Neuron, Vol. 37, 149–157, January 9, 2003, Copyright ©2003 by Cell Press

Coding of Cognitive Magnitude: Compressed Scaling of Numerical Information in the Primate Prefrontal Cortex

Andreas Nieder* and Earl K. Miller
Picower Center for Learning and Memory
RIKEN-MIT Neuroscience Research Center and
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Summary

Whether cognitive representations are better conceived as language-based, symbolic representations or perceptually related, analog representations is a subject of debate. If cognitive processes parallel perceptual processes, then fundamental psychophysical laws should hold for each. To test this, we analyzed both behavioral and neuronal representations of numerosity in the prefrontal cortex of rhesus monkeys. The data were best described by a nonlinearly compressed scaling of numerical information, as postulated by the Weber-Fechner law or Stevens' law for psychophysical/sensory magnitudes. This nonlinear compression was observed on the neural level during the acquisition phase of the task and maintained through the memory phase with no further compression. These results suggest that certain cognitive and perceptual/sensory representations share the same fundamental mechanisms and neural coding schemes.

Introduction

In many prevalent theories of cognition, the information comprising lower-level sensory and higher-level cognitive representations are thought to be fundamentally different. For example, some computational theories of mind (Fodor, 1975; Pylyshyn, 1984) suggest that virtually all cognitive representations are part of a language-based, symbolic system with combinatorial syntax and semantics (i.e., propositional). By contrast, a competing view, the analog coding hypothesis (Shepard and Metzler, 1971; Shepard and Podgorny, 1978; Kosslyn, 1994; Barsalou, 1999), emphasizes a continuum between mental processes; it posits that higher-level cognitive representations are fundamentally similar to lower-level sensory representations and thus should follow the same laws and exhibit similar attributes.

Numerical judgments offer an opportunity for exploring this issue. Although this ability is considered a higher cognitive phenomenon, numerical judgments show magnitude effects like sensory judgments; they are influenced by the numerical distance between two values (the "numerical distance effect") and their absolute magnitude (the "numerical magnitude effect") (Moyer and Landauer, 1967; Dehaene, 1992; Dehaene et al., 1998). At the same time, numerical judgments are clearly different from sensory processes because they are abstract, irrespective of exact physical appearance (e.g., two bicycles and two cows are both "two"). Thus, transduction processes at the level of the sensory epithelium, which can account for magnitude effects in sensory neurons, cannot explain numerical magnitude effects.

So, we tested whether nonverbal numerical judgments follow the same psychophysical laws that sensory judgments follow. This includes the Weber law (Weber, 1850): the amount ΔI (also termed "just noticeable difference") that needs to be added or subtracted to/from the magnitude of a stimulus I is proportional to I, so that the so-called Weber fraction ($\Delta I/I$) is a constant. Fechner (Fechner, 1860) elaborated on this in his stipulation that linear increments in sensation S are proportional to the logarithm of stimulus magnitude I, a relationship known as Weber-Fechner law ($S = k \cdot log(I)$). This nonlinear scaling of stimulus magnitude has also been captured by Stevens' power law ($S = k \cdot I^n$), which postulates that sensation is a power function of the stimulus magnitude (Stevens, 1961). These fundamental laws are largely valid for general sensory phenomena and also account for many properties of sensory neurons; the tuning of sensory neurons are often highly compressed functions of stimulus intensity, typically with logarithmic or weak power-law dependencies (Dayan and Abbott, 2001). Therefore, if a continuum between perceptual and cognitive processes exists, numerical representations should also be represented on a nonlinear, compressed "number scale."

However, as has already been pointed out (MacKay, 1963; Johnson et al., 2002), studying behavior alone cannot tell us whether numerical judgments and other magnitude estimations truly depend on neural representations that follow these laws. The behavioral outcome of an estimation task may simply be the result of multiple, diverse scaling schemes at different processing stages. In other words, at the behavioral level, it may look like magnitude estimations are following Weber-Fechner laws, but the underlying neural code could actually look quite different. Such an argument has been used in models of numerical representation, and, in fact, it remains controversial whether linear (Gibbon, 1977; Gibbon and Church, 1981; Gallistel and Gelman, 2000; Brannon et al., 2001) or nonlinear (Van Oeffelen and Vos, 1982; Dehaene and Mehler, 1992; Dehaene and Changeux, 1993; Dehaene, 2001) coding underlies numerical judgments (Figure 1).

Here, we examine coding schemes of numerical representations in the monkey prefrontal cortex (PFC). The main goal was to determine if the behavioral and neural data are consistent with psychophysical laws that predict compressed scaling of numerical information. We report that both behavioral and neural measures of visual quantities are indeed better described by a nonlinear than a linear coding scheme, and they do seem to follow the Weber-Fechner law. Overall, the results suggest that nonverbal numerical judgments as one example of higher-level cognitive phenomena are fundamentally similar to those underlying lower-level sensory judgments and thus support analog coding over propositional models of cognitive representations.

