

Research Article

Shared System for Ordering Small and Large Numbers in Monkeys and Humans

Jessica F. Cantlon and Elizabeth M. Brannon

Department of Psychological and Brain Sciences and Center for Cognitive Neuroscience, Duke University

ABSTRACT—*There is increasing evidence that animals share with adult humans and perhaps human infants a system for representing objective number as psychological magnitudes that are an analogue of the quantities they represent. Here we show that rhesus monkeys can extend a numerical rule learned with the values 1 through 9 to the values 10, 15, 20, and 30, which suggests that there is no upper limit on a monkey's numerical capacity. Instead, throughout the numerical range tested, both accuracy and latency in ordering two numerical values were systematically controlled by the ratio of the values compared. In a second experiment, we directly compared humans' and monkeys' performance in the same ordinal comparison task. The qualitative and quantitative similarity in their performance provides the strongest evidence to date of a single nonverbal, evolutionarily primitive mechanism for representing and comparing numerical values.*

A strong argument has been made that animals and adult humans share a nonverbal system for representing number as analogue magnitudes (e.g., Brannon, 2005; Brannon & Terrace, 1998, 2000; Cantlon & Brannon, 2005; Feigenson, Dehaene, & Spelke, 2004; Gelman & Gallistel, 2004; Nieder, Freedman, & Miller, 2002; Nieder & Miller, 2003). Although it is undeniable that adult humans can represent and manipulate arbitrary symbols of number in a fashion that far exceeds anything that is seen in nonhuman animals (e.g., predicting the national debt in 2010), there is now considerable evidence that adult humans also possess a nonverbal system for representing number as mental magnitudes. For example, humans show effects of nu-

merical distance (e.g., 1 vs. 9 easier than 1 vs. 2) and magnitude (e.g., 1 vs. 2 easier than 8 vs. 9) when comparing single-digit numbers, double-digit numbers, or random-dot patterns (e.g., Buckley & Gilman, 1974; Dehaene, 1992; Dehaene, Dupoux, & Mehler, 1990; Moyer & Landauer, 1967). Thus, Weber's law describes discriminability between two numbers in the same way that it describes discriminability between values on perceptual continua such as line length, brightness, and weight (e.g., Welford, 1960). Furthermore, humans and nonhuman animals show the same ratio-dependent response signatures in numerical production tasks (Cordes, Gelman, & Gallistel, 2001; Whalen, Gallistel, & Gelman, 1999) and numerical comparison tasks (e.g., Brannon & Terrace, 2002; Pica, Lemer, Izard, & Dehaene, 2004).

A consequence of Weber's law is that the ability to discriminate two values depends on their ratio rather than their absolute values. Studies of both humans and animals have supported the notion that ratio is a critical feature for the successful discrimination of numerosity. However, to test ratio dependence it is necessary to examine performance on a wide range of numerical values. Although this has been done for adult humans' numerical judgments (e.g., Barth, Kanwisher, & Spelke, 2003; Piazza, Izard, Pinel, LeBihan, & Dehaene, 2004; Pica et al., 2004), the vast majority of tests of animals' numerical ability have been limited to values under 10. Furthermore, given a large literature arguing for qualitative differences in the ways humans and animals represent small and large numerical values (e.g., Feigenson, Carey, & Hauser, 2002; Hauser, Carey, & Hauser, 2000; Mandler & Shebo, 1982), it seems critical to test Weber's law for large numerical values.

In the experiments we report here, we tested the ability of rhesus monkeys to apply an ordinal numerical rule (i.e., choose smaller value first) learned with the values 1 through 9 to the larger values 10, 15, 20, and 30. Given previous evidence that monkeys can apply an ordinal numerical rule learned with one

Address correspondence to Jessica Cantlon or Elizabeth Brannon, Center for Cognitive Neuroscience, Duke University, Box 90999, Durham, NC 27708-0999, e-mail: jfc2@duke.edu or brannon@duke.edu.

range of values (1–4) to a novel set of numerical values outside the training range (5–9; Brannon & Terrace, 1998, 2000), we predicted that the monkeys would succeed in applying an ordinal numerical rule to these larger numerical values. In a second experiment, we quantitatively and qualitatively compared the performance of monkeys and adult humans by testing them on their ability to order pairs of the numerosities 2 through 30.

