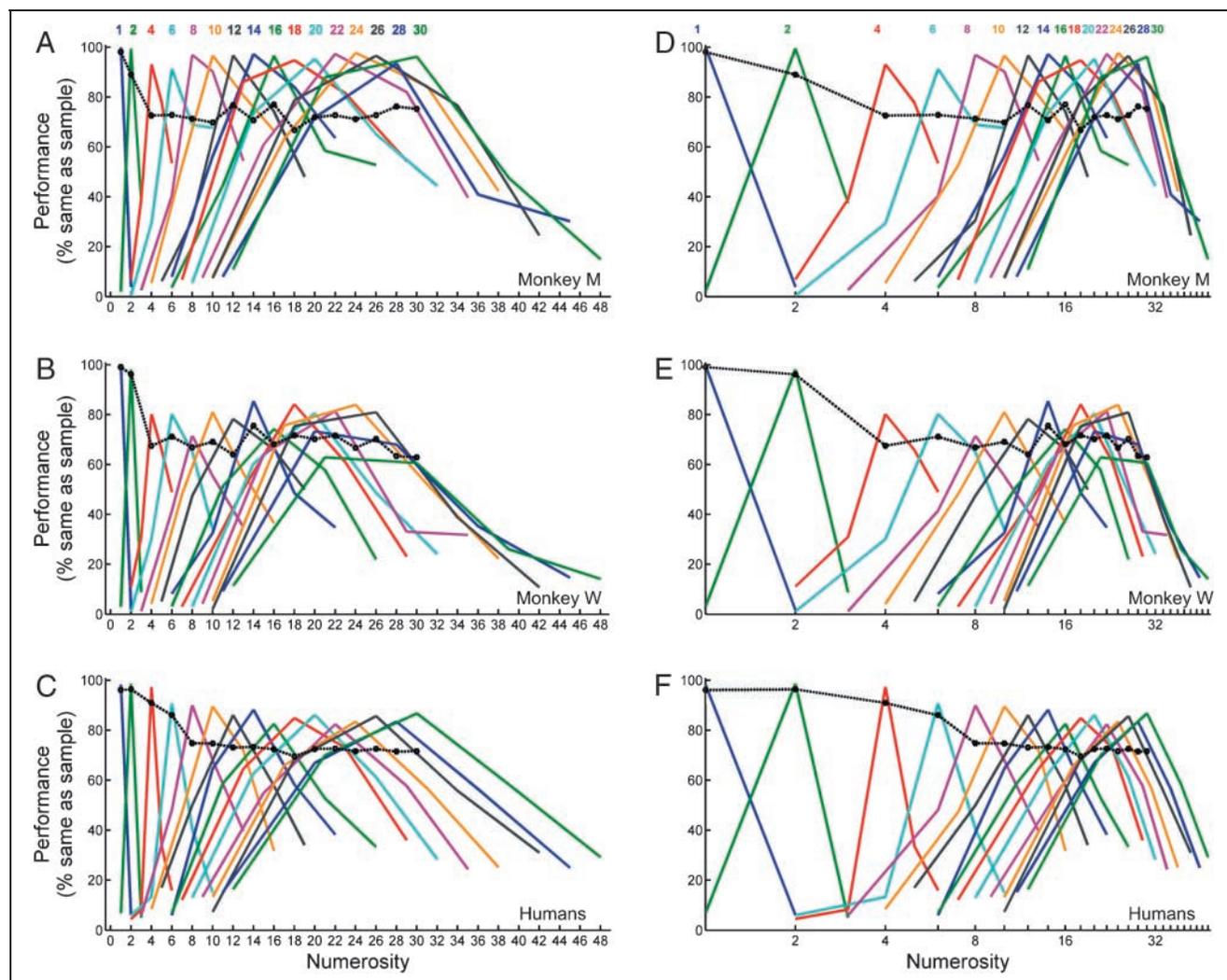

Figure 5. Scaling of performance distributions for human and monkey subjects. Same layout as in Figure 2. The behavioral tuning curves were plotted on linear [(A) Monkey M, (B) Monkey W, (C) humans] and logarithmic scales [(D) Monkey M, (E) Monkey W, (F) humans].