*Correspondence: nieder@mit.edu

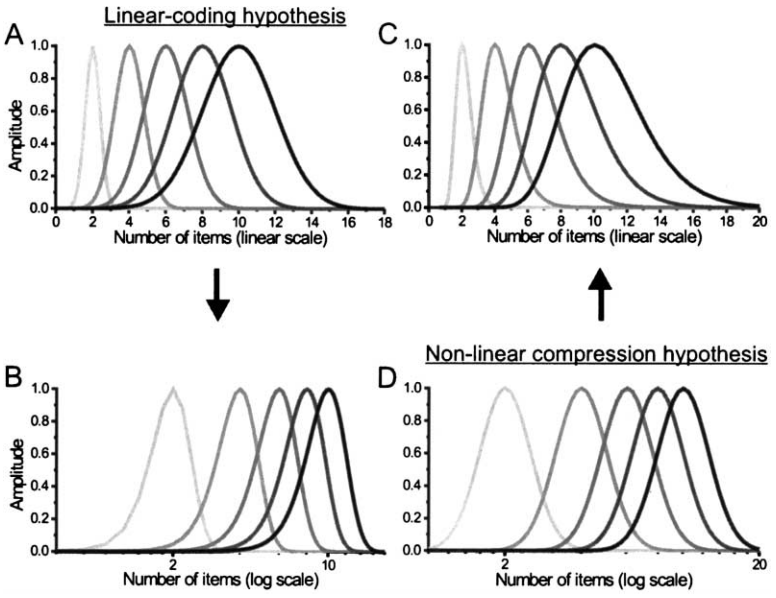

Figure 1. Comparison of the Linear-Coding and Logarithmic-Compression Hypotheses

(A and B) Linear-coding hypothesis. (A) Internal representations consist of symmetric normal distributions on a linear scale that are centered on each number and become progressively wider in proportion to increasing magnitude ("scalar variability"). The ratio of the standard deviation to the mean (i.e., the coefficient of variation) is constant across a range of quantities. (B) The distributions plotted on a logarithmic scale become asymmetric with a shallower slope toward smaller numbers.

(C and D) Logarithmic-compression hypotheses. (D) Quantities are represented on a power-function or logarithmically compressed scale with constant variability across different numbers. Underlying representations will be Gaussian on a log scale. The accuracy of the representations stays invariable with increasing size of a quantity. Thus, the standard deviation, not the coefficient of variation, is constant across quantities. (C) When transformed to a linear scale, the distributions are asymmetric, but now with a shallower slope toward higher numbers relative to the mean. (Amplitudes of all distributions are kept constant for simplicity.) Both models can account for numerical magnitude and distance effects.

Results

Behavior

We trained monkeys on a delayed match-to-numerosity task (Figure 2A) that required them to judge whether successive visual displays contained the same small number of pseudorandomly placed items (Nieder et al., 2002). To perform this task, monkeys needed to abstract the quantity of items from visual displays that varied widely in appearance and then hold that information in

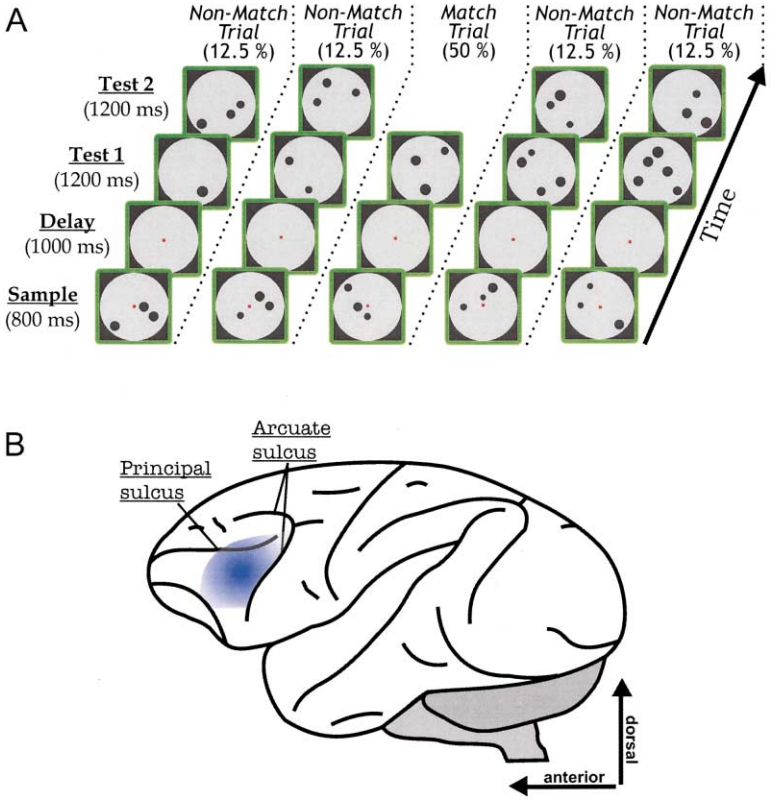

Figure 2. Behavioral Protocol and Recording Site

(A) Delayed match-to-numerosity task. Fixating monkeys were cued for a given numerosity by a sample display. The subjects had to memorize the numerosity in a 1 s delay period and match it to a subsequent test stimulus (either the first or the second test stimulus was correct in 50% of the cases) by releasing a lever. During single-unit recording, the nonmatch stimuli were one numerosity up or down. For the behavioral tests (shown here), however, a larger range of nonmatch numerosities was shown. As an example, the figure displays the protocol for sample numerosity "five."

(B) Area of recording sites (blue) in both monkeys. Mainly neurons from the ventral bank of the prefrontal cortex were recorded.

