

Research Article

NUMERICAL SUBTRACTION IN THE PIGEON: Evidence for a Linear Subjective Number Scale

Elizabeth M. Brannon,¹ Courtney J. Wusthoff,² C.R. Gallistel,³ and John Gibbon^{2,4}

¹Center for Cognitive Neuroscience and Department of Psychology, Experimental, Duke University; ²Department of Psychology, Columbia University; ³Department of Psychology, University of California, Los Angeles; and ⁴Department of Biopsychology, New York State Psychiatric Institute

Abstract—When humans and animals compare two numbers, responding is faster and more accurate with increasing numerical disparity and decreasing numerical size. Researchers explaining these distance and size effects often assume that the subjective number continuum is logarithmically compressed. An alternative hypothesis is that the subjective number continuum is linear, but positions farther along it are proportionately fuzzier, that is, less precisely located. These two hypotheses have been treated as functionally equivalent because of their similar empirical predictions. The current experiment sought to resolve this issue with a paradigm originally developed to address the subjective representation of time (time left). In our adaptation, pigeons were required to compare a constant number with the number remaining after a numerical subtraction. Our results indicate that subjective number is linearly, not logarithmically, related to objective number.

Converging evidence from many different laboratories and paradigms demonstrates that animals represent number (for reviews, see Davis & Perusse, 1988; Gallistel & Gelman, 1992, 2000; Dehaene, 1997). For example, animals can produce a certain number of responses (e.g., Mechner, 1958), discriminate the numerosity of visual or auditory stimuli (e.g., Hicks, 1956; Meck & Church, 1983), represent the ordinal relations between numerosities (Brannon & Terrace, 1998, 2000), and even learn the relationship between arbitrary symbols and numerosities (e.g., Matsuzawa, 1985).

A growing body of data supports the conclusion that nonhuman animals and humans represent number in a similar way (Dehaene, Dehaene-Lambertz, & Cohen, 1998; Gallistel & Gelman, 2000). For example, when human or animal subjects are required to determine which of two numerical values is larger (or smaller), their accuracy and reaction time are systematically influenced by both the numerical distance between the two values (the distance effect) and their numerical size (the size effect; Brannon & Terrace, 1998, 2000; Moyer & Landauer, 1967). Reaction time is longer and accuracy is lower for values that are similar in numerical size (e.g., 2 and 3) than for values that are disparate in size (e.g., 2 and 6). When distance is held constant, accuracy decreases and latency increases as the numerical size of the values being compared increases.

One explanation for the effect of numerical size on numerical ordering is that the internal representation of number is logarithmically compressed (e.g., Dehaene & Mehler, 1992; Moyer & Landauer, 1967; Rule, 1969), so that the representatives of, say, 12 and 13 lie closer together on a mental continuum than do the representatives of 2

and 3. The closer together two numerosities lie on the mental continuum, the harder it will be to determine their ordering. The logarithmic-compression hypothesis is grounded in a tradition of interpreting magnitude-estimation data as evidence for a nonlinear relationship between subjective and objective continua (e.g., Eisler, 1976; Stevens, 1961). However, an alternative lesser-known possibility is that the subjective number scale may be a linear reflection of objective number, but the uncertainty about where exactly a given quantity falls on the mental continuum may increase in proportion to the size of that quantity. Thus, the ordering of two big numbers will be harder to distinguish than the ordering of two small numbers the same distance apart because the positions of the big numbers on the mental number continuum are more fuzzily defined (Fetterman, 1993; Gallistel & Gelman, 1992; Gibbon, 1977; Whalen, Gelman, & Gallistel, 1999). Both hypotheses predict the observed effect of numerical size on ordering (e.g., more difficulty in judging the order of 12 and 13 than 2 and 3).

Recent research suggests that animals represent time and number in similar ways (Meck & Church, 1983; Meck, Church, & Gibbon, 1985; Roberts, 1997; Roberts & Boissvert, 1998). For example, Meck and Church (1983) used a psychophysical choice procedure to train rats to discriminate signals that differed in both duration and number. Although duration and number were confounded in training, test sessions revealed that the rats had encoded both duration and number. Furthermore, the data suggest that a scaling factor of about 200 ms per count relates the subjective number continuum to the subjective time continuum (Meck & Church, 1983; Meck et al., 1985).

Gibbon and Church (1981) showed that one can distinguish between linear and logarithmically compressed continua with a task that requires the subject to base its behavior on the difference between two points on the continuum rather than on their ordering. Any two points that have the same ratio are the same distance apart on a logarithmically compressed continuum. If the subjective number continuum is logarithmically compressed, then two pairs of numbers with the same ratio—for example, the pairs {2,3} and {20,30}—have the same subjective separation; that is, the pairs of subjective representatives of these numbers lie the same distance apart on the subjective number continuum. Thus, if the subjective number scale is logarithmically compressed, when numerical behavior is based on the subjective difference between two numbers, the behavior will be the same whenever the objective ratio of the two numbers is the same, no matter how far apart those two numbers are objectively. Gibbon and Church used this fact to show that the subjective representation of time is linearly related to objective time. In the research we report here, we adapted their paradigm to show that subjective number is linearly related to objective number.

Our task requires subjects to compare a numerical difference that varies from trial to trial with a constant value. Subjects choose whichever seems smaller, the difference or the constant value. If the number scale is logarithmically compressed, then whether the difference is bigger than the constant will depend not on the objective difference

Address correspondence to Elizabeth M. Brannon, Center for Cognitive Neuroscience, Box 90999, Duke University, Durham, NC 27708-0999; e-mail: brannon@duke.edu.

between the two numbers but rather on their objective ratio. Thus, the number that is subjectively equal to a given subjective difference will not increase when we make the values objectively bigger but keep the same ratio between them. Of course, when we scale up two numbers in this way, we make the objective difference between them bigger, so this outcome would be unexpected if the subjective representation of number is linearly related to objective number.

We call the constant number with which differences are to be compared the *standard*, or *S*, number. The first of the two numbers whose difference is to be compared with *S* is called the *initially required*, or *I*, number. The second is called the *tallied*, or *T*, number. At the moment of choice, the subject must judge whether the initially required number of pecks minus the tallied number is less than the standard number (i.e., whether $I - T < S$). If so, the subject chooses the key called the *number-left key*; if not, it chooses the other key, called the *standard key*.