Table 1. Mean Goodness-of-Fit Values (r^2) of Gaussian Fits to Performance Distributions Plotted on Four Different Scales

	Linear	Pow (1/2)	Pow (1/3)	Log
Monkey M	.59 ± .14	.79 ± .11	.83 ± .10	.89 ± .07
Monkey W	.76 ± .15	.89 ± .10	.91 ± .09	.93 ± .07
Humans	.88 ± .06	.96 ± .04	.96 ± .03	.95 ± .02

The goodness-of-fit values (r^2) (mean ± *SEM*) were calculated for the fits to the curves for numerosities 4–20.

2202 trials for Monkey M and 758 for Monkey W. Humans showed no significant difference between standard (76.0 ± 7.5%), area (78.9 ± 7.0%), and density (74.6 ± 7.3%) control trials (Friedman test, $p > .06$). Therefore, standard and control trials were pooled for further analysis of human data. Each human performance curve was derived from a minimum of 1028 trials.

Both monkeys and humans showed a clear numerical distance effect; subjects made most errors when the nonmatch numerosities were adjacent to the match, but performed progressively better with increasing numerical distance between numerosities. In addition, the numerical magnitude effect was also present. Both species required greater numerical distances to discriminate between larger numerosities with equal precision.

Visual inspection of both monkeys' and humans' performances suggested that the distributions were asymmetric when plotted on a linear scale (Figure 5A–C); slopes were more moderate for numerosities larger than the sample numerosity compared to numerosities smaller than the sample. However, plotted on a logarithmic scale (Figure 5D–F), the distributions became more symmetric, suggesting that a nonlinear coding scheme might be more appropriate.

To quantify the symmetry of the discrimination performance curves for different scaling schemes, the data were plotted on a linear scale, a logarithmic scale, and scaled according to power functions with exponents of $\frac{1}{2}$ and $\frac{1}{3}$. The Gaussian distribution was fitted to the filter functions for all tested numerosities. Because the distributions were incomplete for sample numerosities 1 and 2, these data were excluded from statistical analysis. Moreover, data for sample numerosities higher than 20 were excluded because they could not be presented as match and nonmatch stimuli in a balanced way. In these cases, the monkeys may have learned that the highest numerosities were likely to be nonmatches, made fewer errors, which resulted in a distortion of the performance curves.

The mean goodness-of-fit values (Table 1) for all nonlinearly compressed scales were significantly higher (Figure 7A) for both monkeys (Wilcoxon test, $p < .008$) and for human subjects (Wilcoxon test, $p < .02$). For Monkey M, the logarithmic scaling resulted in even higher r^2 values compared to the power function scaling schemes

(Wilcoxon test, $p = .01$). The same picture—skewed distributions on a linear scale, but symmetric functions on a logarithmic scale—emerged for the most detailed performance distribution of the monkeys measured for sample numerosity 15 and a broad range of nonmatch numerosities from 1 to 29 in increments of one item (Figure 6) (see quantitative measures above).

Consistent with the logarithmic coding hypothesis (Nieder & Miller, 2003), the standard deviations of the Gaussian fits (σ) to the behavioral filter curves (Table 2) increased proportionally with an increase of numerosity on a linear scale, but were constant when the data were plotted on nonlinearly compressed scales both for humans (Figure 7C) and for monkeys (Figure 7B). (There was no significant difference between the two monkeys with respect to σ of the Gaussian fit, so the data were pooled for the plot.) In other words, data plotted on the linear scale clearly showed the numerical magnitude effect, and nonlinear scaling compensated for the proportional broadening of the performance curves with increasing stimulus magnitude. Interestingly, the slopes for both monkeys and humans were very similar (see Table 2), indicating numerical distance and size effects of comparable magnitude in both species.