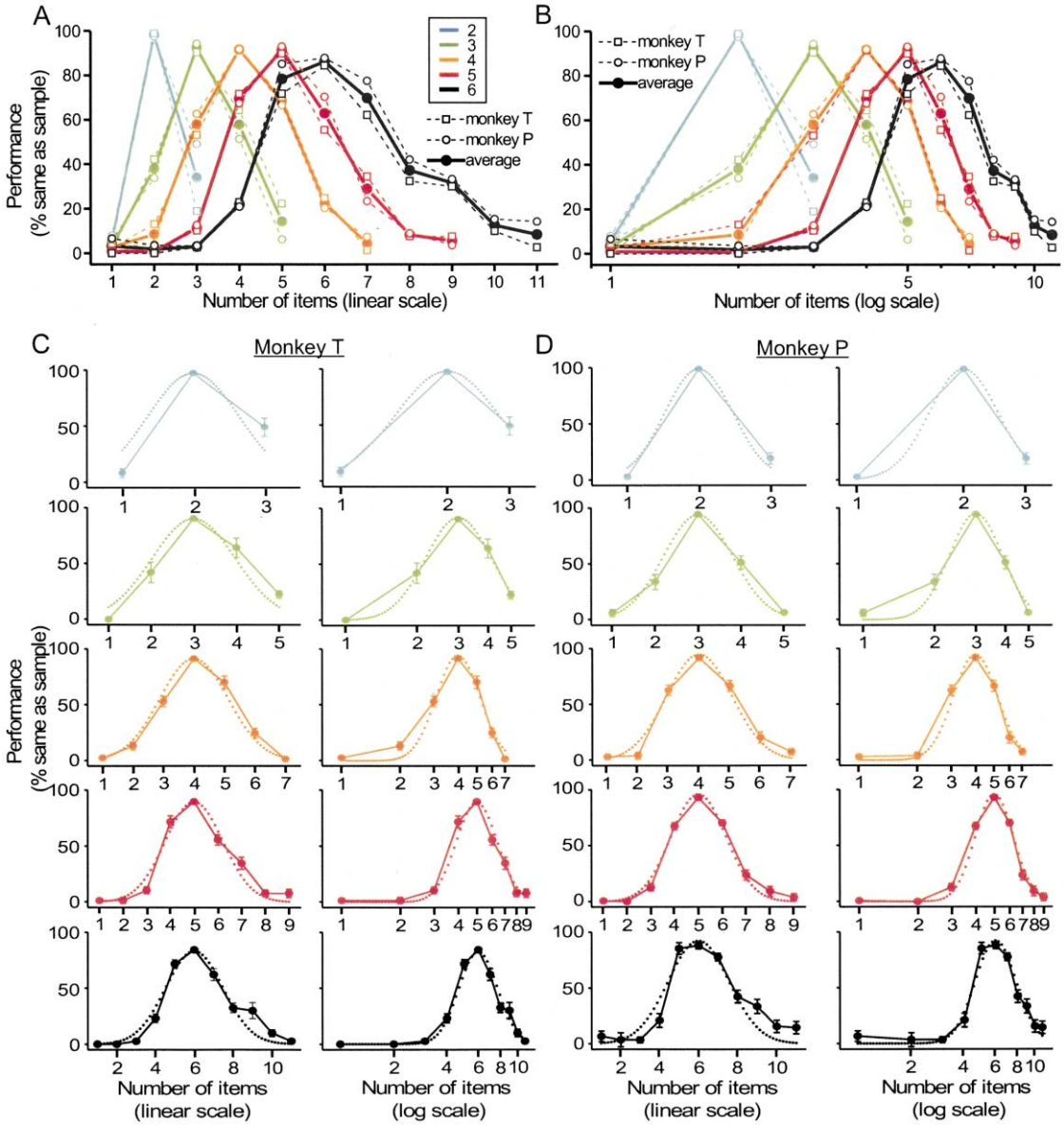

Figure 3. Quantification of Behavioral Performance Curves

The behavioral performance for both monkeys indicated whether they judged the first test stimulus (after the delay) as containing the same number of items as the sample display ("% same as sample"). Colors represent performance curves for a given sample numerosity. Behavioral filter functions are plotted on linear (A) and logarithmic (B) scales. The functions are asymmetric when plotted on a linear scale (note the shallower slope toward higher numerosities) (A), but are symmetric when plotted on the nonlinear logarithmic scale (B). (C and D) Individual distributions for all tested numerosities for both monkeys. Columns to the left in (C) and (D) show data plotted on a linear scale, contrasted by the same functions plotted on a logarithmic scale. To evaluate the symmetry of the behavioral and filter functions in the monkeys, a normal distribution (Gaussian, indicated by dotted lines) was fitted to the measured data, and the goodness-of-fit was derived as quantitative measure (error bars, ±SEM).

memory over a short delay. Performance on sets of control stimuli confirmed that the monkeys were relying on abstract quantity information rather than on the exact appearance of the displays or lower-level visual features (area, circumference, density, or geometric arrangement of the dots) (Nieder et al., 2002). In a new set of behavioral experiments, we tested the monkeys' performance to an expanded range of nonmatch numerosities (Figures 3A and 3B). Monkeys made more errors when the numerosities were adjacent and performed progressively better as numerical distance between two displays increased (numerical distance effect). For larger quantities, the two numerosities had to be numerically more distant for performance to reach the level obtained with smaller quantities and closer numerical distance (numerical size effect).

The distributions of both monkeys' performance were asymmetric when plotted on a linear scale (Figure 3A); slopes were shallower for numerosities higher than the sample numerosity (i.e., the center of each distribution)

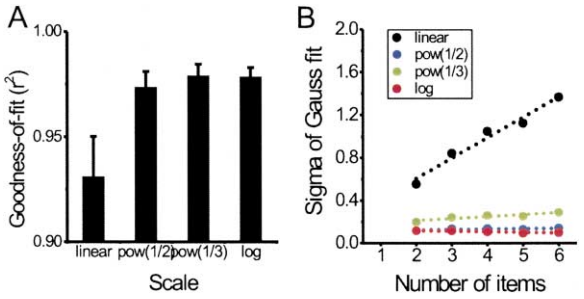

Figure 4. Quantification of Gauss Fits to the Behavioral Data

(A) Goodness-of-fit of Gauss functions fitted to the performance curves plotted on different scales. The goodness-of-fit was significantly better for the three nonlinear scaling schemes (error bars, ±SEM).

(B) The standard deviation (sigma) of the Gauss fits for nonlinear scaling plotted against the center of the Gauss function (which is identical to the numerosity of the match stimulus). Dotted lines indicate linear fits. (The values of sigma are related to the specific compression scheme.)

than for numerosities lower than the sample. However, when plotted on a logarithmic scale, the distributions were more symmetric (Figure 3B), suggesting that a nonlinear coding scheme might be more appropriate for this data. To quantify this, we first determined whether linear or nonlinear scaling models provided a superior fit to this behavioral data. We plotted the data along four scales: a linear scale, a power function with an exponent of 0.5, a power function with an exponent of 0.33, and a logarithmic scale (see Experimental Procedures). To evaluate the symmetry of the behavioral and filter functions in the monkeys, we fitted a normal distribution (Gaussian) to the measured data. The Gaussian was chosen because it represents the standard symmetric distribution and thus provided a means to compare the behavioral functions. The scales become increasingly nonlinearly compressed along this sequence. Gauss functions were fit to the distributions, and the goodness-of-fit of the resulting plots to normal (Gaussian) distribution was derived (Figures 3C and 3D). The more symmetrical the distribution, the better this fit, and, therefore, the better that scale describes the data.

