Symbol addition by monkeys provides evidence for normalized quantity coding

Margaret S. Livingstone[a,1], Warren W. Pettine[a,2], Krishna Srihasam[a], Brandon Moore[a,3], Istvan A. Morocz[b], and Daeyeol Lee[c]

[a]Department of Neurobiology, Harvard Medical School, Boston, MA 02115; [b]Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02116; and [c]Department of Neurobiology, Yale University School of Medicine, New Haven, CT 06510

Edited* by Jon H. Kaas, Vanderbilt University, Nashville, TN, and approved March 28, 2014 (received for review March 4, 2014)

Weber's law can be explained either by a compressive scaling of sensory response with stimulus magnitude or by a proportional scaling of response variability. These two mechanisms can be distinguished by asking how quantities are added or subtracted. We trained Rhesus monkeys to associate 26 distinct symbols with 0–25 drops of reward, and then tested how they combine, or add, symbolically represented reward magnitude. We found that they could combine symbolically represented magnitudes, and they transferred this ability to a novel symbol set, indicating that they were performing a calculation, not just memorizing the value of each combination. The way they combined pairs of symbols indicated neither a linear nor a compressed scale, but rather a dynamically shifting, relative scaling.

macaque | normalization | number sense | value coding

Animals and humans can estimate the number of various items, and the precision of this approximate number sense decreases with magnitude. For example, although it is easy to recognize the difference between 2 and 4 items, it is more difficult to distinguish 22 from 24 items. This dependence of accuracy on magnitude is a property that the approximate number sense shares with more basic sensory processes. Weber (1) observed that in general, across many sensory modalities, the just noticeable difference between two stimuli is proportional to their magnitude. Fechner (2) proposed that Weber's observation could be explained if sensations were physiologically encoded as a logarithmic function of stimulus magnitude, but Stevens (3) argued instead that sensations obey a power law, with perceptual magnitude being proportional to a power function of the stimulus magnitude, with the power usually less than 1. Both a logarithmic and a power-less-than-one relationship between stimulus and internal coding are compressive, with the same physical difference between stimuli producing incrementally smaller internal differences between successively larger pairs of external stimuli. Any kind of compressive scaling would explain a decrease in discriminability with increasing magnitude if the noise in the internal representation is constant.

However, an alternative possibility is that variability in encoding might increase with stimulus magnitude. In fact, the variability in the firing rates of cortical neurons tends to increase with firing rate (4–6). Therefore, to the extent that a stimulus parameter is encoded by the rate of neural firing, an increase in perceptual variability with stimulus magnitude may not require compressive scaling; it is also consistent with a linear neuronal representation with magnitude-dependent variability (7–10).

Neurons that are tuned to numerosity have been recorded in monkey posterior parietal and lateral prefrontal cortex (11–13). The width and asymmetry of such tuning is consistent with a compressed scaling (14). However, neurons tuned to particular numerosities, or numerosity ranges, represent a labeled-line code and therefore are not, themselves, scaled to numerosity, in the sense that either Fechner, or Stevens, meant when they proposed a logarithmic, or power, scale for the sensory response to a graded physical stimulus. What is explicit in Fechner's model, and implicit in models of tuned units, is a stage where sensory response increases with stimulus magnitude. Indeed, neurons whose firing rate depends monotonically on the number of items in an array have been reported in macaque lateral intraparietal area (LIP) (15), but the results do not distinguish between a linear or a logarithmic coding.

Stevens asserted that the only behavioral test that can distinguish linear from logarithmic sensory coding is how sensory magnitudes are added or subtracted (16). Specifically, he pointed out that addition and subtraction can be performed accurately only on linear representations, whereas a compressive representation allows only multiplication and division. If magnitudes are combined at a stage of labeled line coding, as proposed by Dehaene (17), the way magnitudes are combined would not necessarily reflect scaling. However, if magnitudes are combined at a stage where they actually are coded according to a linear or logarithmic scale, then the scale can be distinguished by examining how magnitudes are added or subtracted. If the underlying scale is logarithmic, or otherwise compressed, the combination of two magnitudes should be superadditive (expansive). For example, in a compressed scale, the internal representation of "3" will be more than half the internal representation of "6," so combining two 3s should correspond to more than 6.

Studies in which rats, mice, or pigeons must estimate the time remaining (10, 18) or the number of pecks remaining (9) find that these animals show linear subtraction behavior, consistent with a linear internal scale. However, the concern has been raised that these animals simply learned the correct response for every possible condition (17). Here we ask how monkeys combine pairs of symbols or pairs of dot arrays representing a large range of quantities, from 0 to 25, a range large enough that memorization of all possible pairwise combinations should be prohibitively difficult. Cantlon and Brannon (19) previously showed that monkeys can sum sequentially presented dot arrays, and one chimpanzee has demonstrated the ability to sum Arabic numerals up to a total of 4 (20), but the nature of the internal representation was not explored in either of these studies.

Significance

Symbol-literate monkeys can be trained to combine, or add, pairs of large numbers. They transfer to a novel symbol set, ruling out memorization of each symbol pair. Their addition behavior indicates an underlying relative scaling of magnitude.

Author contributions: M.S.L. and W.W.P. designed research; M.S.L. performed research; I.A.M. contributed new reagents/analytic tools; M.S.L., W.W.P., K.S., B.M., and D.L. analyzed data; and M.S.L. and D.L. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: margaret_livingstone@hms.harvard.edu.

[2]Present address: Colorado University School of Medicine, Aurora, CO 80045.