The Weber fraction values (calculated for performance functions 4 to 20) were equal and constant across all small and large numerosities for both monkeys (Monkey M: mean ± *SD* = 0.60 ± 0.09; Monkey W: 0.51 ± 0.07) (Figure 7D and E). Linear fits to the Weber fractions
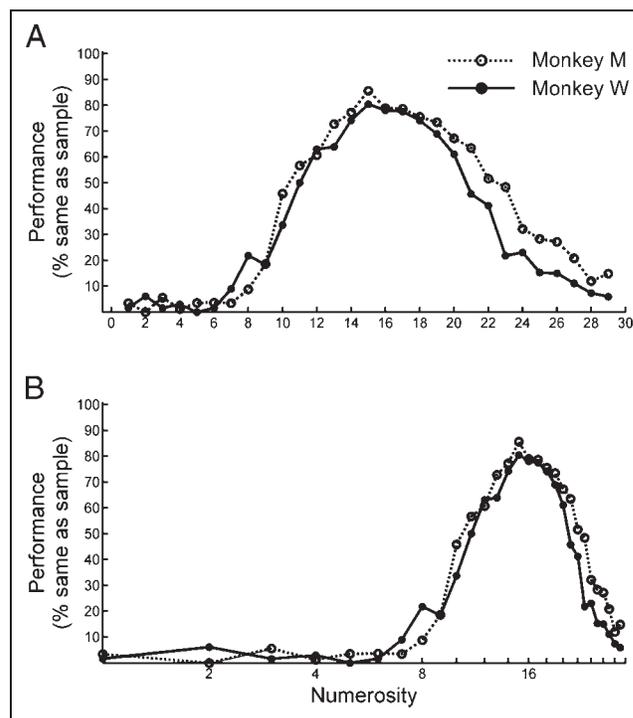

Figure 6. Detailed discrimination performance functions of both monkeys for sample numerosity 15. The functions are plotted on (A) a linear and (B) a logarithmic scale.

Table 2. Linear Fit Slopes to Sigma of Gaussian Function Fitted to Performance Distributions Plotted on Four Different Scales

	Linear	Pow (1/2)	Pow (1/3)	Log
Monkey M	0.374	0.030	0.009	0.001
Monkey W	0.304	0.022	0.006	−0.001
Humans	0.421	0.042	0.016	0.005

The slopes of the linear fits were calculated based on the fits to the curves for numerosities 4–20.

reveal slopes of 0.005 for Monkey M and −0.004 for Monkey W, confirming the constancy of the values across tested numerosities. The Weber fractions derived from the precise performance curves (Figure 6) for numerosity 15 in both monkeys (Monkey M: 0.51; Monkey W: 0.44) were in good agreement with the Weber fractions seen for the broad range of sample numerosities. Only in the transfer trials (Figure 3B), the Weber fractions of Monkey W were slightly elevated (0.56 ± 0.15). In humans, Weber fractions for numerosities 4 and 6 were evidently smaller than those for higher numerosities, indicating higher performance precision (Figure 7F). However, discrimination functions for nu-

merosities 8 to 20 exhibited a constant Weber fraction (0.55 ± 0.03; slope of linear fit: 0.006).

Reaction Times

Analysis of RTs, measured for the correctly discriminated match trials in the large numerosity protocol, showed that the smallest numerosities elicited much faster responses than larger numerosities (Figure 8). This was particularly true for the human subjects and Monkey W. For numerosities 4 and above, the RTs reached a plateau. The RTs of Monkey W (466.7 ± 39.6 msec) and human subjects (455.5 ± 11.95 msec) were very similar, whereas Monkey M reacted more than 100 msec faster to all tested numerosities (312.8 ± 8.1 msec). These faster responses may reflect the speed–accuracy trade-off, as Monkey M's discrimination performance was slightly worse (broader performance curves and higher Weber fractions) compared to Monkey W and human subjects.