Goodness-of-fit values (r^2) were significantly different for the four different scaling schemes ($p = 0.002$, Friedman test) (Figure 4A); all three nonlinear scales resulted in significantly better fits than the linear scale ($p \leq 0.01$, Wilcoxon signed ranks test). The mean goodness-of-fit values for the linear scale, the power function with an exponent of 0.5, the power function with an exponent of 0.33, and the logarithmic scale were 0.93, 0.97, 0.98, and 0.98, respectively. (This effect was not due to truncating higher numbers on a logarithmic scale.) Further, the variance of the distributions for each numerosity (i.e., sigma of the Gauss fit to the performance curves) was constant when the data was plotted on a power function scale with 0.33 exponent (slope of linear fit = 0.004) and the logarithmic scale (slope of linear fit = −0.006) (Figure 4B), which is predicted by a nonlinear coding model of numerosity (Van Oeffelen and Vos, 1982; Dehaene and Mehler, 1992; Dehaene and Changeux, 1993; Dehaene, 2001). Thus, performance data for numerosity judgments is better described using a power function-compressed or logarithmically compressed scale, as opposed to a linear scale; it seems that these data follow the Weber-Fechner law.

Neurophysiological Data

We performed similar analyses on the activity of neurons from the lateral PFC (Figure 2B) of these monkeys obtained during task performance. Data were analyzed separately for the acquisition (sample epoch) and the retention phase (delay epoch).

From a total sample of 352 tested cells, 131 and 111 showed significant tuning to one of the five displayed numerosities during sample presentation and the subsequent memory delay, respectively (ANOVA, $p < 0.01$). Their neural activity formed numerosity filter functions with activity declining progressively with increasing numerical distance from a preferred number (i.e., that eliciting maximal activity) (Nieder et al., 2002). Population mean neural filter functions were constructed by averaging activity across all neurons that preferred a given number. The neuron data mirrored the numerical distance and magnitude effects by the fact that the neural filters were also an inverted V-shape that became less selective (wider) with increasing preferred numerosity.

Much like the behavioral data, the neural filter functions were asymmetric when plotted on a linear scale, but more symmetric when plotted on a logarithmic scale. Figures 5 and 6 show this data separately for the sample and delay epochs, respectively. We applied the same goodness-of-fit tests that were applied to the behavioral data. Once again, the four different scaling schemes resulted in significantly different goodness-of-fit values ($p = 0.02$, Friedman test). For both the sample and delay epochs, the (nonlinear) power function and logarithmic scales provided a better fit to the data than the linear scale ($p < 0.05$, Wilcoxon signed ranks test). Goodness-of-fit for the sample and delay data are shown in Figures 5C and 6C, respectively. For the sample period, the mean goodness-of-fit values for the linear scale, the power function with an exponent of 0.5, the power function with an exponent of 0.33, and the logarithmic scale were 0.72, 0.83, 0.86, and 0.89. For the delay period, the average values were 0.84, 0.92, 0.93, and 0.96, respectively. Also similar to the behavioral data, the variance of neural distributions was more or less constant with increasing preferred numerosity when the data were plotted on a logarithmic scale, but increased with numerosity when the data were plotted on a linear scale (Figures 5D and 6D), once again as predicted by a nonlinear coding model (Van Oeffelen and Vos, 1982; Dehaene and Mehler, 1992; Dehaene and Changeux, 1993; Dehaene, 2001). This was true both for the sample and the delay period (slope of linear fit: sample = −0.001, delay = −0.012). Thus, there was little or no change in the coding scheme between the sample and delay epochs. The nature of the neural representation was established during the initial acquisition process and then maintained over the delay. This will become important when discussing potential "counting" models (see Discussion).

If the behavioral and neural data are truly following the Weber-Fechner law, then it is important to demonstrate

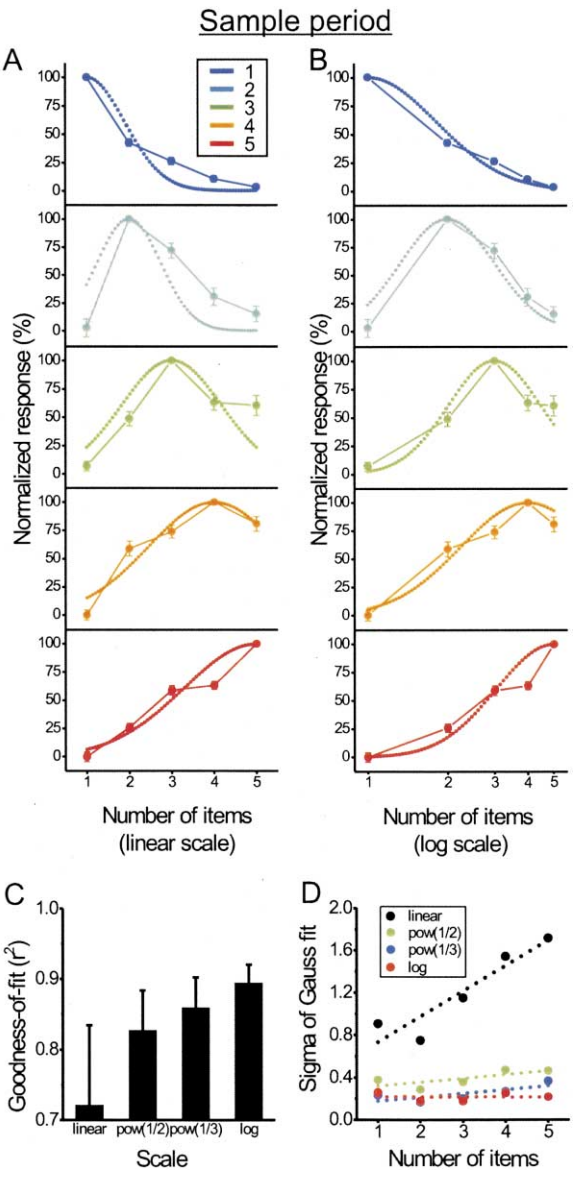

Figure 5. Neural Representation of Numerosities in the PFC during the Sample Period

(A) The neural filter functions are asymmetric on a linear scale (note the shallower slope toward higher values for preferred numerosity "2," for example).