[3]Present address: Vanderbilt Brain Institute, Nashville, TN 37235.

Results and Discussion

To determine how monkeys sum quantities, we taught symbol-literate monkeys (21) an addition task using dots and two distinct symbol sets presented on a touchscreen in their home cage. Three young adult male macaque monkeys had been trained extensively, as juveniles, using a pairwise choice task to associate 26 distinct symbols or up to 25 dots in an array with reward values of 0–25 drops of liquid (Fig. 1, *Upper*). In symbol set 1 Arabic numerals 0–9 represented 0–9 drops, and the letters X Y W C H U T F K L N R M E A J represented 10–25 drops. Symbol set 2, which was learned after the monkeys had mastered addition using symbol set 1, was made by filling 4–5 squares in a 3 × 3 square array: 0–25 drops were represented by the symbols: . The first three plots in Fig. 1 (*Lower*) show the monkeys' choice behavior in each pairwise comparison task. Although the monkeys were rewarded appropriately no matter which side they chose, after training, in all three pairwise comparison tasks, they almost invariably chose the larger option; they were highly accurate at discriminating the stimuli in the pairwise task, especially with symbol set 1, with which they had had several years of experience (21, 22), and were less accurate for symbol set 2, with which they had had less experience.

To investigate how the monkeys combined values, we gave them an addition task, first using dots, then the well-learned symbol set 1. For the dots addition task, the monkeys were presented with two vertically separated dot arrays each inside a circle on one side of the screen, comprising the "sum," and a single dot array on the other side of the screen, the "singleton." Whichever side of the screen he touched, the monkey was rewarded with the number of drops of liquid corresponding to the value (sum or singleton) on that side. Although we made it as clear as possible, using discrete drops accompanied by discrete beeps, that each symbol represented a distinct number of drops, we cannot assume that the monkeys interpreted these symbols as representing numerosity, rather than quantity or hedonic value.

The last three plots in Fig. 1 (*Lower*) show the fraction of times the monkeys chose the sum for each singleton–sum combination for the different tasks. The choice data show that the monkeys usually picked the larger of the two, sum or singleton, except when the sum and the singleton were close in value, when their behavior approached chance. Fig. 2*A* shows that the average percent-correct (larger) choices averaged over all three monkeys were well above chance (50%) for each day of the dots addition task and the two symbol addition tasks.

Our question is not, however, whether the monkeys can perform above chance on a difficult addition task, but how the monkeys combine quantities. To answer this, we first calculated the singleton-equivalent value of each sum magnitude (averaged over all addend combinations) by fitting a logistic function to the choice ratios using maximum likelihood, as shown in Fig. 2*D* for the data from the last month of symbol set 1 addition. The point of subjective equality between each sum and all singletons with which it was paired was taken as the singleton-equivalent value of each sum. Fig. 2*E* plots the singleton-equivalent values for each sum magnitude for the last 30 d of the dots addition task (black), the first 10 d (red) and the last 30 d (blue) of symbol set 1 addition, and the first 10 d of symbol set 2 addition (green). After learning the dots addition task the monkeys valued two sets of dots presented together as equivalent to the numerical sum of the two (Fig. 2*E*, black). When first presented with the symbol set 1 addition task the monkeys on average undervalued the sum of two symbols, compared with the singleton (Fig. 2*E*, red). Indeed, their behavior is roughly equivalent to the choice ratio if they just chose the largest value symbol on the screen, ignoring the smaller value symbol entirely (dotted black line).

We then calculated the contribution of the larger and the smaller addends to the subjective value of the sum separately (Fig. 2 *F* and *G*). In the dots addition task, both addends contributed significantly to the value of the sum ($P < 10^{-10}$), although the contribution of the smaller addend was less than that of the larger addend (larger addend weight = 1.03; smaller addend weight = 0.64). Thus, the monkeys combined the magnitudes of the two dot arrays on the sum side, but undervalued the smaller addends relative to the larger ones. Although this shows that they combined the two addend dot arrays to arrive at an approximately correct sum, we cannot tell whether they first evaluated the magnitude of each addend array and then added them or whether they simply evaluated how many dots in total were on the sum side. On the other hand, they cannot evaluate the sum magnitude directly for the symbols.