DISCUSSION

In this article, we investigated the behavioral characteristics of nonverbal numerical representations in monkeys
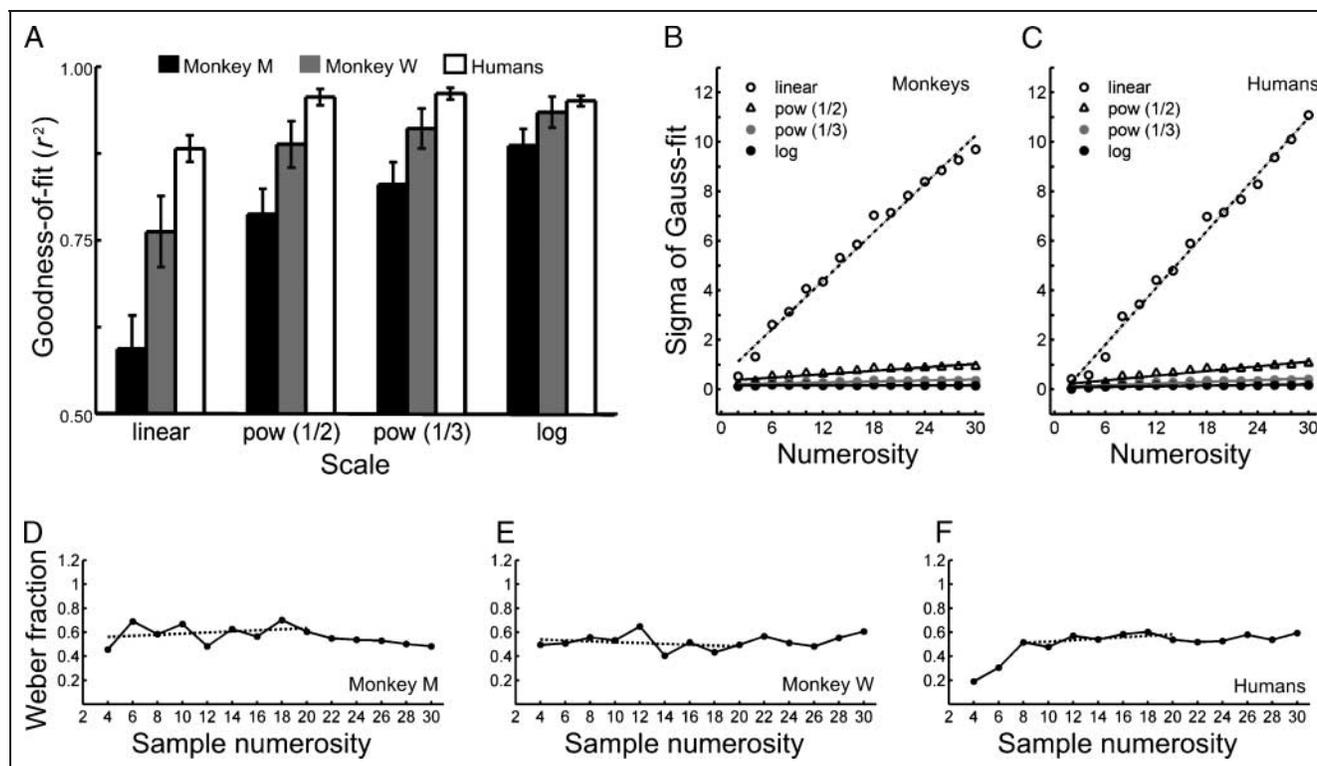
Figure 7. Quantification of performance functions for human and monkey subjects. Gaussian functions were fitted to the performance curves for numerosities 4–20 and plotted on different scales. (A) Mean goodness-of-fit of the Gaussian distributions (error bars ± *SEM*). (B, C) Standard deviations (σ) of the fitted Gaussian functions across all sample numerosities for (B) both monkeys and (C) humans. (D, E, F) Weber fractions (derived from the logarithmic scale) for the behavioral filter functions of (D) Monkey M, (E) Monkey W, and (F) human subjects. The dotted lines represent linear fits to the Weber fractions for sample numerosities 4–20 for the monkeys and 8–20 for human subjects.
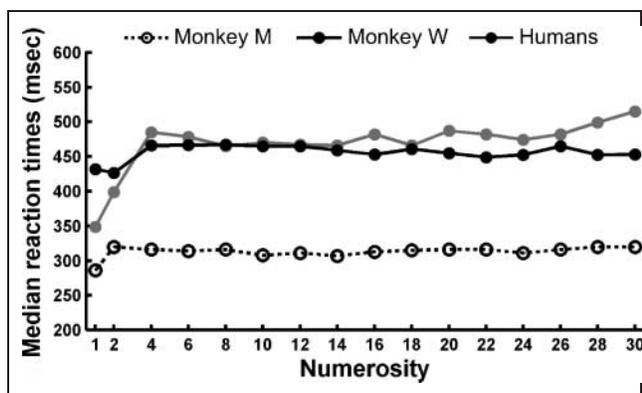
Figure 8. Median reaction times for all tested numerosities (only correct match trials).

and human subjects engaged in an identical large numerosity discrimination task. Our results confirm that monkeys can abstract the set size of items in a multiple-dot display irrespective of the appearance of the displays and low-level visual features. Moreover, monkeys were able to spontaneously generalize their numerosity discrimination performance to quantities they have never seen before. This work demonstrates that language-endowed humans and nonverbal monkeys possess a noisy quantity estimation system with similar precision. Most importantly, the detailed analysis of the performance distributions validates for both species the hypothesis of nonlinear compressed scaling for nonverbal numerosity representation.