(B) Logarithmic transformation of the filter functions results in more symmetric distributions.

(C) Goodness-of-fit for the four different scaling schemes.

(D) Standard deviation values for the scaling schemes across preferred quantities (error bars, ±SEM).

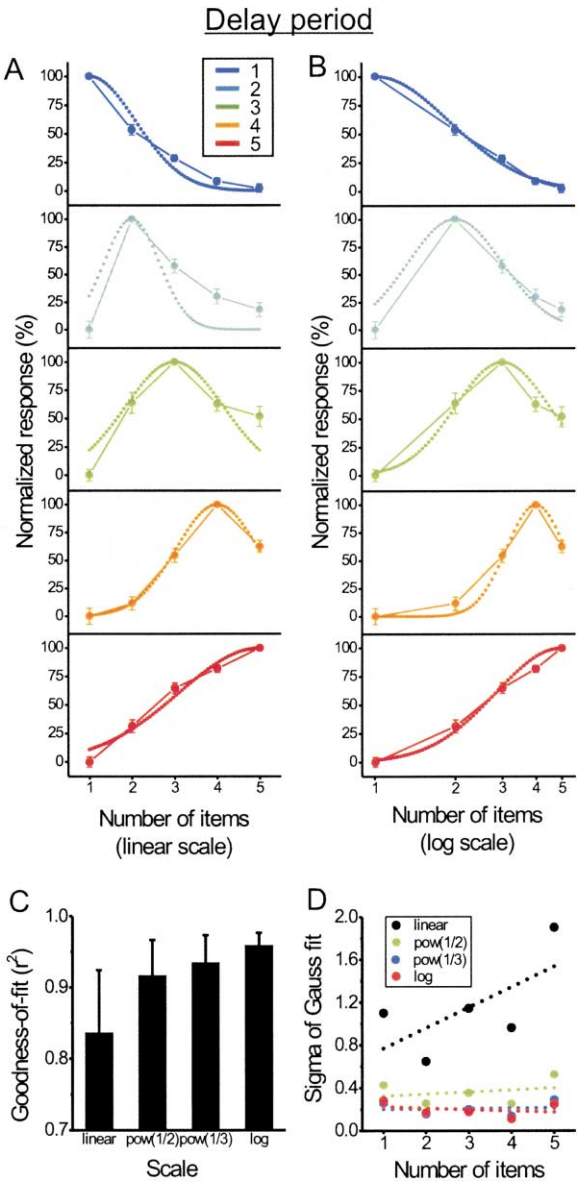

Figure 6. Neural Representation of Numerosities in the PFC during the Delay Period

(A) Again, the neural filter functions are asymmetric on a linear scale.

(B) Logarithmic transformation of the filter functions results in more symmetric distributions.

(C) Goodness-of-fit for the four different scaling schemes.

(D) Standard deviation values for the scaling schemes across preferred quantities (error bars, ±SEM).

that the Weber fractions remain constant over different numerosities. The traditional way to calculate the Weber fraction (Equation 1 in the Experimental Procedures) assumes equal numerical distance of discrimination thresholds for numbers smaller and larger than the sample number. As the above analyses indicate, behavioral and neural distributions are asymmetric on a linear scale; thus, the numerical distances between the sample number *n* and the threshold numbers above and below *n* are unequal (i.e., longer distance toward higher numbers). Instead, we can calculate the Weber fraction separately for numerosities smaller and larger than the sample number *n* (Equations 3 and 4; see Experimental Procedures) (Van Oeffelen and Vos, 1982). According to this formulation, the Weber fraction for the upper and lower part of a distribution should be equal and constant if a logarithmic scale is the best scale. The "asymmetry index" (Equation 5; see Experimental Procedures) additionally tests if the proportions of the lower and upper part of the distributions obey a logarithmic relationship. If they do, the index value should be 1. (The linear and

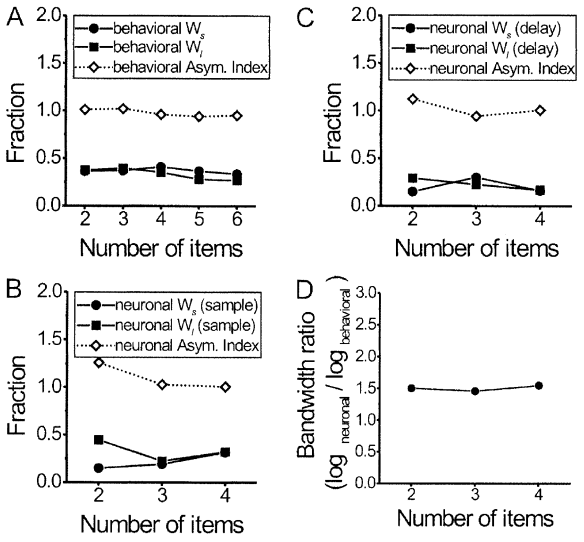

Figure 7. Comparison of Behavioral and Neuronal Data

Weber fractions and asymmetry indices for the behavioral performance (A) and the neuronal data during sample presentation (B) and delay period (C). (D) The bandwidths between the behavioral and the neural filter functions changed proportionally and exhibited a fixed ratio of 1.5.

the log models can give comparable Weber fraction and asymmetry index values if filter bandwidths are extremely sharp, but our measured filter bandwidths are relatively broad).

Figure 7 shows the values of the Weber fractions calculated from numbers smaller than the sample numerosity (W_s) and from numbers larger than it (W_l), as well as the asymmetry index for both the behavioral data and neural data. Consistent with logarithmic coding, the values for W_s (mean 0.37 $\pm$ 0.03 STD) and W_l (mean 0.33 $\pm$ 0.06 STD) for the behavioral data (Figure 7A) are equal and constant across numerosities. Also consistent are the asymmetry indices, which are close to 1 for all tested numerosities (mean 0.97 $\pm$ 0.04 STD).

