Cognition, 44 (1992) 43–74

Preverbal and verbal counting and computation*

C.R. Gallistel and Rochel Gelman

Department of Psychology, University of California–Los Angeles, Los Angeles, CA 90024-1563, USA

Abstract

Gallistel, C.R., and Gelman, R., 1992. Preverbal and verbal counting and computation. Cognition, 44: 43–74.

We describe the preverbal system of counting and arithmetic reasoning revealed by experiments on numerical representations in animals. In this system, numerosities are represented by magnitudes, which are rapidly but inaccurately generated by the Meck and Church (1983) preverbal counting mechanism. We suggest the following. (1) The preverbal counting mechanism is the source of the implicit principles that guide the acquisition of verbal counting. (2) The preverbal system of arithmetic computation provides the framework for the assimilation of the verbal system. (3) Learning to count involves, in part, learning a mapping from the preverbal numerical magnitudes to the verbal and written number symbols and the inverse mappings from these symbols to the preverbal magnitudes. (4) Subitizing is the use of the preverbal counting process and the mapping from the resulting magnitudes to number words in order to generate rapidly the number words for small numerosities. (5) The retrieval of the number facts, which plays a central role in verbal computation, is mediated via the inverse mappings from verbal and written numbers to the preverbal magnitudes and the use of these magnitudes to find the appropriate cells in tabular arrangements of the answers. (6) This model of the fact retrieval process accounts for the salient features of the reaction time differences and error patterns revealed by experiments on mental arithmetic. (7) The application of verbal and written computational algorithms goes on in parallel with, and is to some extent guided by, preverbal computations, both in the child and in the adult.

Correspondence to: C.R. Gallistel or R. Gelman, Department of Psychology, University of California–Los Angeles, Los Angeles, CA 90024-1563, USA.

*The preparation of this paper was supported by NSF Grant BNS-8916220 to R. Gelman.

0010-0277/92/$10.10

Introduction

The gulf between the experimental and theoretical analysis of human cognition on the one side and the experimental and theoretical analysis of animal cognition on the other is not consistent with an evolutionary perspective. The foundations of human preverbal cognition presumably lie for the most part in animal cognition, that is, in mental processes that were already present in the remote non-human ancestors from which we and pylogenetically rather distant genera like rodents and birds both descend. Yet, there are few attempts to link the extensive literature on animal cognition to human cognition (see, however, Diamond & Goldman-Rakic, 1989; Premack & Woodruff, 1978). Cognitive processes for dealing with numerosities and magnitudes have been extensively studied in both the common laboratory animals (rats and pigeons) and in man. Thus, in this domain, there is an experimental basis for a rapprochement. In this paper, we suggest that the foundations of human numerical competence lie in preverbal mechanisms for counting and arithmetic reasoning that we share with genera at least as distant as rodents and birds.

Numerical competence in animals

The category–concept distinction

In discussing numerical competence, it is useful to distinguish between processes such as counting that map from numerosities to mental representatives of numerosities (symbols) and processes such as mental addition that operate on the mental representatives of numerosity. Gelman and Gallistel (1978) called this the estimator–operator distinction. Estimator processes, such as counting, produce the mental representatives of numerosity (termed numerons by Gelman & Gallistel, 1978). They determine the mapping or reference relations between mental entities (numerons) and the numerosities to which they refer. By contrast, operator processes, such as mental addition, process one numeron (unary operators) or two (binary operators) numerons to produce another numeron.

The estimator–operator distinction corresponds to the distinction between numerons as categories and numerons as concepts (Gallistel, in press). While category and concept have often been treated as synonyms, we believe that there are good reasons to distinguish between them, at least in the number domain. A numeron qua category refers to all sets of a given numerosity. By contrast, a numeron qua concept plays a unique role in a system of mental operations isomorphic to at least some of the arithmetic operations. The numerical concept "three" is defined by the role it plays in the mental operations isomorphic to the operations of arithmetic, not by what it refers to, just as a bishop in chess is defined by its role in the game, not by its ecclesiastical referent. "The smallest prime number greater than one" uniquely identifies "three" without saying anything about the referent of "three," just as "the piece that moves along the diagonals" uniquely identifies the bishop without reference to anything external to the game itself.