EXPERIMENT 1

Method

Participants and Apparatus

The subjects were 2 socially housed adult female rhesus macaques (F and M). Both monkeys were kept on water-restricted diets to increase their motivation for juice reinforcement. All experimental procedures were approved by an Institutional Animal Care and Use Committee. The subjects were tested in sound-attenuated booths while seated in Plexiglas primate chairs fitted with a juice-delivery system. Stimuli were presented on a 17-in. touch-screen computer monitor fixed to the inside wall of the soundproof booth. A custom-built program written in RealBasic presented the stimuli and registered responses. Stimuli were presented on a 2×2 touch-screen matrix, with their spatial positions on the screen randomly selected from the four possible ports.

Task and Procedure

Both monkeys were first trained to order all possible pairings of the numerical values from 1 through 9. Subsequently, the novel values 10, 15, 20, and 30 were introduced, creating three categories of pairs (“novel” refers to the values 10, 15, 20, and 30, and “familiar” refers to the values 1 through 9): novel-novel ($n = 6$), familiar-novel ($n = 36$), and familiar-familiar ($n = 36$). No differential reinforcement for correct versus incorrect responses was provided on trials that contained a novel numerical value, so that we could objectively assess whether the monkeys would extrapolate a rule learned with the values 1 through 9 to the novel large values 10, 15, 20, and 30.

Subjects initiated each trial by pressing a small black square presented in the bottom left corner of the screen. They were then required to press each of two stimuli in ascending numerical order. Correct responses were followed by a 1-s chime and pink screen, correctly completed sequences were rewarded with a 0.3-ml squirt of Kool-Aid, and incorrect responses were followed by a 3-s tone and black screen. The intertrial interval was a variable 3 to 5 s.

Over approximately 100 sessions (of about 500 trials each), the subjects were trained to order the numerosities 1 through 9. The subjects were then tested for 10 sessions (approximately 3,000 trials total) with all new stimuli that included the novel numerosities 10, 15, 20, and 30. During testing, 75% of the

stimuli were new exemplars of the 36 pairs of the numerosities 1 through 9 (familiar-familiar pairs), and responses were followed by positive or negative feedback. On 17% of the trials, one of the novel numerical values (10, 15, 20, and 30) was paired with one of the familiar values from 1 through 9 (familiar-novel pairs). The remaining 8% of the trials consisted of paired presentations of the novel numerical values 10, 15, 20, and 30 (novel-novel pairs). When a pair that contained a novel value was presented, positive feedback and a juice reward were given regardless of the order in which the subject responded.¹ Nondifferential reinforcement ensured that the monkeys could not learn about the relationships among novel values during the experiment and thus allowed us to assess the monkeys’ unbiased initial response tendencies.

Stimuli

The training stimuli contained 1 through 9 square-shaped elements of various sizes and colors, whereas the test stimuli contained 1 through 9, 10, 15, 20, or 30 elements (see Fig. 1a). Within a stimulus, each element was randomly placed on a yellow background that varied in size. The surface area of the background was varied such that the larger numerosity was more dense than the smaller numerosity on 50% of the trials, the smaller numerosity was more dense on 30% of the trials, and density was equated between the two numerosities on the remaining 20% of the trials. To ensure that the density control did not lead to the use of stimulus size as a cue, we presented the smaller numerosity on the larger stimulus background on 30% of the trials. The cumulative surface area of the array elements ranged 12-fold for each numerosity so that the larger numerosity had the larger cumulative surface on 45% of the trials and the smaller cumulative surface on 45% of the trials; on the remaining 10% of the trials, cumulative surface area was equated for the two numerosities. The larger numerosity had the larger and smaller cumulative perimeter of the array elements on 76% and 22% of the trials, respectively. The cumulative perimeter of the array elements was equated on 2% of the trials.

Results and Discussion

Both monkeys rapidly learned to order the values 1 through 9. Overall accuracy was 82% for both monkeys at the end of training. Performance was unaffected by the controls for density, surface area, and perimeter. Accuracy was well above chance (binomial tests) when surface area was equated (Monkey F: $M = 83%$, $p < .05$; Monkey M: $M = 86%$, $p < .05$), when density was equated (Monkey F: $M = 82%$, $p < .05$; Monkey M: $M = 85%$, $p < .05$), and when perimeter was equated (Monkey F: $M = 89%$, $p < .05$; Monkey M: $M = 80%$, $p < .05$). The important new result

¹Brannon and Terrace (1998, 2000) gave no reinforcement for these trials; in contrast, we gave full reinforcement. The important commonality is that neither procedure provided differential feedback. Our motivation for switching to full reinforcement was to avoid extinction while obtaining a larger number of test trials.