Similar results were obtained for the neural data, for both the sample (Figure 7B) and delay period (Figure 7C). The neuronal Weber fractions W_s and W_l are equal and constant for the numerosities 2 through 4 (fractions for preferred numerosities "one" and "five" are not included because they are at the end of the tested range and thus some values cannot be derived as the distributions are incomplete). During the sample and delay period, W_s had values of 0.22 ($\pm$0.09) and 0.20 ($\pm$0.08), respectively. Similarly, W_l was 0.33 ($\pm$0.11) for the sample interval and 0.24 ($\pm$0.09) in the delay period. The asymmetry indices in the sample (mean 1.10 $\pm$ 0.14 STD) and delay epoch (mean 1.02 $\pm$ 0.09 STD) were also close to 1, as predicted by the logarithmic-coding hypothesis. Thus, in sum, all of our analyses support that notion of a nonlinearly compressed coding scheme for numerosity information.

Comparison of Behavioral and Neural Data

We also compared behavioral and neural selectivity for different numerosities. The behavioral and neuronal population distributions for each numerosity were plotted on a logarithmic scale and fitted with Gauss functions, and the standard deviation (sigma) of each distribution was estimated from this fit. We then computed a neural-behavioral bandwidth ratio for each numerosity by dividing the neural sigma by the behavioral sigma. Larger sigma means a wider distribution (and therefore less selectivity), so bandwidth ratios greater than one indicate that the neural population was less sensitive than behavior. This revealed that the bandwidth ratios were a constant 1.5 across all numerosities (Figure 7D), indicating greater sensitivity on the behavioral than the neural level. But also note that this ratio remained remarkably constant across different numerosities (slope = 0.02, linear fit) even though at both the behavioral and neural level, sensitivity decreases with increasing numerosity. This suggests a direct relationship between behavioral and neuronal representations.

Discussion

In this study, we report that behavioral and neural measures of visual numerosities are better described by a nonlinear than a linear scale. After nonlinear compression using power functions or a logarithmic scale, the psychophysical and neuronal data showed symmetric distributions and constant variability, at least within the tested numerosity range. Both of these findings are predicted by a nonlinear coding model of numerosity (Van Oeffelen and Vos, 1982; Dehaene and Mehler, 1992; Dehaene and Changeux, 1993; Dehaene, 2001), and thus, these results support that model over a linear coding model.

To evaluate the symmetry of the behavioral and neural filter functions in the monkeys, we fitted a normal distribution (Gaussian) to the measured data. The Gaussian was chosen because it represents the standard symmetric distribution. As indicated by the very high goodness-of-fit values, which were close to the ideal value of 1 for the best scaling schemes, the normal distribution was well suited as a model function. Whether numerical filter functions are best modeled by Gaussians or some other related peak functions, however, is currently unknown. Functions with higher center numerosities and more sampling points can be derived to resolve this question. The differences in goodness-of-fit for the four different scaling schemes were highly significant and consistent across tests, but small. This is not surprising, because large differences could not be expected. Even a slightly skewed distribution resulting from a normal distribution plotted on the "wrong" scale (see predictions in Figure 1) would yield a relatively good goodness-of-fit value. But despite relatively high values, all nonlinear scaling protocols produced systematically higher goodness-of-fit values. Thus, nonlinearly compressed scaling was significantly superior compared to linear scaling.

Both the behavioral and neural data obeyed Weber's law (increasing the numerical distance between numerosities by a given fraction improves discriminability). This clear Weber fraction signature argues for explicit numerical representations in monkeys. The data are not compatible with nonnumerical "object-file" representations (e.g., Feigenson et al. 2002), which would show a strict

limit of quantities that can be discriminated ("set size signature" of object-file representations). More generally, these findings are consistent with the Weber-Fechner law (equal increments in sensory experience are proportional to the logarithm of stimulus magnitude). This suggests that nonverbal cognitive representations (e.g., numerosity) obey the same fundamental laws and neural coding schemes as the purely sensory data from which the Weber-Fechner law was derived. The compression of magnitude in the sensory domain is generally caused by transduction processes at the sensory epithelium. But in the current study, we are dealing with abstract numerical categories, information that generalizes across the exact physical appearance of displays (Nieder et al., 2002). Thus, a simple transduction effect of receptors is unlikely to account for the observed compression. The monkeys' behavioral precision in number discrimination was found to be superior by a factor of 1.5 compared to their neuronal filters' selectivity. This is also reminiscent of sensory physiology, where psychophysical thresholds are typically more sensitive than average neuronal thresholds (Werner and Mountcastle, 1965; Kiang et al., 1965; DeValois et al., 1967; Talbot et al., 1968). This may be explained by the "lower envelope principle," which argues that behavioral detection is supported by the most sensitive individual neurons of a population (Parker and Newsome, 1998).

As a behavioral consequence of these findings, a human or animal subject should be able to discriminate smaller quantities more reliably than higher quantities. To reach equal discriminability for numerosities smaller and larger than the sample numerosity, the numerical distance between sample and test must be larger if the test numerosity is higher than the sample. This effect should become even more obvious for larger magnitudes. Any theoretical approach that attempts to model nonverbal numerical representations in a behavioral/physiological way will have to take a nonlinearly compressed scaling into account (Dehaene and Changeux, 1993).

What might be the advantage of a compressed numerical representation? In psychophysics and sensory physiology, the idea is that compression enlarges the coding space, thus increasing the dynamic range of perception and firing neurons (Dayan and Abbott, 2001). The same could be true for the higher-level representations that mediate numerical abilities. Because we explored representations in an animal without language, our data can only provide evidence for nonverbal representations. But nonverbal numerical abilities are present in humans, both during ontogenesis (e.g., Spelke 2000) as well as in human adults when they are prevented from counting verbally (e.g., Cordes et al., 2001). Further, adult humans, when required to indicate the larger of two Arabic numbers, show a numerical distance effect similar to that seen in animals (Moyer and Landauer 1967). Thus, even symbolic verbal representations share characteristics with the nonverbal representations investigated here. Whether verbal-based representations also follow Weber-Fechner laws will require further testing.