The comparison of monkeys' and humans' performance precision for large numerosities revealed no upper limit. We observed comparable filter functions and similar average performance for humans and monkeys in all three experimental conditions (standard and both controls) for all tested numerosities (1–30). This argues convincingly for a single analog magnitude representation system for small and large numerosities. A recent brain imaging study in 3-month-old infants confirms also that analog representation of numerosities extends across small and large numbers alike (Izard, Dehaene-Lambertz, & Dehaene, 2008). When comparing the visual event-related potentials evoked by unforeseen changes in the cardinality of sets, the authors observed a shared brain response to both small (2 vs. 3) and large (4 vs. 8 and 4 vs. 12) number ranges.

Discrimination performance improved with increasing numerical distance of the test stimulus from the sample quantity (numerical distance effect). Performance distributions became progressively wider for larger set sizes (numerical size effect). These characteristics of numerical representations confirmed the analog magnitude representations for both species. Our findings are in line with studies of Beran and coworkers, who observed very similar performance of rhesus monkeys (Beran, 2007), apes (Beran, 2001, 2004), and humans (Beran, Taglialatela, Flemming, James, & Washburn, 2006) dur-

ing a task in which the subjects were required to select the larger of two sequentially presented sets of items (1–10). In all these species, performance was correlated with the ratio of the set sizes pointing to the distance and magnitude effect and the involvement of an analog magnitude estimation process. Most recently, it has been demonstrated that monkeys can even perform basic arithmetic operations, such as adding numerosities, based on the analog magnitude system (Cantlon & Brannon, 2007). Of course, language-endowed humans using number symbols will always outperform animals in more demanding mathematical tasks. Nevertheless, these data argue that an understanding of numerical quantity is deeply rooted in the primate brain as a fundamental determinant of higher-level numerical cognition.

The response latencies of the human and monkey subjects revealed similar functions for the tested numerosities. An increase in RTs was detected only for numerosities 1 to 2 (humans) and 1 to 4 (monkeys); for larger numerosities a plateau was reached. A direct comparison of small RTs of humans and monkeys is difficult because human RTs for trials with small numerosities might interfere with serial counting processes. Because we did not observe an increase of the RTs for numerosities 6 and larger, it is very likely that numerical information was extracted by parallel mechanisms (Barth et al., 2003). These findings are in line with results of other large numerosity discrimination experiments using limited sample presentation times in rhesus monkeys (Jordan & Brannon, 2006a), chimpanzees, and humans (Tomonaga & Matsuzawa, 2002; Mandler & Shebo, 1982).

Abstract Numerical Quantity Concept

Our dataset clearly shows the monkeys' abstract understanding of the concept of numerical quantity. First, the monkeys (as well as human subjects) did not rely on low-level visual features to solve the task. Completely new sets of stimuli (including area or density control stimuli) were generated for each experimental session, so the sizes and arrangements of items in the displays were randomized. Thus, it was not possible for the subjects to memorize particular nonnumerical features of the stimuli and use them to solve the task. Second, the monkeys were instantly able to discriminate a range of large numerosities they had never been tested on before. In transfer trials, the monkey reliably discriminated novel set sizes of 6, 7, and 8 dots. In addition, both monkeys showed spontaneous generalization to abruptly introduced novel large numerosities with the same discrimination characteristics as for the well-trained small numerosities (albeit more noisy because of the smaller number of trials).