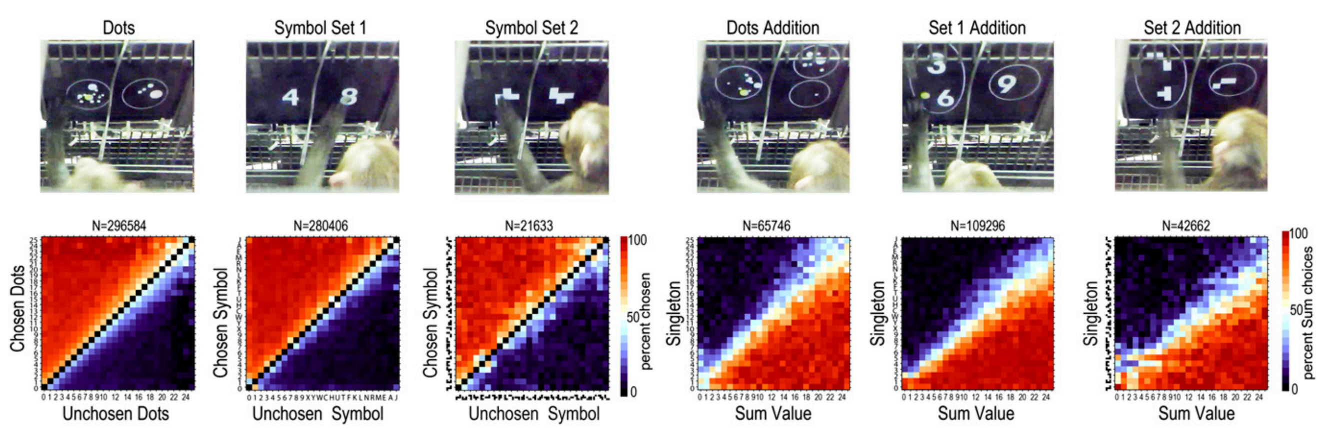


Fig. 1. Tasks. (*Upper*) A monkey performing each of the six tasks. Dots comparison task: The monkey has chosen 11 dots rather than 4 dots and is receiving 11 drops of reward through a stainless steel tube. Symbol set 1 comparison task: The monkey is about to touch the symbol 8 rather than 4. Symbol set 2 comparison task: The monkey is touching the symbol (worth 21 drops) rather than the symbol (worth 3 drops). Dots addition task: The monkey is choosing 8 dots rather than 6 plus 1 dots. Addition with symbol set 1: The monkey has chosen 3 plus 6 instead of 9 (the two choices give equivalent rewards). Addition with symbol set 2: The monkey is about to touch plus (worth 9 +1 3 = 22 drops) rather than (worth 19 drops). (*Lower*) Average choice matrices for each task over a 2-mo period; number of trials indicated above each plot. For the comparison tasks the plot shows the average choice ratio for every possible choice pair. The horizontal and vertical position of each square in the matrix indicates the two choices that were presented, and the color of each square in the matrix indicates the percentage of trials when the monkey chose the vertical option over the horizontal. The choice matrices for the three addition tasks show the average behavior of the same three monkeys over the last 1-mo period on each task for every possible sum and singleton combination. The vertical position of each square represents the value of the singleton option and the horizontal position represents the sum value; the color of each square indicates the percentage of trials when the monkey chose the sum over the singleton.

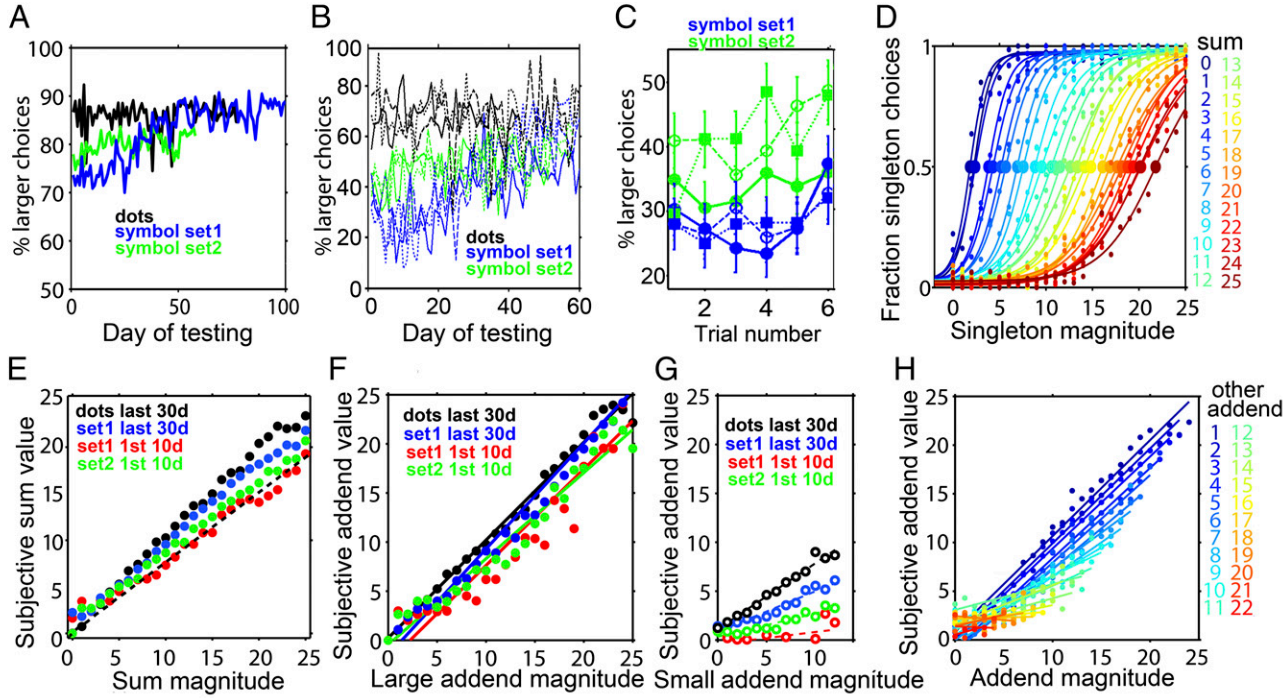


Fig. 2. Behavioral results. (*A*) Percent-correct (larger) choices each day for each task averaged over all three monkeys. (*B*) Percent-correct (larger) choices each day for each task for each monkey (indicated by different line types) for all mandatory calculation combinations (when the sum is larger but neither addend is larger than the singleton). (*C*) Percent-correct (larger) choices ± SEM for the first six trials of each addend combination for the two symbol addition tasks for each monkey for all mandatory calculation combinations. (*D*) Fit of a logistic function to the fraction of singleton choices as a function of singleton magnitude, for each sum value for symbol set 1 addition task, last 30 d; the 50% choice point is taken as the singleton-equivalent value of the sum. (*E*) Singleton-equivalent value (calculated as in *D*) of each sum magnitude for each task. Dotted black line indicates predicted singleton-equivalent sum value if the monkeys simply chose the largest item on the screen. (*F*) Singleton-equivalent value of only the larger of the two addends for each task. (*G*) Singleton-equivalent value of only the smaller of the two addends for each task. (*H*) Singleton-equivalent value of each addend calculated separately for every other addend magnitude with which it was paired, for symbol set 1 addition task, last month of data.