Comparisons of activity between the sample and delay epochs revealed that nonlinear coding was evident during the acquisition phase of the task (the sample epoch) and that there was no further compression of the data in the subsequent memory delay. This is similar to observations from the somatosensory system where it has been proposed that once the neural representation of somatosensory intensity is established during acquisition, information is transmitted across subsequent processing stages with no further compression (Werner and Mountcastle, 1965; Johnson, 2000; Johnson et al., 2002).

Based on these results, we suggest a simple processing scheme. The only nonlinear stage consists of the initial encoding of numerical information, but as a consequence, the maintenance of numerical information (delay period) as well as the behavioral output are also nonlinearly compressed. Encoding of numerical information results in distributions with some variation. This finding is in contradiction to the "accumulator model" for nonverbal number estimation (Gallistel and Gelman, 2000), which assumes that memory is the main source of imprecision in numerosity judgments. Our results, however, are in good agreement with the Dehaene and Changeux (1993) model, which posits that imprecision with increasing numerosity arises from nonlinear compression during the acquisition of numerical information.

In sum, our data provide evidence that some cognitive representations exhibit attributes similar to those found for perceptual/sensory processes, which supports the analog coding hypothesis. Studies of mental imagery also strongly support the idea of analog coding of cognitive information (Shepard and Metzler, 1971; Shepard and Podgorny, 1978; Kosslyn 1994). Other support comes from neuroimaging studies showing that mental representations activate the same neural infrastructure used in perception (Kosslyn et al., 1995, 2001). Thus, perception and cognition may not reflect independent systems in the brain, after all (Barsalou, 1999). From an evolutionary point of view, it might have been more economical to build on existing principles of lower-level sensory information to develop representations of information that we think of as "cognitive."

Experimental Procedures

Behavioral Protocol

Monkeys grasped a lever and fixated a central fixation target to start a trial. A sample display (800 ms) was followed by a memory delay (1000 ms). Next, a test display appeared, which was either a match (it contained the same number of dots as the sample display) or a nonmatch (it contained—with equal probability—more or fewer items). The lowest sample numerosity was "two," the highest "six." For a given sample stimulus, the number of items in nonmatch displays ranged from 1 to sample-numerosity + (sample-numerosity − 1). If the display was a match, monkeys released the lever to receive a juice reward before it disappeared. If the display was a nonmatch, the monkeys held the lever through a second, brief delay until a match appeared that required a lever release for a reward. Trials were randomized and balanced across all relevant features. Monkeys were required to keep their gaze within 1.25 degree of the fixation point during sample presentation and the memory delay (monitored with an infrared eye tracking system, ISCAN).

Stimuli

Black items (diameter range: 0.8°–1.3° of visual angle) were displayed on a gray background (diameter: 8° of visual angle). The exact physical appearance of the displays was varied by randomly placing dots in 24 possible locations on a 5 × 5 matrix centered around the fixation target on which monkeys maintained gaze. Each

dot was also randomly varied between five different sizes. To prevent the monkeys from memorizing the visual patterns of the displays, each quantity was tested with 100 different images per session, and the sample and test displays that appeared on each trial were never identical. All quantities of items were used in each session, and all displays were newly generated for each session by pseudorandomly shuffling all relevant item features. Several control stimuli were applied prior to these experiments to ensure that the monkeys solved the task by truly abstracting quantity, rather than attending to low-level visual features. Controls included displays in which the total area and circumference was equated across different quantities, displays of low and high dot density, displays in which two or more dots formed lines ("linear") and formed polygons ("shapes"), and displays in which the dots were replaced with triangles, squares, and ovals ("variable features"). The behavioral performance of the monkeys to these controls suggested that monkeys were indeed judging quantity (Nieder et al., 2002).

Neurophysiological Recording

The same behavioral protocol as described above was used for recordings. Only the range of nonmatch numerosities was smaller. Nonmatch displays contained—with equal probability—only one more or one less item (except for "one" and "five"). Recordings were made from three hemispheres of the ventral bank of the dorsolateral PFC of the same two adult rhesus monkeys (*Macaca mulatta*) that provided the behavioral data. All procedures were in accordance with the NIH and MIT guidelines for animal experimentation. Arrays of eight to twelve tungsten microelectrodes (FHC) were inserted using a grid (Crist Instruments) with 1 mm spacing. Electrodes were localized using magnetic resonance imaging. Neurons were selected at random. Waveform separation was performed off-line applying principal component analysis (Plexon Systems). Each monkey performed between 500 and 800 correct trials per recording session.

Data Analysis

Sample activity was derived from an 800 ms interval starting 100 ms after stimulus onset. Delay activity was summed over an 800 ms interval starting 200 ms after sample offset. Visual quantities that elicited the largest spike rate were defined as the "preferred numerosity" of a given cell (ANOVA, $p < 0.01$). To create neural filter functions, activity rates were normalized by setting the maximum activity to the most preferred quantity as 100% and the activity to the least preferred quantity as 0%. Filter functions for each applied numerosity were constructed by averaging the normalized spike rates for preferred numerosity "one," "two," and so on for the sample and delay period. Gauss functions were fit to the performance functions of both monkeys separately and the neural filter functions of both the sample and delay period (χ^2 minimization after Levenberg-Marquardt), and the goodness of fit (r^2) was derived as a quantitative measure. The Gaussian was chosen because it represents a standard symmetric distribution; the precise nature of behavioral or neuronal distributions still has to be clarified.