The task works like this: First, the middle key among three keys is illuminated, signaling its activation. Pecking it then produces brief flashes of the light in a food hopper on a variable-ratio schedule, a schedule in which the number of pecks required to produce a flash varies unpredictably about a mean value. With a variable-ratio schedule, the amount of time required to generate a given number of hopper flashes varies considerably. We make use of this fact to deconfound time and number in our data analysis. After the pigeon's pecking of the middle key has produced a variable number, *T*, of hopper flashes, the light on the middle key goes out, and the keys on either side are illuminated, confronting the pigeon with a choice: Which side key should it peck? Whichever side key it pecks first it must then continue to peck in order to collect a food reward, because pecking one side key inactivates the other side key. Continued pecking of a side key generates hopper flashes, and the pigeon gets to eat from the hopper after it has produced the requisite number of flashes for the side key it has chosen. The requisite number is always the same on the standard side key. The requisite number on the number-left side key, however, is $I - T$, where *T* is the number of flashes the pigeon has already tallied by pecking on the middle key. The standard key is the better choice when $I - T > S$; the number-left key is the better choice when $I - T < S$.

The value of *T* at which the subject is equally likely to choose the number-left key and the standard key is called the subjective indifference point. The critical question in this experiment is whether *T* increases when we increase the values of *I* and *S* by the same factor, thereby maintaining their ratio. The indifference point depends on how far apart *I* and *S* are on the subjective continuum, because the subjective magnitude of *T* at the indifference point must be equal to the subjective difference (or separation) between *I* and *S*. The indifference point indicates the difference (the number left) that seems to the subject to be equal to the standard number. If the mental continuum is logarithmically compressed, then this subjective difference will depend on the ratio of the two numbers, not their linear distance. So long as we keep their ratio the same, we should keep the value of *T* at the indifference point the same.

METHOD

Subjects

The subjects were 4 White Carneaux pigeons maintained at approximately 80% ad lib body weight. They were exposed to a regular 12:12 hr light/dark cycle and were tested 5 days a week.

Apparatus

Experimental sessions were conducted in a BRS/LVE Small Environment Cubicle (SEC-002), with minimal external light and noise. Normal illumination was provided by an overhead houselight, except during reinforcement presentation. Stimuli were presented through the BRS/LVE Pigeon Intelligence Panel (PIP-010 through PIP-016), equipped with three illuminated pecking keys. Using an IEEE one-plane readout projector (No. 00010-01-XXXX-1820), we illuminated each key a specific color; the central key was white in each case, and the colors of the left and right keys were yellow and green, counterbalanced for color and position across subjects. A Macintosh computer controlled the experiments and recorded the data.

Procedure

Training

Subjects were first trained to peck two keys on a continuous reinforcement schedule. On each trial, one of two keys was illuminated, and pecking at the illuminated key resulted in 3.5 s of access to the illuminated hopper. One key was green and the other yellow, and the left/right position of the two keys was counterbalanced across subjects. Subjects were then shifted to a VR8 reinforcement schedule (variable-ratio schedule: $M = 8$, range: 4–12).

The next phase of training required the pigeons to keep track of the number of hopper flashes (100–300 ms each), rather than the number of pecks or the time spent pecking. Subjects were reinforced the second time the hopper was illuminated, and because each hopper illumination was produced on a VR8 schedule with a range of 4 to 12, the birds were required to peck anywhere from 8 to 24 times for each reinforcement. Subsequently, the number of flashes to reinforcement was increased to four, requiring anywhere from 16 to 48 pecks per reinforcement. In the next phase of training, the number of flashes to reinforcement was again increased, but only for the number-left key, which then required eight flashes to reinforcement; the standard key still required four flashes. Each flash was provided on a VR8 schedule.

A white middle key was then introduced. On each trial, subjects were required to respond to the middle key to produce a number, *T*, of hopper flashes between 1 and 7. Flashes were again provided on a VR8 schedule. After *T* flashes, the middle key was extinguished, and either the standard or the number-left key became available. If the standard key was illuminated, the subject was required to produce 4 additional flashes for reinforcement regardless of the number of the flash tally from the middle key. However, if the number-left key was illuminated, the subject was required to produce $8 - T$ additional flashes before reinforcement. There were 42 trials per session; each of the seven possible *T* values was tested six times in a random order.

In the final phase of training, subjects were given a choice between the standard and number-left keys on a subset of trials. After pecking produced *T* flashes on the middle key, both the standard and the number-left keys were illuminated. The first response to one key eliminated the availability of the other key. As in the previous phase of training, the standard key required 4 additional flashes to reinforcement, and the number-left key required $8 - T$ flashes to reinforcement. The remaining trials were *forced-choice trials* in that only the standard or number-left key was illuminated, as in the previous phase (no choice). In each session, choice trials were presented on 2 out of every 6 trials for each of the seven middle-key values (i.e., 14 choice

Subjective Number Scale

trials per session). The intertrial interval was 10 s throughout training and testing.

Training continued until the choices the animals made varied systematically as a function of the flashes obtained on the middle key. Training required 25 to 30 days for each subject.

Testing

The first testing condition was essentially identical to the final training phase in that the standard key required 4 flashes to reinforcement and the number-left key required $8 - T$ flashes to reinforcement ($S = 4, I = 8$). The only change was that subjects were given a choice between the standard and number-left keys on 4 of the 6 trials at each of the middle-key values (i.e., on 28 trials). The remaining trials were forced (i.e., either the standard or the number-left key was illuminated at the T th middle-key flash). This test phase continued for 40 sessions.

In the second testing condition, the S and I values were multiplied by a factor of 1.5. Thus, the standard key required 6 flashes to reinforcement, and the number-left key required $12 - T$ flashes to reinforcement ($S = 6, I = 12$). The middle key required 1, 2, 3, 4, 6, 8, or 10 flashes. There were 28 choice trials per session and 14 forced-choice trials per session. Birds were tested on these values for 25 days.

In the third testing condition, the values for the standard and number-left keys were returned to the values in the first testing session ($S = 4, I = 8$). Ten sessions were conducted.

Finally, in the fourth testing condition, the standard and number-left key values were reduced to 3 and $6 - T$, respectively ($S = 3, I = 6$). The middle key required 1, 2, 3, 4, or 5 flashes. There were 28 choice trials per session and 14 forced-choice trials per session. Ten sessions were conducted.