For the first 10 d of symbol set 1 addition (Fig. 2*G*, red), the singleton-equivalent value of the smaller addend was close to zero. The logistic regression model yielded a small-addend weight of 0.10, which was smaller than the large addend weight (0.89) (Fig. 2*F*) but still significantly different from zero (P = 0.003), indicating that when the monkeys were first presented with the addition task, they mostly chose the largest element on the screen, and valued the smaller addend at only 1/10th of its actual value. This is not surprising, because they had been extensively trained on a paired comparison task with the same symbols, in which the optimal strategy was to choose the larger of the two presented options. After 4 mo of training on the symbol set 1 addition task, however, their valuation of the sum of two symbols (Fig. 2*E*, blue) increased, although not quite to the full value of the sum of the two symbol magnitudes.

Thus, the monkeys learned that two symbols on one side of the screen together represent a larger reward than the previously learned value of either symbol alone. They could be performing a calculation; i.e., evaluating the sum as a combination of both addend values, or they could be performing a simpler operation, like valuing two symbols together as "somewhat larger" than the value of the larger symbol alone, or they could have learned the value of each 2-symbol combination (351 position-specific combinations). To decide if, and, if so, how the monkeys were combining the two addends, we calculated the singleton-equivalent value of the smaller and larger addends separately (Fig. 2 *F* and *G*, blue circles). Their valuation of the smaller addend increased after 4 mo of training, such that for the fifth month of symbol set 1 addition the logistic regression model gave the weight of 0.34 (1.01) for the smaller (larger) addend, both of which were significantly larger than zero.

This increase in valuation of the smaller addends supports our conclusion that after 4 mo of daily exposure to symbol set 1 addition, the monkeys learned that the value represented on the sum side was larger than the value represented by either of the two addend symbols alone. Their behavior at the end of the 5-mo period could no longer be explained by a choose-the-largest-symbol strategy (because both the larger and the smaller addend contributed significantly to the behavioral value of the sum), or by any strategy based only on the value of the larger addend, such as simply incrementing it by a fixed amount (because their valuation of the sum depended significantly on the magnitudes of both addends). Furthermore, as with the dots addition task, although their performance approximated addition, the monkeys systematically undervalued the smaller of the two addends. Note that symbols 1–12 can be either the larger or the smaller addend, and that the subjective value of these symbols differed strikingly, by a factor of 3, depending on whether they were presented as the larger or the smaller of the two addends (compare Fig. 2 *F* and *G*).

It is still possible that the monkeys' final improved performance on symbol set 1 addition was achieved by learning the value of every possible addend–addend combination (351 different position-specific combinations), choosing on the basis of memorized value, rather than performing a calculation. To distinguish between memorization, however unlikely, and calculation, we asked whether the monkeys would perform addition with a second symbol set, reasoning that if they performed addition on a second symbol set without extensive training, then they could not be relying on memorized values of each addend–addend combination, but rather had learned to combine the two addends—a calculation. The monkeys learned symbol set 2 using the original two-symbol comparison task for 3 1/2 mo, by the end of which they chose the larger symbol 87% of the time, over all possible symbol pair combinations (Fig. 1). The monkeys then alternated for 1 mo between symbol set 1 addition and the two-symbol comparison task with symbol set 2. Finally, they were presented with the addition task using symbol set 2.

From the first day of testing with symbol set 2 addition, the monkeys chose the larger side more often than they did during the early days of symbol set 1 addition (Fig. 2*A*), and their performance reached a stable asymptote within 10 d, rather than the 50 d it took to reach asymptote for symbol set 1 addition. Their final accuracy in the symbol set 2 addition task was lower than that for symbol set 1 addition, presumably because they had had much less experience (in the comparison task) with this symbol set. Nevertheless, the small addend valuation plot during the first 10 d on symbol set 2 addition (Fig. 2*G*, green) shows that the monkeys valued the smaller addends at 20% of their actual value. Similarly, the smaller addend weight estimated from the logistic regression model during the first 10 d of symbol set 2 addition was 0.2, and was significantly larger than zero ($P < 10^{-18}$). This indicates that the monkeys transferred the addition task to a novel symbol set, even though they experienced each of the 351 possible position-dependent addend combinations on average less than twice per day.

As a further test of whether the monkeys transferred the ability to combine symbols, we define "mandatory calculation" as those trials in which the sum is bigger than the singleton, but the singleton is larger than both addends (i.e., "choose the largest" strategy always gives the incorrect answer). The daily percent-correct (larger) choices for mandatory calculation trials for the first 10 d of symbol set 2 addition were significantly higher than the first 10 d with symbol set 1 addition (χ^2-test, $P < 10^{-14}$); this was true for each monkey individually (Fig. 2*B*) whether calculated for the first 10 d (χ^2-test, $P < 10^{-7}$) or the first 200 or 500 mandatory trials (χ^2-test, $P < 10^{-7}$). To still further ascertain whether this behavior represents true transfer, we looked at their behavior as a function of trial number for each of the possible 132 different addend–addend combinations for all mandatory calculation conditions for the two symbol sets. Fig. 2*C* confirms the transfer of addition behavior, in that the average percent correct over all possible addend–addend combinations was larger for symbol set 2 than for symbol set 1 for the first through the sixth time each addend–addend combination was presented for each monkey individually.