Other studies have also shown that monkeys were able to transfer their numerical knowledge to values they have never seen before. Brannon and Terrace (1998)

trained monkeys to respond to set sizes 1 to 4 in an ascending order and these monkeys subsequently succeed to order novel numerosities 5 to 9 in a transfer test. In addition, Nieder et al. (2006) reported successful transfer of the discrimination behavior in a sequential protocol. Monkeys learned to discriminate sequential numerosities 2 and 4 and succeeded to discriminate sequential numerosity 3 from 2 and 4 in transfer tests. Such findings argue for a true understanding of the cardinality of sets, and thus, the concept of numerical quantity. For pragmatic reasons, we only tested numerosities up to 30 in the current study, but there is every reason to believe that an infinite range of numerosities can be discriminated by the monkeys—of course, at the expense of the decreasing precision due to Weber's Law.

Weber Fraction

Our results show that nonverbal quantity representations in both primate species obey Weber's Law. Characteristic Weber fractions were constant across set sizes (4–20) for Monkey M (0.60) and Monkey W (0.51). This indicates that monkeys use only the analog magnitude system for the explicit representation of numerical quantity. These results are in good agreement with data reported in other studies. For example, Jordan and Brannon (2006a) reported Weber fractions of 0.47 and 0.48 for rhesus monkeys engaged in a forced-choice delayed matching-to-sample protocol. In this protocol, the sample stimulus could adopt any value between 1 and 9, but the test was a forced choice between two fixed values (2 and 8). Moreover, the same work showed that monkeys improved their performance precision due to training with decreasing Weber fractions from 0.58 to 0.32. For rhesus monkeys performing an ordinal comparison task (Cantlon & Brannon, 2006), a Weber fraction of 0.38 was found. The variation in the reported Weber fractions in rhesus monkeys might be due to different task demands, and it is likely influenced by the stage of training of the animal on the specific task.

Compared to the two monkeys, the performance of the human subjects in our study was very similar for large numerosities (beyond 6). Most importantly, the average Weber fractions of humans (0.55) closely matched the values of the two monkeys (0.51 and 0.60). Interestingly, Piazza et al. (2004) found considerably lower Weber fractions of 0.17 for human adults discriminating multiple-dots stimuli. A possible reason for this difference might be the higher task demand in the current study. Whereas Piazza et al. only used numerosities 16 and 32 as sample, our task design required the subjects to discriminate almost any given numerosity from the other. Cantlon and Brannon (2006) found a Weber fraction of 0.26 in humans performing the same ordinal comparison task as the monkeys. However, the more precise performance of human subjects in Cantlon and Brannon compared to monkeys was accompanied

by longer RTs for humans (on average an additional 100 msec). This indicates that a tradeoff between speed and accuracy might be an important reason for varying Weber fractions.

Humans showed a clear performance advantage, and thus, smaller Weber fractions for small numerosities. A likely reason is that, although stimulus presentation times were short, humans were able to symbolically enumerate the items in stimulus displays with small quantities. Alternatively, humans may, indeed, possess a separate and precise system to represent small numerosities (Mandler & Shebo, 1982). To address this issue, shorter sample presentation times need to be tested for human subjects.

Scaling

Most importantly, our study shows that the nonverbal representation of numerosities in both humans and monkeys is described best on a nonlinear scale. Plotted on a linear scale, performance distributions were asymmetric. In contrast, nonlinear scaling of the filter functions resulted in symmetric peak functions which were reflected in significantly higher goodness-of-fit values and constant standard deviations of the fitted Gaussian distributions for all tested numerosities.

In more detail, our results indicate that the logarithmic scale (Fechner's Law) described the data even better than power function scales (Stevens' Law). The goodness-of-fit values of the monkeys' performance functions were highest for the logarithmic scaling. In addition, the standard deviations of log-transformed discrimination functions remained more stable across the tested numerosities compared to the representations on power function scales. These findings for large numerosities are fully consistent with a previous report on small numerosities (Nieder & Miller, 2003).

The monkeys' behavioral numerosity representations are more precise than the average neuronal tuning functions of single cells (Nieder & Merten, 2007). The population of numerosity selective neurons showed a mean neuronal Weber fraction of 1.20 for numerosities 1 to 30. In other words, the averaged precision of all cells is considerably lower than numerosity discrimination performance. However, this finding is consistent with the "lower envelope principle" (Parker & Newsome, 1998), which argues that discrimination thresholds are based on the most sensitive neurons of the population and not on the population average. In agreement with the behavioral data, the neuronal tuning functions obeyed the Fechner Law and were best described on a logarithmically compressed scale.