It may help to emphasize the importance of this distinction to note that the history of mathematics has been driven in no small measure by the tension between the conceptual and the categorical use of numbers. Conceptual manipulations such as subtracting a bigger from a smaller number or taking the square root of a negative number produced symbols with no apparent referent, symbols like "-2" and "$\sqrt{-1}$." It was initially thought that these symbols could not stand for anything outside the symbol system. There were no categories of non-symbolic entities (e.g., numerosities or magnitudes) to which these symbols could refer. Hence, the manipulations that produced them were either forbidden or only allowed as intermediate steps on the way to an expression that did not include such "nonsensical" numbers. In time, however, mappings from these symbols to non-numerical entities such as bank balances, directed magnitudes, and points in the plane were discerned. The discovery or invention of categories of things to which these symbols could refer brought these symbols and the conceptual operations that produced them out of the mathematical closet. This history also led mathematicians to realize the importance of distinguishing between numbers as concepts and numbers as categories.

Animals may be said to have number categories insofar as they can be shown to base their behavior on the numerosity of a set independent of its other attributes. Such an ability suggests that they map from all instances of a given numerosity to a mental representative of that category of sets, a numeron that represents that category. This numeron makes it possible for the animal to respond in the same way to all instances of a given numerosity, even to instances never before encountered. By contrast, animals may be said to have a concept of number insofar as they may be shown to mentally manipulate numerons in processes that are isomorphic to some or all of the operations that define the system of arithmetic: ordering, addition, subtraction, multiplication, and division.

Analog isomorphisms

The things we are most apt to think of as representatives of numerosity are sound patterns like /three/ or ink patterns like "3" or the patterns of *on* and *off* states in the bit registers of a digital computing device. The physical characteristics of these symbols are not generally such that simple physical operations performed on the symbols themselves can be isomorphic to arithmetic operations. No simple physical operation with the ink pattern "3" and the ink pattern "2" yields the ink pattern "5" – the pattern that represents the combined numerosity of the two numerosities to which "3" and "2" refer. Thus, these symbols do not readily or transparently become entities in a physical system whose operations are isomorphic to the arithmetic operations. In this sense, the numerical symbols we are most familiar with are arbitrarily related to their referents. But symbols need not be as arbitrarily related to what they symbolize as are the bit patterns in digital computation, let alone the ink patterns for the arabic numerals. In analog computation, the symbols, which are generally magnitudes such as currents or voltages, are chosen because they readily enter into physical processes isomorphic to the arithmetic operations.

A histogram is a familiar example of the use of magnitudes to represent numerosities: the higher the column in a histogram, the greater the numerosity of the set represented by that column. Like all magnitudes, these magnitudes are readily incorporated into a system of operations isomorphic to the system of arithmetic. The column that represents the combined numerosity of sets 1 and 2 is the column you get by placing the column for set 1 on top of the column for set 2. The column that represents the more numerous of two sets is the first column contacted by a horizontal line lowered from the top of the graph. If you form a rectangle whose height is that of column 1 and whose width is the height of column 2, then hold constant the area of the rectangle while adjusting its width to the standard column width, you get a column whose height represents the numerosity of the set formed by multiplying the numerosities represented by columns 1 and 2.

The system just described – histogram arithmetic or the histogrammic calculator – is an analog system isomorphic to arithmetic. Its symbols are magnitudes, the heights of the columns. Its operations are processes involving those magnitudes – processes chosen to be isomorphic to the arithmetic operations. It and other analog computers are just as much symbol-processing devices as are the more familiar digital computers. We want to suggest, on the basis of both animal and the human data, that the preverbal processes that underlie both the animal and the human capacity to represent numerosities and reason arithmetically are analogous to histogram arithmetic. These preverbal processes generate analog mental variables (ultimately, of course, neurophysiological variables) that function as the mental/neural representatives of numerosity (numerons). These analog variables participate in mental/neural processes that were chosen, we believe, via natural selection during a remote epoch in evolutionary history because their isomorphism to arithmetic operations conferred decisive adaptive advantages on their possessors. Arithmetic is fundamental to the scientific description of the word in which animals attempt to survive and reproduce. It seems reasonable to suppose that the ability to create similar descriptions of the world within the organ that governs animal actions would confer substantial competitive advantages on the possessors of those abilities. The question is, do the experimental data support such a supposition?