RESULTS

The central findings were that the pigeons' choice between the two side keys was controlled by the number of flashes generated by pecking the middle key and that increasing the I and S values by varying amounts had a linear effect on the indifference point; the bigger the objective difference between I and S , the higher the indifference point, contrary to the predictions of the logarithmic-compression hypothesis.

Figure 1 shows the probability of choosing the number-left key as a function of the number of flashes on the middle key in the first condition, where $I = 8$ and $S = 4$. When the tally of flashes generated by pecking on the middle key was low, the pigeons chose the standard key; when it was high, they chose the number-left key. A cumulative normal ogive fit to these data accounted for 99% of the variance. The indifference point is the tally (number of middle-key flashes) at which this ogive crosses the .5 level. When the tally was less than the indifference point, the birds preferred the standard key; when it was greater than the indifference point, they preferred the number-left key.

Figure 2a shows the same function (probability of choosing the number-left key as a function of the tally of middle-key flashes) for each of the four conditions. Each function reflects average values for the 4 birds for the last five sessions of the condition.¹ The indifference

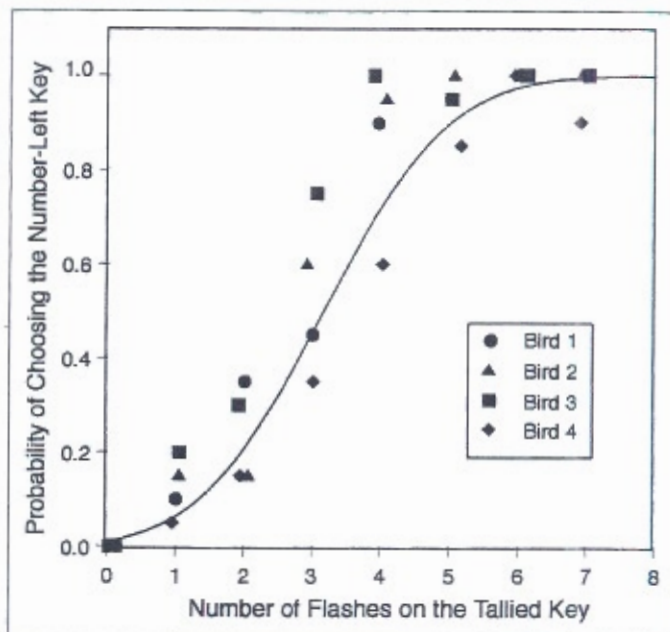


Fig. 1. Psychometric functions relating probability of choosing the number-left key to the tallied number of flashes on the middle key for each of the 4 birds for the last 5 days of the first testing condition. The fitted line is the best-fit model by a cumulative normal distribution.

point shifted to the right, toward higher tallies, when the S and I values were increased to 6 and 12 and to the left, toward lower tallies, when they were reduced to 3 and 6. Figure 2b shows that the functions for the four conditions superpose when the probability of choosing the number-left key is plotted as a function of the proportion of middle-key flashes. This proportion is arrived at by dividing the number of tallied flashes by the maximum number of flashes that could occur on the tallied key for that condition.

Cumulative normal functions fit to the data in Figure 2 accounted for 99% of the variance in each case, so these curves accurately represent the data. The tallies at which these curves cross the .5 level give the indifference points for the conditions. Figure 3 plots these indifference points for each of the 4 birds, as a function of the value of S , hence also as a function of the value of I , which was twice S in each condition. For each bird, the indifference point appears to be a linearly increasing function of S , and thus of $S - I$. Regression analyses revealed that for each bird, the slope of the function in Figure 3 was significantly different from zero ($p < .05$); the intercepts were not significantly different from zero ($p > .1$). Thus, increasing or decreasing the S and I values by the same factor produced a proportional increase or decrease in the indifference point.

Furthermore, there was a significant shift in the indifference point from the last 5 days of the first condition, where $I = 8$ and $S = 4$, to the first session of the second condition, where $I = 12$ and $S = 6$ (2.8 vs. 3.593), $t(3) = -3.09, p = .05$.

It is already known that subjective time increases as a linear function of objective time (Gibbon & Church, 1981). Our goal in this study was to test whether for number also, subjective representations are a linear function of objective magnitude. Because pigeons peck at an approximately constant rate, the time spent pecking a key tends to covary with the number of flashes thereby generated. Furthermore, it has been

1. Although data throughout this article reflect the last five sessions of each condition, there is one exception. Data for Bird 4 in the second testing condition ($S = 6, I = 12$) were taken from the 16th through 20th sessions (rather than 21st through 25th) because his indifference point decreased in the last five sessions and was not representative of his previous performance.