The Weber threshold is defined as the 50% correct discrimination between two stimuli. The Weber fraction for visual numerosity is (Van Oeffelen and Vos, 1982)

$$W_b = \frac{(max - min)}{min}, \quad (1)$$

where *max* and *min* are the upper and lower cutoff values of the performance/neural-filter functions. However, if numerosities are represented on a compressed (nonlinear) numerical scale, the threshold of the smaller number n_s is numerically less remote from the sample number n than the threshold of the larger number n_l. Thus, the Weber fraction for numbers smaller than n is

$$W_b = \frac{(n - n_s)}{n_s}. \quad (2)$$

The Weber fraction for numbers larger than n is

$$W_b = \frac{(n_l - n)}{n}. \quad (3)$$

As the Weber fraction is a constant, the following relation must hold:

$$\frac{(n - n_s)}{n_s} = \frac{(n_l - n)}{n}, \quad (4)$$

Because of the asymmetric relation of the discrimination functions, the following special relation arises, which will be called 'asymmetry index':

$$\frac{n_s}{n} \cdot \frac{n_l}{n} = 1 \quad (5)$$

These predictions for asymmetric representation of numerical information are tested in the current paper. Note that the absolute values of the Weber thresholds are not directly comparable for the neuronal and behavioral condition fractions because they were derived for neuronal data at 75% of the maximum response due to incompletely closed distributions at the 50% level.

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 a NIMH grant (1-R01-MH65252-01) and the RIKEN-MIT Neuroscience Research Center. The authors wish to thank Marlene Wicherski for valuable comments on an earlier version of the manuscript, and three anonymous reviewers for helpful suggestions.

Received: August 22, 2002
Revised: November 27, 2002

References

Barsalou, L.W. (1999). Perceptual symbol systems. Behav. Brain Sci. *22*, 577–660.

Brannon, E.M., Wusthoff, C.J., Gallistel, C.R., and Gibbon, J. (2001). Numerical subtraction in the pigeon: evidence for a linear subjective number scale. Psychol. Sci. *12*, 238–243.

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

Dayan, P., and Abbott, L.F. (2001). Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems (Cambridge, MA: MIT Press).

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

Dehaene, S. (2001). Subtracting pigeons: logarithmic or linear? Psychol. Sci. *12*, 244–246.

Dehaene, S., and Changeux, J.P. (1993). Development of elementary numerical abilities: a neural model. J. Cogn. Neurosci. *5*, 390–407.

Dehaene, S., and Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. Cognition *43*, 1–29.

Dehaene, S., Dehaene-Lambertz, G., and Cohen, L. (1998). Abstract representation of numbers in the animal and human brain. Trends Neurosci. *21*, 355–361.

DeValois, R.L., Abramov, I., and Mead, W.R. (1967). Single cell analysis of wavelength discrimination at the lateral geniculate nucleus in the macaque. J. Neurophysiol. *30*, 415–433.

Fechner, G.T. (1860). Elemente der Psychophysik, Volume 2 (Leipzig: Breitkopf & Härtel).

Feigenson, L., Carey, S., and Hauser, M. (2002). The representations underlying infants' choice of more: object files versus analog magnitudes. Psychol. Sci. *13*, 150–156.

Fodor, J. (1975). The Language of Thought (New York: Crowell).

Gallistel, C.R., and Gelman, R. (2000). Non-verbal numerical cognition: from reals to integers. Trends Cogn. Sci. *4*, 59–65.

Gibbon, J. (1977). Scalar expectancy theory and Weber's Law in animal timing. Psychol. Rev. *84*, 279–335.

Gibbon, J., and Church, R.M. (1981). Time left: linear versus logarithmic subjective time. J. Exp. Anim. Behav. Process. *7*, 87–107.

Johnson, K.O. (2000). Neural coding. Neuron *26*, 563–566.

Johnson, K.O., Hsiao, S.S., and Yoshioka, T. (2002). Neural coding and the basic law of psychophysics. Neuroscientist *8*, 111–121.

Kiang, N.Y.-S., Watanabe, T., Thomas, E.C., and Clarke, L.F. (1965). Discharge Patterns of Single Fibers in the Cat's Auditory Nerve (Cambridge, MA: MIT Press).

Kosslyn, S.M. (1994). Image and Brain: The Resolution of the Imagery Debate (Cambridge, MA: MIT Press).

Kosslyn, S.M., Thompson, W.L., Kim, I.J., and Alpert, N.M. (1995). Topographical representations of mental images in primary visual cortex. Nature *378*, 496–498.

Kosslyn, S.M., Ganis, G., and Thompson, W.L. (2001). Neural foundations of imagery. Nat. Rev. Neurosci. *2*, 635–642.

MacKay, D.M. (1963). Psychophysics of perceived intensity: a theoretical basis for Fechner's and Stevens' laws. Science *139*, 1213–1216.

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

Nieder, A., Freedman, D.J., and Miller, E.K. (2002). Representation of the quantity of visual items in the primate prefrontal cortex. Science *297*, 1708–1711.

Parker, A.J., and Newsome, W.T. (1998). Sense and the single neuron: probing the physiology of perception. Annu. Rev. Neurosci. *21*, 227–277.

Pylyshyn, Z. (1984). Computation and Cognition: Toward a Foundation for Cognitive Science (Cambridge, MA: MIT Press).

Shepard, R.N., and Metzler, J. (1971). Mental rotation of three-dimensional objects. Science *171*, 701–703.

Shepard, R.N., and Podgorny, P. (1978). Cognitive processes that resemble perceptual processes. In Handbook of Learning and Cognitive Processes, W.K. Estes, ed. (Hillsdale, NJ: Erlbaum), pp. 189–237.

Spelke, E.S. (2000). Core knowledge. Am. Psychol. *55*, 1233–1243.

Stevens, S.S. (1961). To honor Fechner and repeal his law. Science *133*, 80–86.

Talbot, W.H., Darian-Smith, I., Kornhuber, H.H., and Mountcastle, V.B. (1968). The sense of flutter-vibration: comparison of the human capacity with response patterns of mechanoreceptive afferents from the monkey hand. J. Neurophysiol. *31*, 301–334.

Van Oeffelen, M.P., and Vos, P.G. (1982). A probabilistic model for the discrimination of visual number. Percept. Psychophys. *32*, 163–170.

Weber, E.H. (1850). Der Tastsinn und das Gemeingefühl. In Handwörterbuch der Physiologie, Vol. 3, Part 2, R. Wagner, ed. (Braunschweig, Germany: Vieweg), pp. 481–588.

Werner, G., and Mountcastle, V.B. (1965). Neural activity in mechanoreceptive cutaneous afferents: stimulus-response relations, Weber functions, and information transmission. J. Neurophysiol. *28*, 359–397.