Animals categorize sets on the basis of their numerosity

Number discrimination

Several different experimental paradigms have been developed to test the ability of an animal to discriminate on the basis of the numerosity of a set. In some, the set whose numerosity is the basis of discriminative behavior is the set of responses the animal has made since the last reward. In others, it is the number of elements in a set of simultaneously or sequentially presented stimuli. In a paradigm developed by Mechner (1958), on some percentage of the trials a rat had to leave off pressing one lever (the A lever) and press another lever (the B lever) to get its reward, but only after making a required number of consecutive presses on the A lever. Premature abandonment of the A lever incurred a penalty. For a block of trials in which some number, N, of presses on the A lever was required to arm the B lever, Mechner plotted the rat's probability (relative frequency) of abandoning the A lever as a function of the number of consecutive presses it had made on that lever (Figure 1a). The resulting plots were approximately normally distributed around a number somewhat higher than the number of required presses. The difference between the modal value of the distribution and the required number of presses was a systematic function of the penalty: the greater the penalty, the larger the difference. However, for any given penalty, the difference remained constant as Mechner increased N, the required number of presses. Of course, time on the A lever covaried with the number of presses of that lever. However, in a subsequent experiment, Mechner and Guevrekian (1962) showed that doubling the rat's rate of responding by increasing their state of food deprivation had no effect on the modes of these distributions. If the abandonment of the first lever were based on the time spent on that lever, doubling the rate of responding would double the modal value of the distribution.

Platt and Johnson (1971) used a paradigm in which rats pressed a lever that silently armed a feeder after some number, N, of presses. The value of N was constant within a block of trials but varied between blocks. The armed feeder was activated when the rat broke off pressing the lever and poked its head into a feeding alcove, interrupting a photocell beam. If the rat interrupted the beam before making the requisite number of presses on the lever, there was a penalty: the response counter was reset to zero. The data in Figure 1b show the rats' probability of interrupting the beam at the entrance to the feeding alcove as a function of n, the number of presses on the lever, for various values of N, the

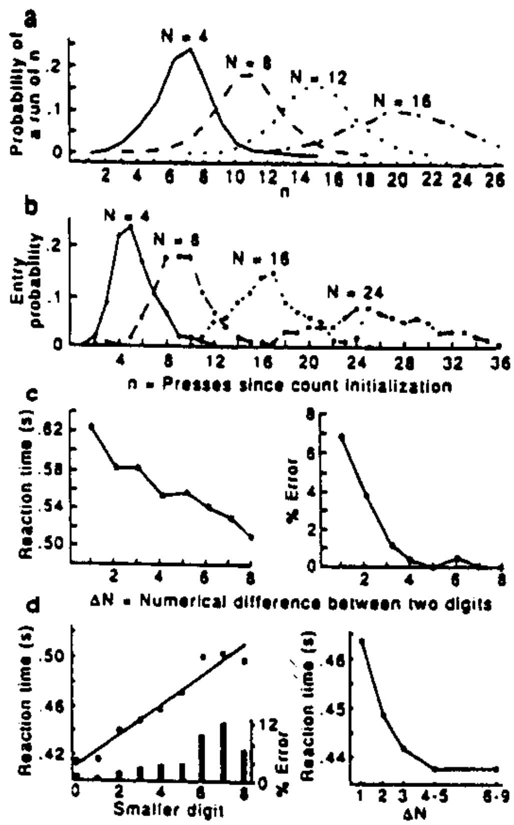

Figure 1. *Evidence that both animals and humans represent numerosities by mental magnitudes. (a) Rat data: the probability of a run of length n on lever A prior to a switch to lever B, as a function of the number, N, of consecutive presses on A required to arm B. (Redrawn from Mechner, 1958, p. 113, by permission of the author and the publisher.) (b) Rat data: the probability of breaking off to enter the food delivery area as a function of n, the number of presses made since the initialization of the response counter, for various values of N, the required number of presses. ... wn from Platt and Johnson, 1971, by permission of the authors and the publisher.) (... n data: reaction time (left panel) and error rates (right panel) in judging which of ... s is larger, as functions of their numerical difference. (Redrawn from Moyer and ...auer, 1967, by permission of the authors and the publisher.) (d) Human data: reaction time (filled circles) and error rates (bars) in judging which of two digits is larger, as functions of the numerically smaller digit (left panel). Residual reaction time (after factoring out the contribution from the magnitude of the smaller digit) in judging which of two digits is larger, as a function of their numerical difference (right panel). (Redrawn from Parkman, 1971, by permission of the author and the publisher.)*

required number of presses. The modal value of n matches N from $N = 4$ to $N = 24$, the full range tested.