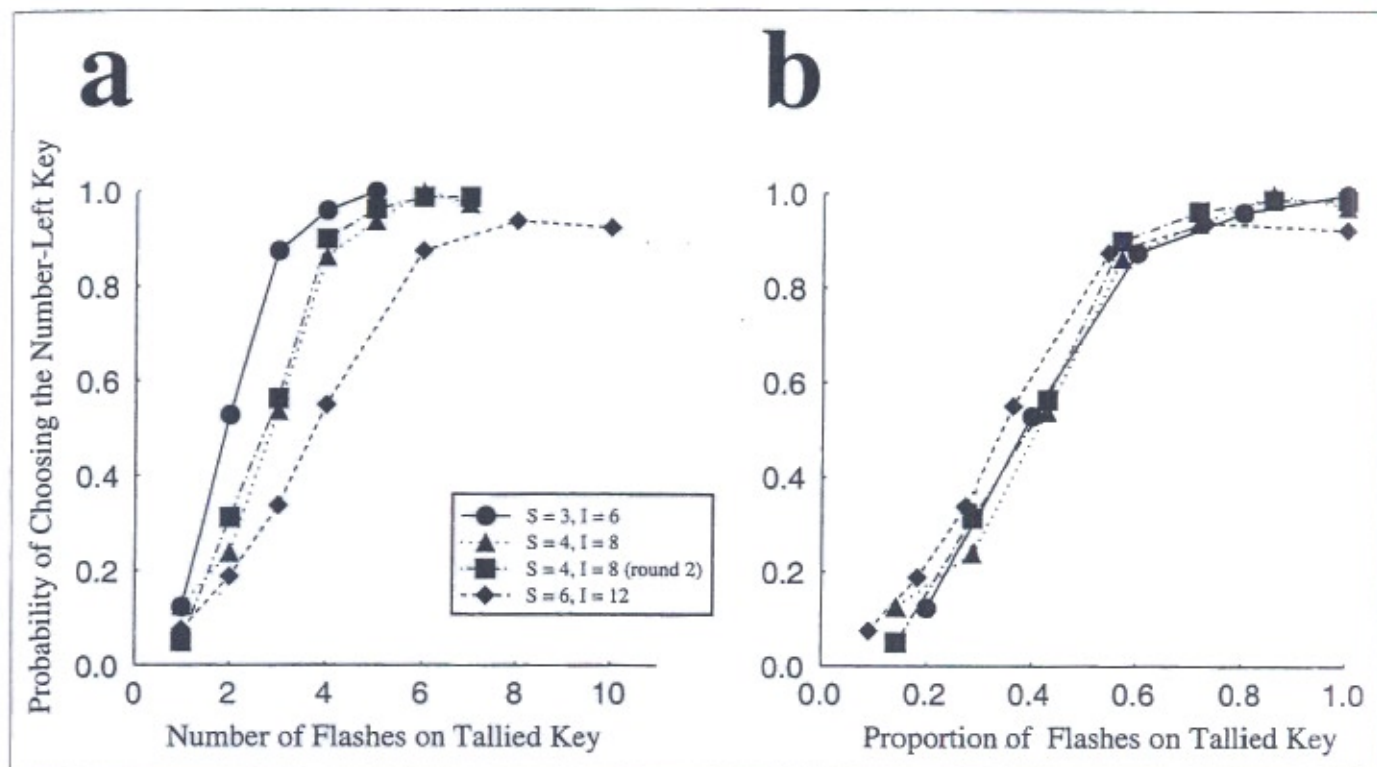


Fig. 2. Psychometric functions relating the probability of choosing the number-left key to the tallied number of flashes (a) and the proportion of tallied flashes (b) on the middle key. Each function reflects one of the four testing conditions. The proportion of tallied flashes is arrived at by dividing the number of tallied flashes by the maximum number of flashes that could occur on the tallied key for that condition.

demonstrated that timing sometimes contributes to numerosity discrimination in pigeons (Fetterman, 1993). It was therefore essential to show that the pigeons' choices were controlled by the number of hopper flashes generated, and not by the time spent generating them. That was the rationale for delivering flashes on a variable-ratio schedule. Such a schedule produces considerable variation in the number of pecks, hence the amount of pecking time, required to generate any given number of hopper flashes.

Figure 4a shows the cumulative distribution functions for the time spent pecking on the middle key when T ranged from one to seven flashes. The point where each cumulative function begins to rise gives the minimum amount of time spent generating the requisite number of flashes; the point where each function attains asymptote gives the maximum amount of time so spent. Clearly, for each value of T , there was great variation in time spent on the tallied key. Most important, there was a great deal of overlap in these time distributions for different numbers of flashes (different tallies). Thus, we determined whether the pigeons' behavior was controlled by time or by the number of flashes generated by plotting the probability of choosing the number-left key after a given tally of hopper flashes as a function of the time spent generating that tally (Fig. 4b). If time influenced pigeons' choice, these functions would rise, but in fact they are more or less flat, so variation in time did not affect choice. By contrast, for each increase in the middle-key flash tally, there was a marked increase in the probability of choosing the number-left key (Fig. 2a). Thus, choices were controlled by the flash tally, not by the time spent generating that tally.

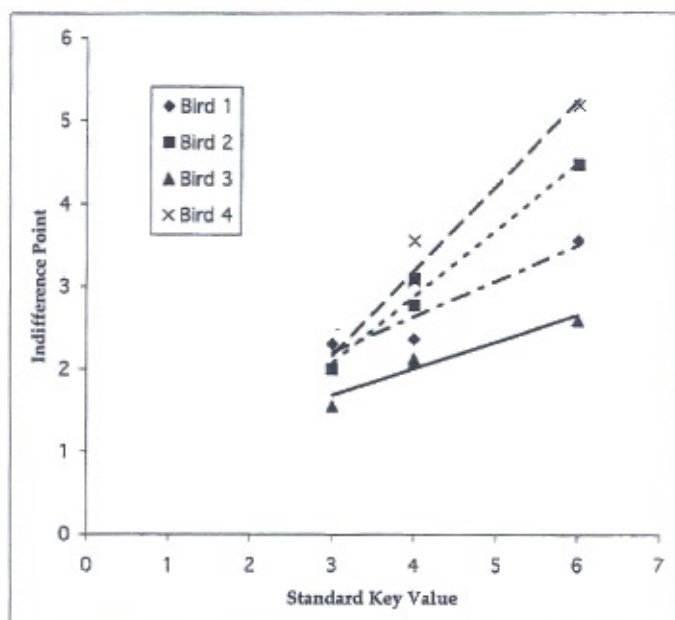


Fig. 3. Indifference point for each of the four conditions for each bird. The lines reflect the best-fit linear models for the 4 birds.