We conclude from the results so far that the monkeys learned to combine pairs of symbols in such a way that the combination was valued at more than the magnitude of either individual addend, but less than the numerical sum of the two symbols. Because the monkeys transferred the task to a novel symbol set, we conclude that they did not simply memorize the value of every possible pair of symbols, but rather performed some kind of calculation. Because both the large and the small addends contributed significantly to the singleton-equivalent value of the sum, we conclude that the calculation was not simply "choose the largest," or "value the larger addend at some fixed increment or fraction of its magnitude," but rather to combine the two addends. Therefore, we sought to characterize the nature of internal representation used for the calculation by using a maximum likelihood method to find for various models of how the monkeys might represent the sum values the Bayesian information criterion (BIC). For this purpose, the data from the last 30 d of symbol set 1 addition were used because that is the data set where the monkeys were most clearly performing addition.

In the linear model, the sum of the two addends was compared with the value of the singleton. In the logarithmic model, which is based on Fechner's original proposal (2) that sensory magnitudes are represented internally by the log of the stimulus magnitude, the internal representation of each addend or singleton magnitude is given by the log of the magnitude +1 (to avoid taking the log of 0). The square root model is another proposed compressed scale that can account for Weber's law (14, 16). In this model, the internal representation is simply the square root of the stimulus magnitude. We also tested a max model, in which the monkeys simply choose the largest option on the screen. For all these models, the only free parameter in the model is the slope of the probit psychometric function fit to the choice probability for each addend–addend–singleton combination (*Experimental Procedures*). To test whether the best-fitting scale is indeed compressive, we compared the two compressed scales (log and square root) to both a linear model and a power model. For the power model, a best-fitting value for the exponent that is less (greater) than 1 indicates that the internal representation is compressive (expansive). Accordingly, the free parameters of the power model were the exponent and the slope of the probit function.

The predicted internal representation for each addend or singleton magnitude is shown in Fig. 3*A* for each of these models. The internal representation functions of both the log model and the square root model are compressive (concave downward), with progressively smaller increments between the internal representations for progressively larger magnitudes, and of course the internal representation function for the linear model is neither compressive nor expansive. The average internal representation function for the max model is expansive (concave upward), because small addends are more likely to be the smaller, ignored, of the two addends, and thus are on average undervalued compared with large addends. Of these simple models, the power model had the lowest BIC; the exponent for the best-fitting power model was 1.24, indicating that the underlying scale must be expansive, not compressive, which is not consistent with Fechner's hypothesis that a compressive internal representation underlies the Weber law behavior of decreasing discriminability with increasing magnitude.

Fig. 3*B* shows the predicted sum value (average point of subjective equality between each sum value and all possible singleton values) for each of the simple models in Fig. 3*A*, assuming that internal representations are combined additively. Combining linear internal representations (Fig. 3*A*, black dots) yields an additive combination (Fig. 3*B*, black dots.). Combining logarithmic internal representations (Fig. 3*A*, green dots) yields an expansive (superadditive) combination (Fig. 3*B*, green dots). Therefore, given that by inspection the shape of the monkeys' sum value behavior (Fig. 2*E*) is overall linear (for dots), or

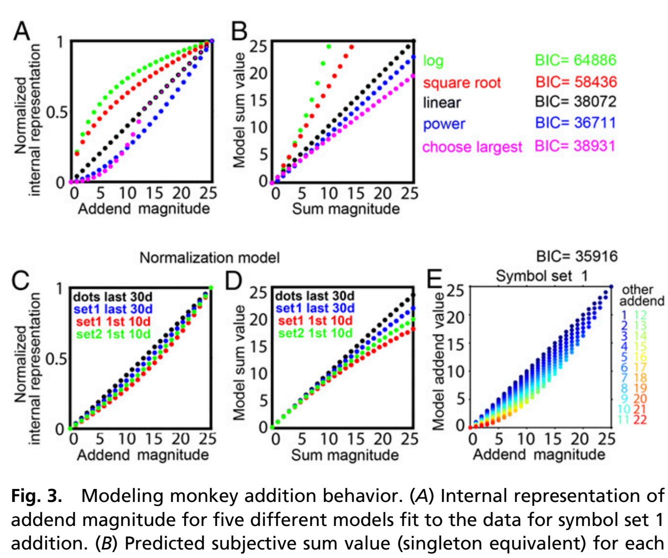


Fig. 3. Modeling monkey addition behavior. (*A*) Internal representation of addend magnitude for five different models fit to the data for symbol set 1 addition. (*B*) Predicted subjective sum value (singleton equivalent) for each model. (*C*) Internal representation of addend magnitude for the normalization model calculated using the parameter k obtained from the fit to different data sets as indicated. (*D*) Predicted sum value for the normalization model calculated using the same parameter k. (*E*) Predicted addend values for each addend magnitude, calculated independently for each other addend with which the addend could be paired, as indicated, using the parameter k obtained from the fit to the symbol set 1 last 30-d data set.

slightly concave downward (for symbol set 1 addition), we conclude that the underlying internal representation of dots must be approximately linear, and for symbols must be slightly expansive, not compressive, consistent with the best-fitting exponent for the power model being greater than 1. One could argue that by rewarding the sum as the linear addition of the two addends, we taught the monkeys to do linear addition, rather than teaching them that the combination of two symbols should be superadditive. Because their initial behavior on the symbol addition task, their persistent behavior over time on the symbol addition task, and their initial behavior on a second symbol set addition task all showed subadditive valuation of the sum, we are inclined to think that superadditive combining would not be expected even if it were rewarded as such.