Both data sets in Figure 1 show two characteristics. The distributions have appreciable spread, even for numerosities as small as 4. No matter how small the numerosity, the likelihood that a rat will confound it with a nearby numerosity is not negligible. Secondly, animal number discrimination conforms at least qualitatively to Weber's law. The greater the reference numerosity, the more imprecisely the rats distinguish between it and nearby numerosities.

These Weber characteristics of animal number discrimination data are reminiscent of magnitude discrimination data, both human and animal. The striking similarity between number discrimination data and magnitude discrimination data is demonstrated in an experiment by Meck and Church (1983), in which rats discriminated on the basis of the number of elements in a stimulus set or the duration of the sequence of elements. Meck and Church trained rats to choose between two levers on the basis of the sound sequence heard just before the levers appeared. One sequence consisted of a half-second of white noise, a half-second of silence, a second half-second of noise and a second half-second of silence, followed by the appearance of the two levers. Thus, this sequence consisted of two cycles (two noises), with a total sequence duration of 2 s. The other training sequence consisted of eight of these same noise–silence cycles, so that the duration of the sequence was 8 s. Note that on training trials the number of sounds in the sequence and the duration of the sequence covaried. Following the 2-sequence, one lever (the "2-lever") was "correct;" following the 8-sequence, the other lever (the "8-lever") was. A press on the incorrect lever was never rewarded, but a press on the correct lever was only rewarded 50% of the time. By accustoming the rats to trials on which even correct choices were not rewarded, Meck and Church could interpolate test trials on which neither lever was correct (neither lever was rewarded), without disrupting the performance of the discrimination.

After the discrimination between the two training sequences had reached asymptote, two different series of never-rewarded test trials were interpolated among continued training trials. In one series, sequence duration was constant at 4 s, while the number of cycles varied between two and eight. The rats' choice of the 8-lever was plotted as a function of the number of cycles (filled circles in Figure 2). In the second series, the number of cycles was constant at four, while the duration of the sequence varied between 2 and 8 s. The rats' choice of the 8-lever was plotted as a function of the duration of the sequence (open circles in Figure 2). The two plots in Figure 2 – the numerical discrimination plot and the duration discrimination plot – are statistically indistinguishable and they are equally well fit by the same model of the underlying process. The model is Gibbon's (1981) "sample known exactly with similarity decision rule," which generated the curve through the data in Figure 2.

The data in Figure 2 show, first, that in a training regime in which stimulus duration and stimulus numerosity covaried, the rats learned the relation between both variables and the correct lever. When sequence duration could not be used to predict the correct lever, they chose on the basis of the number of cycles in the sequence. When the number of cycles could not be used, they chose on the basis

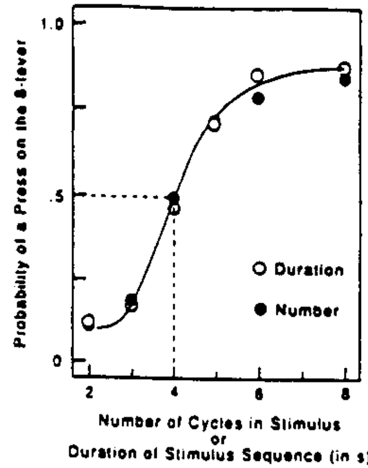

Figure 2. *Probability of choosing the 8-lever as a function of either the number of cycles (noises) in the stimulus (with the duration of the sequence fixed at 4 s) or as a function of the duration of the sequence (with the number of cycles fixed at 4). The curve through the data was calculated from Gibbon's (1981) "sample known exactly with a similarity decision rule" model. (Redrawn from Meck and Church, 1983, p. 323, by permission of the authors and the publisher.)*

of sequence duration. Thus, the data confirm many other results showing that both duration and numerosity are salient aspects of a stimulus (Gallistel, 1990). Second, the results show that the rat's representation of the numerosity of the cycles in the sequence and its representation of the duration of the sequence are indistinguishable from a psychophysical standpoint. Meck and Church (1983) explain this by a model, to be presented in more detail shortly, in which it is assumed that numerosity is represented by the same mental magnitudes that represent temporal durations (just as the columns in a column graph may represent either numerosity or duration). Their model further assumes that the mental magnitudes representing numerosity (the columns) have the same scalar variance property as the mental magnitudes representing duration. The standard deviation of the population of magnitudes that represent a given numerosity (or a given duration) increases in proportion to the mean of the population.