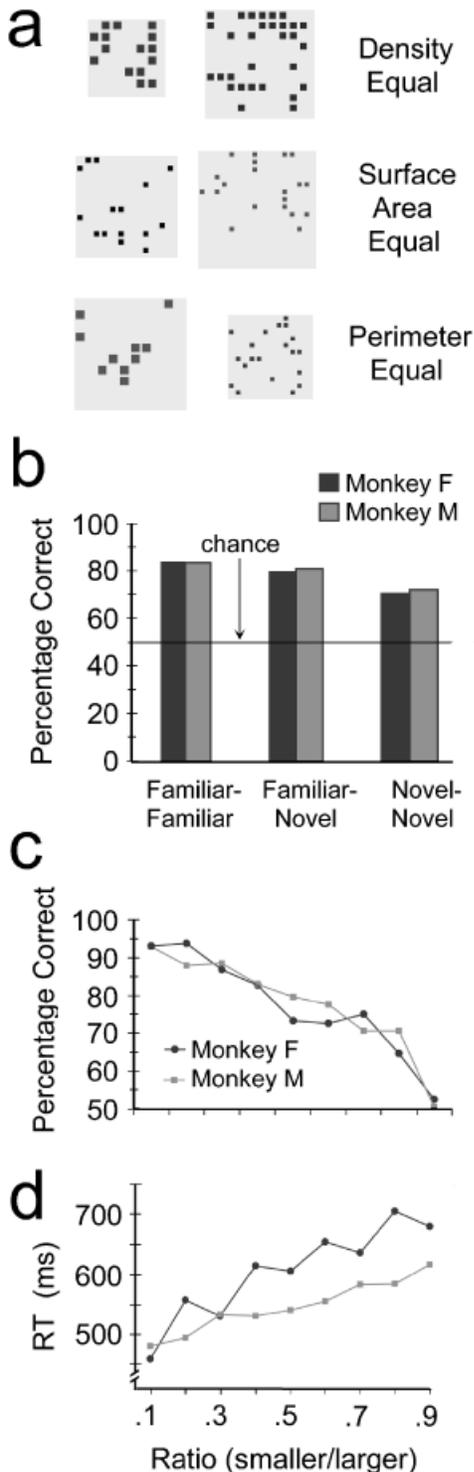


Fig. 1. Examples of the stimuli (a) and results (b–d) from Experiment 1. Density, cumulative surface area, and perimeter were sometimes larger and sometimes smaller for the larger numerosity relative to the smaller numerosity; density, cumulative surface area, and perimeter were also equated for the larger and smaller numerosities on a small proportion of trials. The graph in (b) presents accuracy for monkeys on familiar-familiar pairs, familiar-novel pairs, and novel-novel pairs (“familiar” refers to the values 1–9, and “novel” refers to the specific values 10, 15, 20, and 30). Chance performance was 50%. The lower graphs plot accuracy (c) and reaction time (RT; d) for each monkey as a function of the ratio of the numerical values in a pair.

from this study is that both monkeys responded significantly above chance to pairs of the novel large values under conditions of nondifferential reinforcement (Fig. 1b; single-sample t tests), $t(5) = 6.06, p < .01, \eta^2 = .38$, for Monkey F and $t(5) = 4.40, p < .01, \eta^2 = .79$, for Monkey M. After training to respond to pairs of the values 1 through 9 in ascending numerical order, the monkeys inferred that they should choose the smaller and then the larger value when faced with pairs of the large values 10, 15, 20, and 30. These data demonstrate that monkeys can represent and compare large values without laboratory training and that the ordinal numerical continuum a monkey represents is unlikely to have an upper limit.

Figures 1c and 1d show that accuracy and latency of responding were systematically related to the ratio between the numerical values being compared. Accuracy decreased as the ratio between the magnitudes of the stimuli increased (Monkey F: $R^2 = .91, p < .01$; Monkey M: $R^2 = .88, p < .01$), and latency increased as the ratio increased (Monkey F: $R^2 = .86, p < .01$; Monkey M: $R^2 = .96, p < .01$). Thus, the limitation on a monkey’s numerical discrimination capacity is imposed by the ratio of the values compared rather than absolute set size.