Despite these overall similarities in the nonverbal discrimination performances between human and nonhuman primates, the logarithmic scaling scheme was less pronounced in humans. The goodness-of-fit values (r^2) were high for all scales and the (significant) difference

in absolute mean r^2 values for linear versus nonlinear representations were rather small. Assuming the findings in monkeys mirror the original numerosity representations in non- and preverbal primates, analog magnitude representations in enumerate humans seem to experience a shift toward linear scaling. We suspect that this shift might stem from a gradual cultural transformation of a nonverbal logarithmic scheme to a linear scheme during the course of mathematical education. This assumption is supported by work by Siegler and colleagues, who examined the representation of numerical quantity in children and adults asking them to estimate positions of numbers on a number line. They found a shift from reliance on logarithmic to linear representations of numerical magnitudes between kindergartners and second graders using the 0–100 number lines (Booth & Siegler, 2006; Siegler & Booth, 2004) and between second and sixth graders and adults using the 0–1000 number lines (Siegler & Opfer, 2003). In addition, the same effect was found for other types of estimation tasks, like computational, numerosity, and length measurement estimation (Booth & Siegler, 2006; Siegler & Booth, 2004). Children with minor mathematical training relied on logarithmic numerical scales that are probably derived from nonverbal quantity representations. On the other hand, advanced mathematical education led to linear representations of numerical magnitudes. To test whether the number line representation might be modified with age and numerical experience, and whether language and cultural mathematic influences might have altered the scaling in adults, children with different levels of numerical experience should be tested on the same numerosity discrimination protocol. Support for this idea comes from studies using the bisection protocol. Testing children (4-, 5-, and 6-year-olds) and rhesus monkeys on a very same task revealed a similar representation of numerosity as analog magnitudes obeying Weber's Law (Beran, Johnson-Pynn, & Ready, 2007; Jordan & Brannon, 2006b).

In conclusion, the striking similarities in numerosity discrimination performance of both primate species corroborate the view that numerical cognition has not emerged de novo in humans, but has rather built on a biological precursor system (Dehaene, 1997; Danzig, 1954) and that humans and monkeys share an ancient, nonverbal quantification system (Cantlon & Brannon, 2006). Our study is an evidence for the representation of small and large numerosities via a single analog magnitude system, best described on a logarithmic scale.

Acknowledgments

Reprint requests should be sent to Andreas Nieder, Department of Animal Physiology, Zoological Institute, University of Tuebingen, Auf der Morgenstelle 28, 72076 Tübingen, Germany, or via e-mail: andreas.nieder@uni-tuebingen.de.

REFERENCES

Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number representations in adults. *Cognition, 86,* 201–221.

Beran, M. J. (2001). Summation and numerousness judgments of sequentially presented sets of items by chimpanzees (Pan troglodytes). *Journal of Comparative Psychology, 115,* 181–191.

Beran, M. J. (2004). Chimpanzees (*Pan troglodytes*) respond to nonvisible sets after one-by-one addition and removal of items. *Journal of Comparative Psychology, 118,* 25–36.

Beran, M. J. (2007). Rhesus monkeys (*Macaca mulatta*) enumerate large and small sequentially presented sets of items using analog numerical representations. *Journal of Experimental Psychology: Animal Behavior Processes, 33,* 42–54.

Beran, M. J., Johnson-Pynn, J. S., & Ready, C. (2007). Quantity representation in children and rhesus monkeys: Linear versus logarithmic scales. *Journal of Experimental Child Psychology, 100,* 225–233.

Beran, M. J., Taglialatela, L. A., Flemming, T. M., James, F. M., & Washburn, D. A. (2006). Nonverbal estimation during numerosity judgements by adult humans. *Quarterly Journal of Experimental Psychology (Colchester), 59,* 2065–2082.

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology, 42,* 189–201.

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., Wusthoff, C. J., Gallistel, C. R., & Gibbon, J. (2001). Numerical subtraction in the pigeon: Evidence for a linear subjective number scale. *Psychological Science, 12,* 238–243.

Cantlon, J. F., & Brannon, E. M. (2006). Shared system for ordering small and large numbers in monkeys and humans. *Psychological Science, 17,* 401–406.