Transfer on the basis of numerosity

The numerosity of a set is independent of the sensory attributes of its members. This is why, in empiricist theories of mind, number is a highly abstract and therefore derivative property of a set. Indeed, if Locke was correct in claiming that there is nothing in the mind that was not first in the senses, then it is unclear how number ever comes into the mind, because the numerosity of a set is not something that acts on a sensory receptor. Do the experimental results on animal number discrimination suggest that animals categorize sets on the basis of their numerosity independent of the sensory attributes of the elements in the set?

Fernandes and Church (1982) taught rats to choose between levers on the basis of the number of noise bursts they heard just before the levers appeared. When they substituted light flashes for noise bursts, the discrimination transferred. Church and Meck (1984) taught rats to choose one lever following either a 2-flash sequence or a 2-burst sequence and the other following either a 4-flash or 4-burst sequence. When they interpolated never-rewarded test trials on which the rats were given 2 flashes and 2 bursts in tandem, the rats chose the 4-lever, that is, they chose on the basis of the numerosity of the set of 4 elements composed from two subsets of 2 elements, even though each subset alone induced the opposite choice on training trials. When the rats heard one burst and saw one flash in tandem, they chose the 2-lever, even though the compound set and both of its subsets (single flash and the single burst) did not figure in training. Both of these results are striking examples of transfer of training based on the numerosity of the stimulus sequence, without regard to the sensory characteristics of the members of that sequence. Capaldi and Miller (1988) trained rats to adjust their runway performance on the basis of the number of rewarded trials in a sequence and showed that performance was unaffected by substituting novel rewards, even though the rats could be trained to discriminate solely on the basis of the number of sequential rewards of a particular kind. Matsuzawa (1985) taught a chimpanzee to choose the arabic numeral representing the numerosity of a set of red pencils, then tested transfer to sets composed of other items. Transfer was immediate and total. In short, there are numerous demonstrations that animals abstract the numerosity of a set independent of the sensory attributes of its members.

The preverbal counting model

Meck and Church (1983) proposed and tested a model of the mechanism that maps from the numerosity of a set to the mental magnitude that represents it. Their counting mechanism is diagrammed in Figure 3. It is composed of a source for a stream of impulses, a pulse former that gates the stream of impulses to an accumulator for a fixed duration (the duration of the counting pulse) whenever an event or object is counted, an accumulator that sums the impulses gated to it, and a readout mechanism that dumps the magnitude in the accumulator to memory when the last event or object has been counted. The operation of this mechanism conforms to the principles that define counting processes (Gelman & Gallistel, 1978). The mechanism pairs states of the accumulator (numerons) with the items in the set being counted. The pairing is one–one, because the pulse former gates a burst of impulses to the accumulator once and only once for each item in the enumerated set. The order in which the states of the accumulator are used is stable (does not vary from one count to the next), because the ordering of magnitudes is isomorphic to the ordering of numbers. The final state of the accumulator – the final numeron in the sequence of numerons used in a count – is

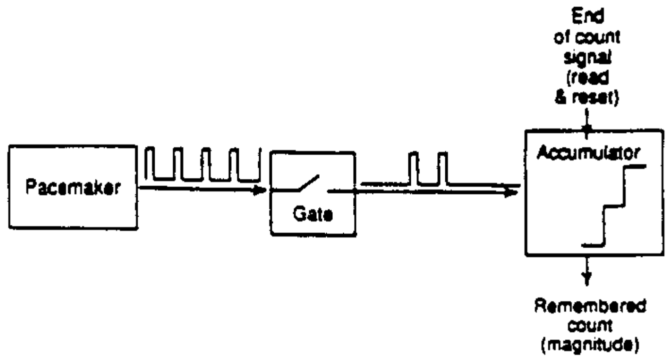

Figure 3. *The model for the animal counting mechanism proposed and experimentally tested by Meck and Church (1983) and Meck, Church, and Gibbon (1985). Each closing of the gate increments the magnitude in the accumulator. The magnitude in the accumulator at the end of the count is the mental representative of the numerosity of the counted set.*