Although the monkeys' valuation of the sums was a function of both addends, they clearly did not perform accurate addition, because they systematically undervalued the smaller addends. This cannot be explained simply by always undervaluing the symbols representing small rewards, because the same symbol could show a subjective value close to its actual value when it was the larger of the two addends, but a subjective value that was a fraction of its actual value when it was the smaller of the two addends. That is, the monkeys' subjective valuation of each symbol was context dependent, as has been previously described for value coding in midbrain dopamine neurons (23), orbitofrontal cortex neurons (24, 25), LIP neurons (26), and for monkey behavioral value choices (27). This suggests a relative valuation, or normalization. Fig. 2*H* shows the singleton-equivalent value of each symbol set 1 addend (last 30 d) separately for every other addend it could be combined with to give a sum ≤ 25; this plot shows that the monkeys' valuation of each addend is systematically reduced by the increasing magnitude of the other addend. This means the monkeys are basing their valuation of each symbol on its relative value compared with the other symbol simultaneously presented on the same side, not its absolute value.

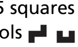
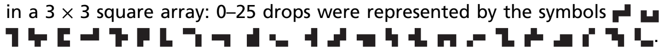
To model a representation in which the value of an addend depends on the magnitude of the other addend, we used normalization, for which biologically plausible mechanisms have been proposed (28). We first fit the symbol set 1 addition data (last 30 d) with a full normalization model in which the internal representation of each quantity is weighted by a hyperbolic function of the p norm of the remaining two quantities (*Experimental Procedures*). The value of the power in the best-fitting model was large ($>10^{11}$), suggesting that the normalization was effectively accomplished by a maximum. Therefore, we used a simpler model in which each quantity was weighted by a hyperbolic function of the maximum of the other two quantities. The most parsimonious model was obtained when the weight for the singleton was set a priori to 1, resulting in a value of BIC = 35,916, which is smaller than any of the models without normalization (Fig. 3*B*). Fig. 3*C* shows the calculated internal representation for each addend using the parameter k obtained by fitting this model to each data set as indicated, and Fig. 3*D* shows the calculated sum value. Fig. 3*E* shows that the predicted addend values are reduced by increasing the other addend in a manner similar to the monkey behavior (compare Fig. 2*H*).

Reference-dependent discriminability is a long-established principle in psychophysics (1), economics (29), and neural coding (30, 31) that could be explained by compressive scaling (logarithmic or power-less-than-one). A century of psychophysics has amassed evidence for a compressive relationship between many kinds of sensory stimuli and perceived sensation, but neurophysiology has shown that although neuronal signaling might be compressed relative to stimulus magnitude, in general the compression of neuronal responsiveness is a dynamic process, involving mechanisms like adaptation, lateral inhibition, and gain control (5, 32). A normalization process could account for the apparent compressed scaling observed in many behavioral studies, as well as the ability to discriminate proportionately over a range of magnitudes (25, 28). In this study, we used addition behavior to ask whether the relative sensitivity, the scaling of just noticeable difference with magnitude, of symbolically represented reward is better explained by a logarithmic, or other compressive internal scale, compared with scalar variability. Instead we found a dynamically shifting, relative scaling. Our result brings the coding of symbolically represented magnitude into agreement with direct measurements of neuronal coding of value.

Experimental Procedures

Touch-Screen Task. The three monkeys were housed in one quad cage with a computer-driven touch screen (Elo TouchSystems) mounted in one quadrant. Software for stimulus presentation, reward delivery, data collection, and data analysis was written in MATLAB (MathWorks). Each monkey spent 2–4 h per day alone, with food, in the training cage, 7 d per week. They were allowed to work to satiety each day, usually performing >500 trials per day. During training/testing the monkey was presented with two options on the two sides of the touch screen; the monkey touched the screen and was rewarded with a number of liquid drops corresponding to the magnitude represented on whichever side of the screen he touched. The liquid was delivered via a stainless steel tube mounted in front of the screen. To make the numerosity of each symbol-associated reward as clear as possible, the liquid was dispensed in discrete drops at 4 Hz by the opening of a computer-driven solenoid, and each drop was accompanied by a beep.

Dots comparison task. Two sets of dots were presented on either side of the screen; each set of dots was enclosed in a 9-cm-diameter circle. The dots were placed at random positions within the circle, and they were of random size and color. When two dots overlapped, the smaller dot always occluded the larger, not vice versa, and the two were constrained to be different colors.

Symbol comparison task. We used two distinct symbol sets of 26 symbols each, 5 cm in height, each set representing 0–25 drops. In symbol set 1 Arabic numerals 0–9 represented 0–9 drops, and the letters X Y W C H U T F K L N R M E A J represented 10–25 drops. Symbol set 2 was generated by filling 4–5 squares in a 3×3 square array: 0–25 drops were represented by the symbols 
.
For a given symbol set, two symbols were presented simultaneously on either side of the touch screen, and the monkey was rewarded with the number of drops corresponding to the symbol value on whichever side of the screen the monkey first touched. Except for the symbols representing zero, the monkey would be rewarded no matter which side he touched, but the monkeys much more often chose the larger of the two options.