To more systematically test whether ratio dependence extends to a wide range of numerical values, and to explore the parallels between monkeys and humans, we next directly compared performance of adult humans and monkeys on a numerical comparison task.

EXPERIMENT 2

Method

Participants and Apparatus

Two monkeys (F and M from Experiment 1) and 11 Duke University students (6 male and 5 female) participated in this experiment. The apparatus used for monkeys was as described for Experiment 1. Human subjects responded via a computer mouse rather than a touch screen.

Task and Procedure

The general task design was the same as that described for Experiment 1; however, monkeys did not receive a training phase, but were only tested, in five 350-trial sessions. Human subjects were tested in a single 40-min session of 550 trials without prior training; they were verbally instructed to choose the stimulus with the smaller number of elements on each trial. Both monkeys and humans were rewarded for selecting the numerically smaller and then the larger of two arrays that were presented simultaneously on a computer screen. Correct responses were followed by positive visual and auditory feedback (and Kool-Aid for monkeys); incorrect responses were followed by a 3- to 5-s time-out. Human subjects were instructed not to count and to respond to the stimuli as quickly as possible once they appeared on the screen.

Stimuli

All possible numerical pairs of the even values 2 through 30 (105 pairs total) were presented with equal frequency. As in Experiment 1, stimuli consisted of arrays of square elements on a yellow background. The stimulus controls for Experiment 2 were as described for Experiment 1.

Results and Discussion

Figures 2a and 2b show accuracy and latency for humans and monkeys as a function of the ratio (smaller/larger) of the numerosities being compared. For both monkeys and humans, accuracy and latency were controlled by this ratio, showing the pattern predicted by Weber's law: Accuracy decreased with increasing ratio (monkeys: $R^2 = -.94, p < .01$; humans: $R^2 = -.91, p < .01$), and latency increased with increasing ratio (monkeys: $R^2 = .97, p < .01$; humans: $R^2 = .99, p < .01$). Further, accuracy at a given ratio was constant across number pairs, confirming that the ratio between numerical values, rather than objective magnitude or distance, modulated performance. Monkeys and humans exhibited no linear increase or decrease in accuracy among different numerical pairs having the ratio of 2:3 (monkeys: $R^2 = .001, p = .97$; humans: $R^2 = .01, p = .87$) or among different numerical pairs having the ratio of 1:2 (monkeys: $R^2 = .01, p = .81$; humans: $R^2 = .0001, p = .98$).

Overall, monkeys responded significantly faster than did humans (549 ms vs. 674 ms), $t(16) = 3.37, p < .01, \eta^2 = .42$. However, accuracy was similar for monkeys (80%) and humans (87%); in fact, as a function of ratio, accuracy did not differ significantly between the two species, $t(16) = 0.87, p = .40, \eta^2 = .05$. To calculate the sensitivity of each species' ability for making numerical comparisons, we estimated the internal Weber fraction, which is a constant that reflects the proportion by which two numbers must differ to be discriminable. We used a model put forth by Pica et al. (2004), which is the first model to estimate internal Weber fractions for the output of an ordinal comparison task.² The model assumes (a) that any given numerical value is represented as a Gaussian distribution, and (b) that the precision with which an organism can discriminate two values is determined by the degree to which their distributions overlap. The model estimates the degree to which the distributions of any two numerical values overlap given a specific internal Weber fraction.³ By comparing the predicted error rate at

²It is important to note that the model does not allow a determination of whether number is represented on an internal scale that is logarithmically compressed or on an internal scale that is linear with objective number but has variability that increases proportionally with magnitude.

³The following equation produces the predicted error rate ($p_{\text{comparison}}$) for comparisons of numerical values ($n1$ vs. $n2$) given an estimate of the internal Weber fraction (w ; Pica et al., 2004):

$$p_{\text{comparison}} = \int_0^{\infty} e^{-\frac{1}{2} \left(\frac{x + \text{Abs}(n1 - n2)}{w \sqrt{n1^2 + n2^2}} \right)^2} \frac{1}{\sqrt{2\pi} w \sqrt{n1^2 + n2^2}} dx = \frac{1}{2} \text{erfc} \left(\frac{\text{Abs}(n1 - n2)}{\sqrt{2} w \sqrt{n1^2 + n2^2}} \right)$$

different ratios with the observed error rates for each numerical pair, one can determine the internal Weber fraction that produces the best goodness of fit (r^2).