used as a representative of the numerosity of the set (the cardinal principle). Also, the current content of the accumulator is used as a representative of the numerosity of the set so far counted in decision processes that involve comparing a current count to a remembered count. In these processes, the magnitudes previously read out of the accumulator into long-term memory represent previously experienced numerosities, while the magnitude currently in the accumulator represents the currently experienced numerosity.

The counting mechanism in Figure 3 is a minor modification of the timing mechanism proposed by Gibbon (1981), which generates the mental magnitudes that represent temporal duration. When the mechanism is used to generate magnitudes that represent the duration of an interval, the gate closes at the beginning of the interval and opens at the end, so that the magnitude in the ac[illegible]lator is proportionate to the duration of the interval. Meck and Church ([illegible]) call this the "Run" mode. This timing mechanism becomes a counting m[illegible]ism when the gate closes for a short fixed interval once for each stimulus in the sequence being counted, so that the magnitude in the accumulator at the end of the sequence is proportionate to the number of elements in the sequence. (Meck and Church call this the "Event" mode.)

Meck and Church estimated that the duration of a closure in the Event mode was about 0.2 s. This set the stage for an audacious experimental test of the hypothesis that both numerosity and duration were represented by the same kinds of mental magnitudes. The idea was to train rats on a duration discrimination, then present them with stimulus sequences whose durations lay far beyond the range of durations on which the rats had been trained, but whose numerosities were such as to generate mental magnitudes estimated to be of the same size as the magnitudes generated during duration training. The experiments tested for transfer from the duration training to the numerosity tests based on similarity in the mental magnitudes used to represent these two different aspects of a stimulus. That is, would rats base their decision on the magnitude of the representatives independent of what those representatives represented? To return to the column graph analogy for a moment: when rats had been taught to respond one way for a certain mental column height and another way for a different mental column height, under conditions where those mental columns represented durations, would they then read the column heights the same way when the columns represented numerosities rather than durations? The answer, surprisingly, was yes (Meck & Church, 1983; Meck, Church & Gibbon, 1985). This is strong evidence that the preverbal representatives of numerosity are magnitudes.

Animals reason arithmetically

When we say that animals reason arithmetically we mean that their brains process numerons (representatives of numerosity) in operations isomorphic to the operations of arithmetic. If the magnitudes that represent numerosities enter into processes equivalent to ordering, addition, subtraction, multiplication and division, then, by our definition, the animals reason arithmetically. Evidence for these kinds of operations with numerons come from the analysis and modeling of the decision processes that underlie performance in more complex discrimination tasks involving number, duration and rate (number divided by duration).

For example, in the data from Mechner (1958) shown in Figure 1a, there is a fixed numerical difference between the median number of presses at which the animal broke off to try the other lever and the number of presses required to arm the other lever. Under the particular conditions that generated the data in Figure 1a, the median number of presses was always three more than the required number. When Mechner manipulated the probability that the rat would have to abandon lever A and move to lever B to collect its food, this difference changed. When the probability was high, this difference was low; when the probability was low, this difference was high. Thus, the decision process underlying this behavior appears to involve computing the difference between the current number of presses and the required number, then comparing this difference to a criterion (subtraction followed by ordering). On the other hand, in the data from Meck and Church (1983) shown in Figure 2, where the rat was asked in effect to which of two training numerosities (2 and 8) the numerosity presented on a test trial was more similar, the rats' decision process apparently rested on the computation of two ratios and their comparison (ordering). The indifference point, where the animals' choices split fifty–fifty between 2 and 8 was 4, which is the geometric mean or equiratio point. When the ratio between 8 and the test number was smaller than the ratio between the test number and 2, the rats preferred the

8-lever; when the first ratio was greater than the second, they preferred the 2-lever. The model that generated the curve in Figure 2 is based on the assumption that the animal's decision process involves the comparison of these ratios. It is also based on the scalar variance assumption, the assumption that the standard deviation of the population of magnitudes assigned to represent a given numerosity is proportionate to the mean magnitude, as is evident in the discrimination data in Figure 1a and 1b.