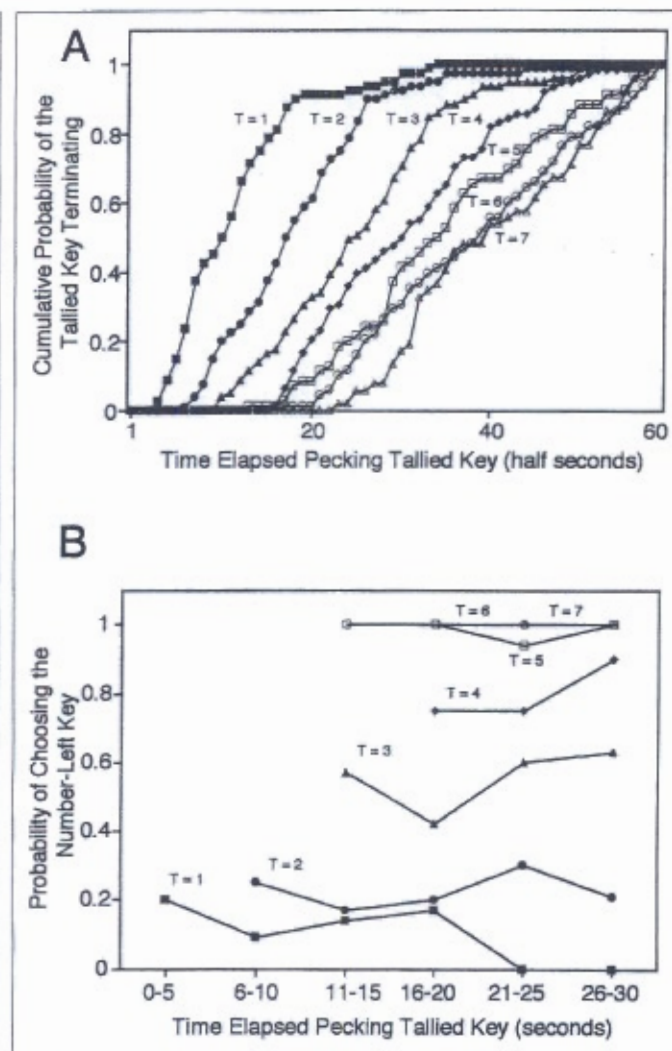


Fig. 4. Cumulative probability distributions for the time spent pecking the middle key for each tallied number (T ; a) and conditional probability of choosing the number-left key as a function of the amount of time spent pecking the middle key (b). The amount of time spent pecking the middle key was a continuous variable but was averaged in discrete time bins for simplification. The data reflect average values for the 4 birds for the last 5 days of the first testing condition ($S = 4$, $I = 8$).

DISCUSSION

In this experiment, pigeons were required to estimate the numerical difference between two values ($I - T$) and compare that difference to a constant number, S . If, as has commonly been assumed, the subjective representation of numerosity were logarithmically compressed, then the tally number T at which the pigeons judged the difference between I and T (the number left) to equal S would be constant for a constant ratio between I and S , even though the objective difference between I and S was made bigger by increasing both numbers by a common factor. This prediction of the logarithmic-compression hypothesis is clearly false.

The alternative model, based on Gibbon's (1977) scalar expectancy

theory, is that the subjective continuum is linearly (indeed, almost proportionately) related to the objective continuum, but that the variability of the magnitudes read from memory increases in proportion to their mean value. The magnitudes representing the fixed numerosities I and S must live in memory, and be read from memory on every trial.

In scalar expectancy theory, the variability in the values read from memory accounts for the trial-to-trial variability in which key the subject chooses after a given flash tally. For any given intermediate tally, the subject sometimes chooses the standard key and sometimes chooses the number-left key. However, the greater the tally, the more likely the subject is to choose the number-left key. In scalar expectancy, the trial-to-trial variability is explained by assuming that on each trial, the subject reads from numerical memory subjective magnitudes for S and I . Because memory is noisy, the magnitudes read vary from trial to trial. On a trial in which the magnitude read for S happens to be unusually low and the magnitude read for I is unusually high, the tally required in order for the subject to choose the number-left key will be unusually high. On trials in which the opposite variations occur, the tally required will be unusually low. That is why, for a given intermediate tally, the subject sometimes chooses the number-left key and sometimes chooses the standard key. Gibbon and Fairhurst (1994) have shown that in order to predict the superposability of the cumulative distributions shown in Figure 2b, it is necessary to assume (a) that the variability in the magnitudes read from memory is proportional to the target magnitudes and (b) that the subjective variable that determines which choice the subject makes is the ratio of the two quantities being compared (in this case, the ratio of $I - T$ and S).

An Alternative Explanation

As in the time-left paradigm, the number-left paradigm used here assumes that subjects subtract their count of the flash tally (their subjective tally) from their memory for the initial value, I . Subjects must then compare the result of this subtraction to their subjective reference value for the standard (also from memory). An alternative explanation, also considered by Gibbon and Church (1981), involves a kind of paired-associate learning. High tallies on the middle key are associated with short delays of reinforcement on the number-left key, whereas low tallies are associated with long delays. Perhaps the birds learn simply to avoid the number-left key after low tallies and to choose it after high tallies. Although we cannot entirely rule out this possibility, our data suggest that it is unlikely, because the indifference point shifted significantly in the first session when the S and I values were multiplied by a factor of 1.5. Although it is possible that this observed change in the indifference point reflects rapid learning, the fact that there were only seven trials at each of the tallies makes this explanation unlikely. Nevertheless, it would be of interest to determine whether the observed shift in the indifference point would occur in un-reinforced test sessions.

Implications for the Neural Representation of Number

The logarithmic-compression hypothesis states that the size and distance effects in numerical ordering judgments result from a logarithmically spaced subjective number scale. Our results suggest that, psychologically, the number scale is organized linearly with respect to real number. Although our results suggest that behavior in the current task is mediated by a linear representation of number, the possibility

exists that the neural implementation of this linear number representation might itself be logarithmic and then transformed to a linear scale before numerical ordering and subtraction can occur (Dehaene, this issue). Alternatively, the brain may implement a second, logarithmic, subjective number scale that is independent of the linear subjective number scale accessed in the current task. Neither our study nor any other published studies provide data that can address either of these two hypotheses. Given the current state of knowledge, we view the idea that number is represented in the brain both linearly and logarithmically as unparsimonious. This debate underscores the need for further research on the neurobiology of number representation.

In summary, our findings support the conclusion that the subjective number scale is not logarithmic. This implies that the size and distance effects, which inspired the logarithmic-compression hypothesis, are rather to be attributed to the fact that the variability in the signals read from numerical memory is proportional to their mean value. This property is commonly called scalar variability (Gibbon, 1977). We believe our experiment also demonstrates, almost incidentally, that pigeons can do numerical subtraction.

Acknowledgments—We gratefully thank Steve Fairhurst for writing the programs to execute the experiment and analyze the data. We also thank Corby Dale for being a relief pigeon runner in times of need and Michael Platt for helpful comments on the article. This research was supported by a National Institutes of Health MERIT award (MH41649-09) to John Gibbon and a National Research Service Award (MH12065-02) to Elizabeth M. Brannon.