The estimates of the internal Weber fractions that produced the best goodness of fit (r^2) for the data were .38 for monkeys and .26 for humans. The fitted lines in Figure 2c are composed from the predicted values produced by the model. Although the estimated internal Weber fractions suggest that monkeys were less sensitive to numerical differences than were adult humans, this difference should be viewed cautiously. The monkeys spontaneously responded an average of more than 100 ms faster than the humans even though the humans were instructed to respond as quickly as possible; thus, the enhanced accuracy for humans may reflect a speed-accuracy trade-off. In addition, as suggested by the graph in Figure 2d, the difference in accuracy between monkeys and humans as a function of ratio was not statistically different from the difference in accuracy between the two most divergent humans, $t(16) = 1.55, p = .14, \eta^2 = .13$. In fact, the difference in accuracy between monkeys and humans was smaller than the difference between the most accurate and least accurate humans. The sensitivity of monkeys on this task was therefore within the range of sensitivity exhibited by human subjects.

One possible explanation for the similarity in numerical sensitivity between the monkeys and humans is that unlike the human subjects, the monkeys received extensive laboratory training on this task during Experiment 1. This training could have heightened numerical sensitivity in the monkeys relative to the humans. However, we found no increasing trend in accuracy over the eighteen 100-trial blocks (the 18th block contained only 50 trials) for monkeys, $R^2 = .03, p = .47$, or the eleven 50-trial blocks for humans, $R^2 = .13, p = .10$: Accuracy was 80% and 81% for monkeys and 86% and 88% for humans for the first and second halves of testing blocks, respectively. Asymptotic performance appears to have been achieved relatively early in the testing period for both subject groups, which suggests that performance differences are not attributable to differing degrees of familiarity with the task or stimuli. Given that the humans reached asymptotic performance early in the experimental session without extensive training, it is important to note that the humans and monkeys may have differed in their preexperimental experience comparing large numerosities. Overall, our results indicate that monkeys' performance was not only qualitatively similar to humans' performance, but also quantitatively comparable to humans' performance when humans are instructed to respond quickly.

GENERAL DISCUSSION

In Experiment 1, monkeys trained to order pairs of the numerosities 1 through 9 (by responding to the smaller numerosity first and the larger numerosity second) spontaneously ordered pairs of the values 10, 15, 20, and 30 in the same ordinal