Cantlon, J. F., & Brannon, E. M. (2007). Basic math in monkeys and college students. *PLoS Biology, 5,* e328.

Danzig, T. (1954). *Number, the language of science.* New York: Free.

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

Dehaene, S. (1997). *The number sense.* New York: Oxford University Press.

Dehaene, S. (2001). Subtracting pigeons: Logarithmic or linear? *Psychological Science, 12,* 244–246; discussion 247.

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

Emmerton, J., & Renner, J. C. (2006). Scalar effects in the visual discrimination of numerosity by pigeons. *Learning & Behavior, 34,* 176–192.

Fechner, G. T. (1860). *Elemente der psychophysik* (Vol. 3). Leipzig: Breitkopf and Härtel.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8,* 307–314.

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

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review, 84,* 279–335.

Gibbon, J., & Church, R. M. (1981). Time left: Linear versus logarithmic subjective time. *Journal of Experimental Psychology: Animal Behavior Processes, 7,* 87–107.

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science, 306,* 496–499.

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

Izard, V., Dehaene-Lambertz, G., & Dehaene, S. (2008). Distinct cerebral pathways for object identity and number in human infants. *PLoS Biology, 6,* 1–10.

Jordan, K. E., & Brannon, E. M. (2006a). Weber's Law influences numerical representations in rhesus macaques (*Macaca mulatta*). *Animal Cognition, 9,* 159–172.

Jordan, K. E., & Brannon, E. M. (2006b). A common representational system governed by Weber's law: Nonverbal numerical similarity judgments in 6-year-olds and rhesus macaques. *Journal of Experimental Child Psychology, 95,* 215–229.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24,* 175–219.

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

Nieder, A. (2005). Counting on neurons: The neurobiology of numerical competence. *Nature Reviews Neuroscience, 6,* 177–190.

Nieder, A., Diester, I., & Tudusciuc, O. (2006). Temporal and spatial enumeration processes in the primate parietal cortex. *Science, 313,* 1431–1435.

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., & Merten, K. (2007). A labeled-line code for small and large numerosities in the monkey prefrontal cortex. *Journal of Neuroscience, 27,* 5986–5993.

Nieder, A., & Miller, E. K. (2003). Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron, 37,* 149–157.

Nieder, A., & Miller, E. K. (2004a). Analog numerical representations in rhesus monkeys: Evidence for parallel processing. *Journal of Cognitive Neuroscience, 16,* 889–901.

Nieder, A., & Miller, E. K. (2004b). A parieto-frontal network for visual numerical information in the monkey. *Proceedings of the National Academy of Sciences, U.S.A., 101,* 7457–7462.

Parker, A. J., & Newsome, W. T. (1998). Sense and the single neuron: Probing the physiology of perception. *Annual Review of Neuroscience, 21,* 227–277.

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron, 44,* 547–555.

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science, 306,* 499–503.

Roberts, W. A. (2005). How do pigeons represent numbers? Studies of number scale bisection. *Behavioural Processes, 69,* 33–43.

Roberts, W. A. (2006). Evidence that pigeons represent both time and number on a logarithmic scale. *Behavioural Processes, 72,* 207–214.

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development, 75,* 428–444.

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science, 14,* 237–243.

Stevens, S. S. (1861). To honor Fechner and repeal his law. *Science, 133,* 80–86.

Tomonaga, M., & Matsuzawa, T. (2002). Enumeration of briefly presented items by the chimpanzee (*Pan troglodytes*) and humans (*Homo sapiens*). *Animal Learning & Behavior, 30,* 143–157.

van Oeffelen, M. P., & Vos, P. G. (1982). A probabilistic model for the discrimination of visual number. *Perception & Psychophysics, 32,* 163–170.

Weber, E. H. (1850). Der Tastsinn und das Gemeingefühl. In R. Wagner (Ed.), *Handwörterbuch der physiologie* (Vol. 3, part 2, pp. 481–588). Braunschweig: Vieweg.

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representations. *Psychological Science, 10,* 130–137.

Xu, F. (2003). Numerosity discrimination in infants: Evidence for two systems of representations. *Cognition, 89,* B15–B25.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74,* B1–B11.