REFERENCES

- Brannon, E.M., & Terrace, H.S. (1998). Ordering of the numerosities 1-9 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.
- Davis, H., & Perusse, R. (1988). Numerical competence: From backwater to mainstream of comparative psychology. *Behavioral Brain Sciences*, 11, 602-615.
- Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representation of numbers in the animal and human brain. *Nature Neuroscience*, 21, 355-361.
- Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, 43, 1-29.
- Eisler, H. (1976). Experiments on subjective duration 1868-1975: A collection of power function exponents. *Psychological Bulletin*, 83, 1154-1171.
- Fetterman, J.G. (1993). Numerosity discrimination: Both time and number matter. *Journal of Experimental Psychology: Animal Behavior Processes*, 19, 149-164.
- Gallistel, C.R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, 44, 43-74.
- Gallistel, C.R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences*, 4, 59-65.
- Gibbon, J. (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 the Experimental Analysis of Behavior*, 7, 87-107.
- Gibbon, J., & Fairhurst, S. (1994). Ratio versus difference comparators in choice. *Journal of the Experimental Analysis of Behavior*, 62, 409-434.
- Hicks, L.H. (1956). An analysis of number-concept formation in the rhesus monkey. *Journal of Comparative and Physiological Psychology*, 49, 212-218.
- Matsuzawa, T. (1985). Use of numbers by a chimpanzee. *Nature*, 315, 57-59.
- Mechner, F. (1958). Probability relations within response sequences under ratio reinforcement. *Journal of the Experimental Analysis of Behavior*, 1, 109-122.
- Meck, W.H., & Church, R.M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, 9, 320-334.
- Meck, W.H., Church, R.M., & Gibbon, J. (1985). Temporal integration in duration and number discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, 11, 591-597.
- Moyer, R.S., & Landauer, T.K. (1967). Time required for judgments of numerical inequality. *Nature*, 215, 1519-1520.
- Roberts, W.A. (1997). Does a common mechanism account for timing and counting phenomena in the pigeon? In C.M. Bradshaw & E. Szabadi (Eds.), *Time and behavior: Psychological and neurobiological analyses* (pp. 185-215). New York: Elsevier Science.
- Roberts, W.A., & Boissvert, M.J. (1998). Using the peak procedure to measure timing and counting processes in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 24, 416-430.
- Rule, S.J. (1969). Equal discriminability scale of number. *Journal of Experimental Psychology*, 79, 35-38.
- Stevens, S.S. (1961). To honor Fechner and repeal his law. *Science*, 133, 80-86.
- Whalen, J., Gelman, R., & Gallistel, C.R. (1999). Non-verbal counting in humans: The psychophysics of number representation. *Psychological Science*, 10, 130-137.

(RECEIVED 5/22/00; REVISION ACCEPTED 9/8/00)

Commentary

SUBTRACTING PIGEONS: Logarithmic or Linear?

Stanislas Dehaene

Institut National de la Santé et de la Recherche Médicale, Commissariat à l'Energie Atomique, Orsay, France

Brannon, Wusthoff, Gallistel, and Gibbon (this issue) suggest that pigeons can subtract numbers and that this implies a linear encoding of numbers in the pigeon's brain. I am sympathetic with the idea that many animals possess an elementary sense of number (Dehaene, 1997). However, although the experiment Brannon et al. conducted convincingly demonstrates that number, rather than time, was the parameter that drove the animals' behavior, the experiment fails to unambiguously demonstrate either subtraction or linear encoding.

Brannon et al. conceptualize their task as requiring mental subtraction. After seeing a variable number T of light flashes, pigeons must choose between two keys. One, called the standard key, leads to waiting for a fixed number S of additional flashes before being rewarded. The other, the number-left key, leads to waiting for $I - T$ flashes, where I is another fixed number. Because pigeons learn to select the key associated with the smaller number, S or $I - T$, Brannon et al. postulate that pigeons can compute $I - T$.

A more economical hypothesis, however, is that the pigeons merely represent the first number of flashes, T , and learn to associate each value of T with the key that leads to the fastest reward. Crucially, statistics about the associations between numbers, responses, and rewards can be compiled without subtraction, simply by noting how favorable is the outcome of pecking one key or the other. Thus redescribed, the task becomes similar to a previous experiment in which rats were rewarded for pressing a left lever when hearing two tones and a right lever when hearing eight tones (Meck & Church, 1983), a task that was simulated by a simple neural network (Dehaene & Changeux, 1993). To show that the task Brannon et al. used can also be acquired by simple association learning, I ran some simulations of a minimal network architecture capable of learning number-response associations (see Fig. 1). A schedule similar to the pigeons' was used. First came a training phase, in which activity of the output units was forced to be correct (40 blocks of 42 trials each, $S = 4$, $I = 8$). After training came several test phases during which two thirds of the trials were choice trials in which existing connections determined the network's output (Test 1, 40 blocks, $S = 4$, $I = 8$; Test 2, 25 blocks, $S = 6$, $I = 12$; Test 3, 10 blocks, $S = 4$, $I = 8$; Test 4, 10 blocks, $S = 3$, $I = 6$).

The network's performance, monitored during the last five blocks of each test period, reproduced the pigeons' performance reported by Brannon et al. The probability of choosing the number-left key varied systematically as a cumulative normal function of the input number T . Moreover, increasing or decreasing the values of I and S had a linear effect on the location of the indifference point. Brannon et al. interpret the latter property as implying a linear coding of numbers. However, in the simulation, this property was observed whether the numbers were coded linearly or logarithmically. In the linear scheme, the Gaussian distributions of activation coding for different numbers T were spaced linearly and had a linearly increasing width. In the logarithmic

scheme, the distributions were spaced logarithmically and had a fixed width. The two schemes led to similar performance (see Fig. 2). In particular, the indifference point increased linearly with S , even when the coding scheme was logarithmic.

Interestingly, the indifference point was systematically shifted in the direction of smaller numbers. An ideal observer would respond randomly when $T = I - S$ (and therefore $T = S$, because $I = 2S$ in all blocks): This is the point where the two choices S and $I - T$ are indifferent. The results of both Brannon et al. and the simulations, however, always gave a lower indifference point. This deviation seems hard to explain if pigeons compute separate estimates of S and $I - T$, but follows naturally from associative learning. Under both the linear and the logarithmic coding schemes, increasingly larger numbers are coded by increasingly overlapping distributions and hence activate increasingly overlapping sets of connections. Thus, the network generalizes asymmetrically in the direction of the larger numbers. A similar shift of the indifference point was observed in Meck and Church's (1983) experiment with rats, strengthening the hypothesis that the same association mechanisms are at work in both cases.