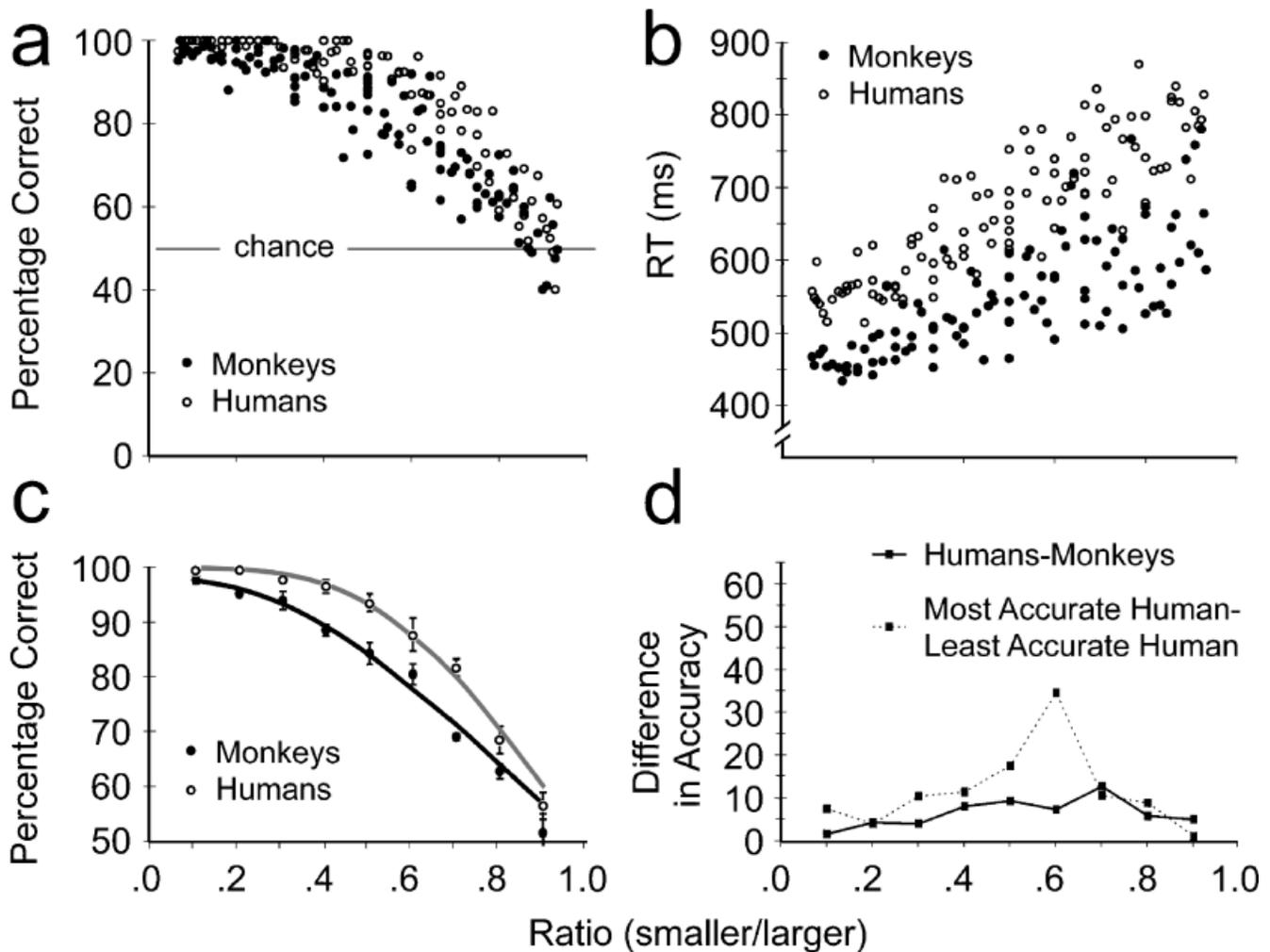


Fig. 2. Results from Experiment 2. The upper graphs plot accuracy (a) and reaction time (RT; b) as a function of the ratio of the numerical values in a pair for monkeys and humans (all possible pairs of the even values 2 through 30 were tested). The graph in (c) shows accuracy for monkeys and humans as a function of the ratio of the numerical values being compared. Dark lines show the error rate predicted by the model described in Pica, Lemer, Izard, and Dehaene (2004). The graph in (d) presents the difference in accuracy between monkeys and humans and between the most accurate and least accurate human participant as a function of ratio.

direction without any laboratory experience with these values. These results confirm that monkeys are not limited to making comparisons among familiar numerical values and further suggest that there is no upper limit on the numerical values that monkeys can represent.

In Experiment 2, monkeys performed a numerical comparison task in a way that was both qualitatively and quantitatively similar to the way adult humans performed the task. For both groups, accuracy and latency were controlled by the ratio of the values being compared across the wide range of values from 2 through 30.

For monkeys, the estimate of sensitivity to numerical differences obtained from the current data is quite similar to a previous estimate using a same/different task and a smaller numerical range (for numerosities 1–6, discriminability ratio = .35; Nieder & Miller, 2003). In contrast, greater sensitivity has been found for adults in some previous tasks. For example, in a recent

study using a limited set of numerical values within the range 10 through 80 (12 pairs at four ratios), Pica et al. (2004) estimated the internal Weber fraction to be .17 for French-speaking adults. An important difference between their study and ours is that we gave instructions emphasizing speed and used finer increments in numerical values, thereby creating much more difficult discriminations (e.g., 26 vs. 28). One or both of these factors could have resulted in the larger estimate of the internal Weber fraction for adult humans in our study.

The difference in estimates of the internal Weber fraction obtained in different studies raises the interesting question of the role of experience in shaping sensitivity to numerical differences. If experience has an impact on the internal Weber fraction, then we would predict that the estimated internal Weber fraction would decrease over development and that within the laboratory, extended training could increase a monkey's sensitivity. In fact, between 6 and 9 months of age, human

infants progress from discriminating a ratio of 1:2 (e.g., 8 vs. 16) to discriminating values with a ratio of 2:3 (e.g., 8 vs. 12; Lipton & Spelke, 2003). Thus, experience appears to be a critical factor for refining numerical sensitivity early in development. It may also be the case that experience shapes numerical sensitivity over the entire life span.