Addition Task. Two values between 0 and 25 were chosen by a random number generator; the side that would represent the singleton was chosen randomly, and for the other side, two addends were chosen randomly from all possible combinations that could represent the sum.

Dots addition task. The monkeys were presented with one set of dots on one side of the screen (the singleton) and two sets of dots on the other side (the sum, made up of two "addends").

Symbol addition task. The monkeys were presented with two symbols on one side of the screen (the sum side) and one symbol on the other side (the singleton). In the symbol addition task, the two addend symbols on the sum side were always contained within a single oval, to encourage the monkeys to recognize the two-symbol combination as a single choice option. The monkeys first learned the dots comparison task, then symbol set 1 comparison task, then dots addition, then symbol set 1 addition, followed by symbol set 2 comparison, and lastly symbol set 2 addition.

Analysis of Behavioral Data. Although the monkeys were rewarded no matter which side of the screen they touched (except for value zero), they usually chose the larger side; we therefore will refer to larger choices as "correct." We also calculated percent correct for the situations we defined as "mandatory calculation" conditions, for those conditions when the sum was larger than the singleton, but neither addend was larger than the singleton.

To find the subjective value (singleton-equivalent value) of each addend–addend sum (averaged over all addend combinations) we fit a logistic psychometric function (with a lapse rate, gamma) to the choice ratios; the parameters (slope, mu, and gamma) were estimated using maximum likelihood, as shown in Fig. 2*D*. The point of subjective equality between each sum and all singletons with which it was paired was taken as the singleton-equivalent value of each sum.

To quantify the contribution of the small and large addends to the sum value, for each addend magnitude individually, we fit a logistic function to the difference between the singleton and the other addend magnitude. We also fit the following logistic regression model to the probability of choosing the singleton, p(single). Logit p(single) = $a_0 + a_1 X_{single} + a_2 X_{small} + a_3 X_{large}$, where

X_{single}, X_{small}, and X_{large} denote the magnitude of the singleton, smaller addend, and larger addend, respectively, and a_0–a_3 the corresponding regression coefficients. From this model, the relative contribution of small and large addend in the unit of singleton (referred to as small and large addend weights) can be estimated as $-a_2/a_1$ and $-a_3/a_1$, respectively.

Modeling Choice Behavior. To identify the nature of the internal representation most consistent with the observed addition behaviors, we calculated the BIC for a probit psychometric function combined with several possible functions. Representing the singleton, small addend, large addend as X_{single}, X_{small}, and X_{large}, respectively, and the internal representations of singleton and sum as Y_{single} and Y_{sum}, respectively, we first considered the following five simple models for the monkeys' internal representation of presented addend and singleton magnitudes:

Log model:

$$Y_{single} = \ln(X_{single} + 1) \quad \text{vs.} \quad Y_{sum} = \ln(X_{small} + 1) + \ln(X_{large} + 1),$$

Square root model:

$$Y_{single} = X_{single}^{1/2} \quad \text{vs.} \quad Y_{sum} = X_{small}^{1/2} + X_{large}^{1/2},$$

Linear model:

$$Y_{single} = X_{single} \quad \text{vs.} \quad Y_{sum} = X_{small} + X_{large},$$

Power model:

$$Y_{single} = X_{single}^{\alpha} \quad \text{vs.} \quad Y_{sum} = X_{small}^{\alpha} + X_{large}^{\alpha},$$

Max model:

$$Y_{single} = X_{single} \quad \text{vs.} \quad Y_{sum} = \max(X_{small}, X_{large}).$$

For each of these models, the probability of choosing the singleton was given by the normal cumulative distribution function, i.e., p(singleton) = normcdf $\{\beta(Y_{single} - Y_{sum})\}$. All model parameters were estimated using the fminsearch function in MATLAB (MathWorks Inc.).

We also tested several different normalization models in which the internal representation of each quantity was normalized by a function of the two other quantities. We used the p norm to investigate systematically how this normalization process was influenced by the two other quantities. In other words,

$$Y_{single} = k_S \cdot X_{single}, \ \& \ Y_{sum} \equiv Y_{small} + Y_{large} = k_1 \cdot X_{small} + k_2 \cdot X_{large},$$

where k_{single}, k_1, and k_2 are given by the hyperbolic function of the p norm of the other two magnitudes, namely, $k_S = 1/\left(1 + k\left(X_{small}^{z} + X_{large}^{z}\right)^{1/z}\right)$, $k_1 = 1/\left(1 + k\left(X_{single}^{z} + X_{large}^{z}\right)^{1/z}\right)$, $k_2 = 1/\left(1 + k\left(X_{single}^{z} + X_{small}^{z}\right)^{1/z}\right)$. Therefore, the free parameters of this full normalization model were k and z in addition to the slope parameter β in the probit function. The BIC for this model was 35,971, but it gave a large value of z ($>10^{11}$), indicating that the p norm effectively performed a max operation. We therefore fit a simpler model in which normalization is accomplished by the maximum of the remaining magnitude, to the same data. Namely,

$$k_S = 1/\left(1 + k \cdot \max(X_{single}, X_{large})\right), \ k_1 = 1/\left(1 + k \cdot \max(X_{single}, X_{large})\right),$$
$$k_2 = 1/\left(1 + k \cdot \max(X_{single}, X_{small})\right).$$

The best fit for this model gave a BIC = 35,961, a better fit, with one fewer parameter.