Brannon et al. reject associative learning because they observed a shift in the indifference point in the first block of 42 trials after the introduction of new values of S and I , a result suggesting immediate generalization and therefore an abstract understanding of task requirements. However, even those trials continued to be rewarded and hence could promote new learning. It is an empirical issue whether pigeons can adapt their behavior with only 42 rewarded trials. In the simulation, for a high-learning-speed parameter, the indifference point shifted significantly from the last five blocks of Test 1 to the first block of Test 2 (3.10 vs. 3.41, 16 simulations), $t(15) = 1.93$, one-tailed, $p = .036$, although it did not immediately reach its final value of 4.83. Such a partial shift, also seen in the actual data, again seems to support an associative learning process more than a logical arithmetic computation. In order to settle this issue, a classical test of generalization using unrewarded trials seems compulsory.

Elementary arithmetic computations have been demonstrated in monkeys (Hauser, Carey, & Hauser, 2000; Hauser, MacNeilage, & Ware, 1996). It thus seems possible that subtraction may eventually be demonstrated in pigeons. However, would this prove that numbers are encoded linearly? Brannon et al. presume that if numbers are encoded logarithmically, subjects would not be able to subtract normally, but would only manage to subtract the logarithms of numbers. This does not seem to be a valid inference. Subtraction can certainly be implemented whether the numbers are coded on linear or logarithmic number lines. As I have shown here, multiple neural implementations may lead to the same function. In particular, linear coding with scalar variability and logarithmic coding with fixed variability lead to the same metric of number similarity, and therefore to the same behavior. More generally, just as examination of a computer's performance cannot tell us how the computer encodes numbers in its microchips, behavioral evidence cannot reveal how representations are implemented, but merely shows what their functional relations are.

Address correspondence to Stanislas Dehaene, INSERM Unit 334, Service Hospitalier Frédéric Joliot, CEA/DRM/DSV, 4 Place du Général Leclerc, 91401 Orsay, France; e-mail: dehaene@shfj.cea.fr.

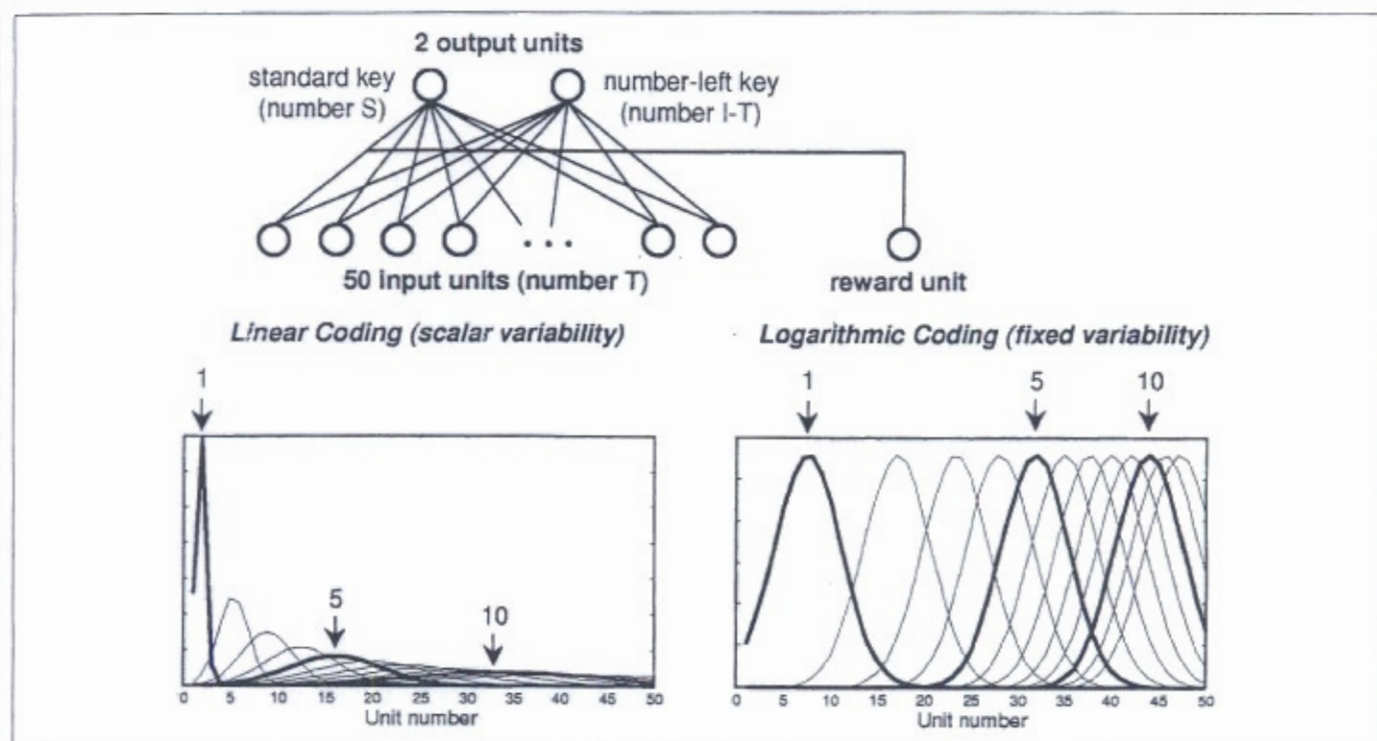


Fig. 1. Network structure and coding schemes. The illustration at the top shows the network structure used in the simulation. Fifty input units, coding for the input number T , are connected directly to two output units coding for the two possible response choices, S and $I - T$. Each output unit computes the sum of its inputs, with additional uniformly distributed noise. The output unit that has the greater sum is considered to be the network's response to the number T (its activation is then set to 1). After responding, the network receives a reward that, in the present case, is an exponentially decreasing function of the number associated with the selected key, either S or $I - T$. The connection weights are modified after each trial according to a Hebbian rule modulated by reward (as in Dehaene & Changeux, 1993). The graphs show the two number-coding schemes that were compared. In linear coding (left), each input number T is represented by a distribution of activation whose location and width are both proportional to T , with the width-to-location ratio set to a fixed Weber fraction (here arbitrarily chosen to be 0.30). In logarithmic coding (right), each input number T is represented by a distribution with a fixed width, but whose central location on the number map is a logarithmic function of T . Note that in the two cases, the metric of number similarity, as measured by the amount of overlap between the distributions coding for two numbers, is essentially the same.