In conclusion, our findings suggest that monkeys represent values as large as 30 as mental magnitudes and rely on a mental comparison process that is controlled by Weber's law. Although the majority of previous studies have employed numerical values under 10, there is no theoretical reason to do so, and in fact it is unlikely that there is any upper limit on monkeys' ability to represent numerical values from visual arrays of elements. The fact that monkeys easily extended a rule learned with the values 1 through 9 to the values 10, 15, 20, and 30 suggests that monkeys do not treat values under and over 10 qualitatively differently. Finally, by testing monkeys and humans in the same task with the same stimuli on a wide range of numerical values, we have provided the strongest evidence to date in favor of a nonverbal system, shared by animals and humans, that allows both representation and comparison of numerosities and that is controlled by Weber's law.

Acknowledgments—We thank Kerry Jordan, Evan MacLean, Pierre Rojas, Kelley Safford, and members of the Brannon lab for help with data collection and discussion of this work. We also thank Stanislas Dehaene for helpful discussion of the data presented in this manuscript. This work was supported by a National Science Foundation Graduate Student Fellowship to J.F.C. and by a John Merck Fund Fellowship, National Institute of Child Health and Human Development Grant R01 (HD49912), and a National Science Foundation CAREER Award to E.M.B.

REFERENCES

- Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number representations in adults. *Cognition*, *86*, 201–221.
- Brannon, E.M. (2005). The independence of language and mathematical reasoning. *Proceedings of the National Academy of Sciences, USA*, *102*, 3177–3178.
- 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.
- Brannon, E.M., & Terrace, H.S. (2002). The evolution and ontogeny of ordinal numerical ability. In M. Bekoff, C. Allen, & G. Burghardt (Eds.), *The cognitive animal: Empirical and theoretical perspectives on animal cognition* (pp. 197–204). Cambridge, MA: MIT Press.
- Buckley, P.B., & Gillman, C.B. (1974). Comparisons of digits and dot patterns. *Journal of Experimental Psychology*, *103*, 1131–1136.
- Cantlon, J.F., & Brannon, E.M. (2005). Semantic congruity affects numerical judgments similarly in monkeys and humans. *Proceedings of the National Academy of Sciences, USA*, *102*, 16507–16511.
- Cordes, S., Gelman, R., & Gallistel, C.R. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychological Bulletin and Review*, *8*, 698–707.
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*, 1–42.
- Dehaene, S., Dupoux, E., & Mehler, L. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 626–641.
- Feigenson, L., Carey, S., & Hauser, M.D. (2002). The representations underlying infants' choice of more: Object files versus analog magnitudes. *Psychological Science*, *13*, 150–156.
- Feigenson, L., Dehaene, S., & Spelke, E.S. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*, 307–314.
- Gelman, R., & Gallistel, C.R. (2004). Language and the origin of numerical concepts. *Science*, *306*, 441–443.
- Hauser, M.D., Carey, S., & Hauser, L.B. (2000). Spontaneous number representation in semi-free-ranging rhesus monkeys. *Proceedings of the Royal Society of London: Biological Sciences*, *267*, 829–833.
- Lipton, J.S., & Spelke, E.S. (2003). Origins of number sense: Large-number discrimination in human infants. *Psychological Science*, *14*, 396–401.
- Mandler, G., & Shebo, B.J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General*, *111*, 1–22.
- Moyer, R.S., & Landauer, T.K. (1967). Time required for judgments of numerical inequality. *Nature*, *215*, 1519–1520.
- Nieder, A., Freedman, D.J., & Miller, E.K. (2002). Representation of the quantity of visual items in the primate prefrontal cortex. *Science*, *297*, 1708–1711.
- Nieder, A., & Miller, E.K. (2003). Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron*, *34*, 149–157.
- Piazza, M., Izard, V., Pinel, P., LeBihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human parietal cortex. *Neuron*, *44*, 547–555.
- Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian Indigene Group. *Science*, *306*, 499–503.
- Welford, A.T. (1960). The measurement of sensory-motor performance: Survey and re-appraisal of twelve years' progress. *Ergonomics*, *3*, 189–230.
- Whalen, J., Gallistel, C.R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science*, *10*, 130–137.

(RECEIVED 3/22/05; REVISION ACCEPTED 8/1/05;
FINAL MATERIALS RECEIVED 8/12/05)