We found an even better fit using a simpler model in which only the addends are normalized by each other's magnitude, namely, by setting $k_S = 1$ (i.e., $Y_{single} = X_{single}$). This model gave a BIC = 35,916.

ACKNOWLEDGMENTS. This work was supported by National Institutes of Health Grants EY16187 and DA029330.

1. Weber EH (1834) *De Pulsu, Resorptione, Auditu et Tactu Annotationes Anatomicae et Physiologicae* (CF Koehler, Leipzig, Germany).
2. Fechner GT (1860) *Elemente der Psychophysik* (Breitkopf und Hèartel, Leipzig, Germany). German.
3. Stevens SS (1961) To honor Fechner and repeal his law: A power function, not a log function, describes the operating characteristic of a sensory system. *Science* 133(3446): 80–86.
4. Dean AF (1981) The variability of discharge of simple cells in the cat striate cortex. *Exp Brain Res* 44(4):437–440.
5. Barlow HB (1965) Optic nerve impulses and Weber's law. *Cold Spring Harb Symp Quant Biol* 30:539–546.
6. Lee D, Port NL, Kruse W, Georgopoulos AP (1998) Variability and correlated noise in the discharge of neurons in motor and parietal areas of the primate cortex. *J Neurosci* 18(3):1161–1170.
7. Cantlon JF, Cordes S, Libertus ME, Brannon EM (2009) Comment on "Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures". *Science* 323(5910):38, author reply 38.
8. Whalen J, Gallistel C, Gelman R (1999) Nonverbal counting in humans: The psychophysics of number representation. *Psychol Sci* 10(2):130–137.
9. Brannon EM, Wusthoff CJ, Gallistel CR, Gibbon J (2001) Numerical subtraction in the pigeon: Evidence for a linear subjective number scale. *Psychol Sci* 12(3):238–243.
10. Gibbon J (1977) Scalar expectancy theory and Weber's law in animal timing. *Psychol Rev* 84:279–325.
11. Nieder A, Diester I, Tudusciuc O (2006) Temporal and spatial enumeration processes in the primate parietal cortex. *Science* 313(5792):1431–1435.
12. Nieder A, Freedman DJ, Miller EK (2002) Representation of the quantity of visual items in the primate prefrontal cortex. *Science* 297(5587):1708–1711.
13. Nieder A, Miller EK (2004) A parieto-frontal network for visual numerical information in the monkey. *Proc Natl Acad Sci USA* 101(19):7457–7462.
14. Nieder A, Miller EK (2003) Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron* 37(1):149–157.
15. Roitman JD, Brannon EM, Platt ML (2007) Monotonic coding of numerosity in macaque lateral intraparietal area. *PLoS Biol* 5(8):e208.
16. Stevens SS (1961) The psychophysics of sensory function. *Sensory Communication*, ed Rosenblith WA (MIT Press, Cambridge, MA).
17. Dehaene S (2001) Subtracting pigeons: Logarithmic or linear? *Psychol Sci* 12(3): 244–246, discussion 247.
18. Cordes S, King AP, Gallistel CR (2007) Time left in the mouse. *Behav Processes* 74(2): 142–151.
19. Cantlon JF, Brannon EM (2007) Basic math in monkeys and college students. *PLoS Biol* 5(12):e328.
20. Boysen ST, Berntson GG (1989) Numerical competence in a chimpanzee (Pan troglodytes). *J Comp Psychol* 103(1):23–31.
21. Livingstone MS, Srihasam K, Morocz IA (2010) The benefit of symbols: Monkeys show linear, human-like, accuracy when using symbols to represent scalar value. *Anim Cogn* 13(5):711–719.
22. Srihasam K, Mandeville JB, Morocz IA, Sullivan KJ, Livingstone MS (2012) Behavioral and anatomical consequences of early versus late symbol training in macaques. *Neuron* 73(3):608–619.
23. Tobler PN, Fiorillo CD, Schultz W (2005) Adaptive coding of reward value by dopamine neurons. *Science* 307(5715):1642–1645.
24. Kobayashi S, Pinto de Carvalho O, Schultz W (2010) Adaptation of reward sensitivity in orbitofrontal neurons. *J Neurosci* 30(2):534–544.
25. Padoa-Schioppa C (2009) Range-adapting representation of economic value in the orbitofrontal cortex. *J Neurosci* 29(44):14004–14014.
26. Louie K, Grattan LE, Glimcher PW (2011) Reward value-based gain control: Divisive normalization in parietal cortex. *J Neurosci* 31(29):10627–10639.
27. Louie K, Khaw MW, Glimcher PW (2013) Normalization is a general neural mechanism for context-dependent decision making. *Proc Natl Acad Sci USA* 110(15):6139–6144.
28. Carandini M, Heeger DJ (2012) Normalization as a canonical neural computation. *Nat Rev Neurosci* 13(1):51–62.
29. Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263–292.
30. Hartline HK, Wagner HG, MacNichol EF, Jr. (1952) The peripheral origin of nervous activity in the visual system. *Cold Spring Harb Symp Quant Biol* 17:125–141.
31. Kuffler SW (1953) Discharge patterns and functional organization of mammalian retina. *J Neurophysiol* 16(1):37–68.
32. Uttal WR (1973) *The Psychobiology of Sensory Coding* (Harper & Row, New York), p xvi.