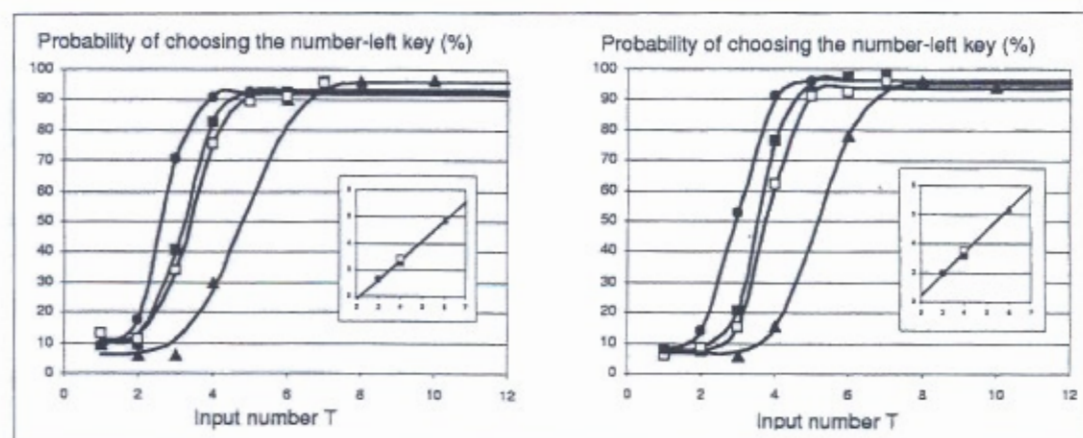


Fig. 2. Results of the simulations using linear (left) and logarithmic (right) coding. The probability of choosing the number-left key is shown as a function of the input number T (filled squares: Test 1, $S = 4$, $I = 8$; filled triangles: Test 2, $S = 6$, $I = 12$; open squares: Test 3, $S = 4$, $I = 8$; filled circles: Test 4, $S = 3$, $I = 6$). Insets show the linear relation between the indifference point and the standard number, S .

REFERENCES

- Dehaene, S. (1997). *The number sense*. New York: Oxford University Press.
- Dehaene, S., & Changeux, J.P. (1993). Development of elementary numerical abilities: A neuronal model. *Journal of Cognitive Neuroscience*, 5, 390-407.
- Hauser, M.D., Carey, S., & Hauser, L.B. (2000). Spontaneous number representation in semi-free-ranging rhesus monkeys. *Proceedings of the Royal Society of London B*, 267, 829-833.
- Hauser, M.D., MacNeilage, P., & Ware, M. (1996). Numerical representations in primates. *Proceedings of the National Academy of Sciences, USA*, 93, 1514-1517.
- Meck, W.H., & Church, R.M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, 9, 320-334.

(RECEIVED 9/8/00; ACCEPTED 9/8/00)

Commentary

RESPONSE TO DEHAENE

C.R. Gallistel,¹ Elizabeth M. Brannon,² John Gibbon,^{3,4} and Courtney J. Wusthoff³¹Department of Psychology, University of California, Los Angeles; ²Center for Cognitive Neuroscience and Department of Psychology: Experimental, Duke University; ³Department of Psychology, Columbia University; and⁴Department of Biopsychology, New York State Psychiatric Institute

Dehaene's model avoids the assumption of numerical processing only by putting it in the tutor instead of in the net. The attempt is ironic, because the units in the net process their inputs arithmetically. Furthermore, treating the different values of T as unrelated stimuli leads to a radically nonparsimonious model of what must be learned.

In Dehaene's simulation, the value of a reward is multiplied by an exponentially decaying function of the number of flashes that the subject had to generate after its choice and before obtaining its reward. The computation of the discount factor and its scaling of the value of the reward is carried out by the tutor, but if such processes are causally relevant to the behavior observed, then they have to be located in the brain of the subject. Absent that brain, the effect of the flashes on the value of a food reward is moot. It is not more parsimonious to assume that the brain negates the flash count, exponentiates the negation, and then multiplies the base value of a reward by the result than it is to assume that the brain subtracts. Implementing Dehaene's model in a multiunit neural net is still less parsimonious. Because every unit in a net is assumed to sum signed inputs, only one such unit is required to implement the subtraction that we assume.

Finally, in Dehaene's model, the response to each number of T flashes is an independently acquired habit. Nets are universal function approximators (Hornik, Stinchcombe, & White, 1989), which means they will learn an arbitrary mapping from inputs to outputs. Dehaene's net will learn, for example, to choose the standard key after an odd number of flashes and the number-left key after an even number. We know that this is not how numbers function for animal subjects. (See Honig, 1993, for a demonstration that pigeons have great difficulty

with odd-even discriminations and Brannon & Terrace, 2000, for a demonstration that monkeys cannot learn to touch arrays of different numerosities in arbitrary nonmonotonic orders but readily learn to touch them in a monotonic numerical order.) Thus, in Dehaene's model, the more numbers there are, the more numerous are the "habits" (S-R relations) that must be acquired. When the values of I and S are increased, the net must learn a different mapping for each of the T values it had previously mastered plus new mappings for the newly introduced values of T . Our subjects' rapid adjustments to new values of I and S are not consistent with the assumption that making this adjustment involves learning a large number of new discriminations. We would welcome, as Dehaene suggests, future demonstrations of numerical subtraction in our number-left paradigm with unrewarded trials; however, given the data available, we view Dehaene's associative explanation unparsimonious and unlikely.

REFERENCES

- 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.
- Honig, W.K. (1993). Numerosity as a dimension of stimulus control. In S.T. Boysen & E.J. Capaldi (Eds.), *The development of numerical competence: Animal and human models* (pp. 61-86). Hillsdale, NJ: Erlbaum.
- Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2, 359-366.

(RECEIVED 11/7/00; ACCEPTED 11/7/00)

Address correspondence to Randy Gallistel, Rutgers Center for Cognitive Science and Department of Psychology, Rutgers University, 152 Frelinghuysen Rd., Piscataway, NJ 08854-8020; e-mail: galliste@rucss.rutgers.edu.