In their numerous studies of duration discrimination, Gibbon, Church, and Meck (1984) have used a variety of tasks involving different, often complex decision processes, and they have extensively analyzed their data from the standpoint of different models of the underlying decision processes. The models required to explain their data involve the addition, subtraction, division, and ordering of the mental magnitudes representing durations. As we saw above, there is evidence that the same magnitudes are used to represent numerosity. It seems reasonable to suppose that the same arithmetic operations may be applied to these magnitudes when they represent numerosity as are applied when they represent duration.

A reason for assuming that the same set of arithmetic operations are available for processing both the magnitudes that represent numerosity and the magnitudes that represent duration is that many behaviorally important computations involve the arithmetic combination of the mental representations of these two different variables, as, for example, in the computation of rate of return. There is a well-documented tendency for animals to apportion their time between different foraging locations in proportion to the relative rates of return that they have experienced from these locations (Godin & Keenleyside, 1984; Harper, 1982; Herrnstein, 1970; Smith & Dawkins, 1971). The rate of return from a foraging patch is the number of morsels obtained while in that patch divided by the duration of one's stay and multiplied by the average magnitude of the morsels. Attempts to model the sensitivity to the rate of return from a patch with "rule-of-thumb" mechanisms that do not involve the performance of the computation just spelled out have not been successful (Gallistel, 1990, Ch. 11; Lea & Dow, 1984). The successful models of this phenomenon explicitly or implicitly assume that the animal accurately represents the (local) rate of return (Gallistel, 1990, Ch. 11).

The most direct demonstration of arithmetic competence with numbers in a nonverbal animal is the experiment of Boysen and Berntson (1989). Like Matsuzawa (1985), they taught a juvenile chimpanzee to choose the arabic numeral corresponding to a given numerosity of oranges in the range from 1 to 5. When she was induced to search in several different sites for oranges and then pick a numeral, she spontaneously picked the numeral that specified the total number of oranges she had seen. This might simply mean that she continued her count from site to site. However, when they put arabic numerals in the hiding sites in lieu of oranges, the chimpanzee spontaneously picked the numeral that represented the sum of the values represented by the numerals she had seen.

Implications for adult numerical competence

Having briefly surveyed the evidence on animal numerical competence and the mechanisms thought to underlie it, we now develop several hypotheses about the relevance of this preverbal numerical competence to human verbal competence with numbers and the acquisition of that competence by children. In this section, we suggest that when we learn to count, we also learn a bidirectional mapping between the preverbal magnitudes that represent numerosity and the number words. We use this mapping, we suggest, in every aspect of verbally assayed numerical competence: in subitizing, in judging the order of two digits, and in retrieving the digit addition and digit multiplication facts.

The bidirectional mapping hypothesis

The important characteristics of the bidirectional mapping between the digits and the coresponding preverbal magnitudes are portrayed in Figure 4. The preverbal magnitudes are represented by the black columns immediately above the quoted digits. The digits are in quotes because this is a mapping from the numerals or words for these digits, not from the numerical values to which they refer. The mapping from preverbal magnitudes to digits is given by the digitally labeled intervals on the mental number line to the left of the columns. The interval within which the top of a column falls determines the digit that a subject will produce if he is required (by, for example, time limitations) to produce a digit on the basis of a preverbal representation of numerosity rather than on the basis of a verbal count. In other words, a numerically competent human must not only learn to generate or find a preverbal magnitude appropriate to a given digit (the mapping from digits to preverbal magnitudes), he must learn to partition the range of preverbal magnitudes and assign the appropriate digits to the resulting intervals (the mapping from preverbal magnitudes to digits).

A pivotal assumption about the mapping from digits to preverbal magnitudes is that there is variability in the magnitudes to which a digit maps and this variability obeys Weber's law: the standard deviation of the distribution of magnitudes to which a digit maps increases in proportion to the mean magnitude. In Figure 4, the fade from black to white at the top of a column represents the variability in the magnitudes to which the digit at the bottom of the column maps. There is negligible probability that the magnitude to which the digit maps will be less than the level at which the column begins to fade. There is a 50% chance that the

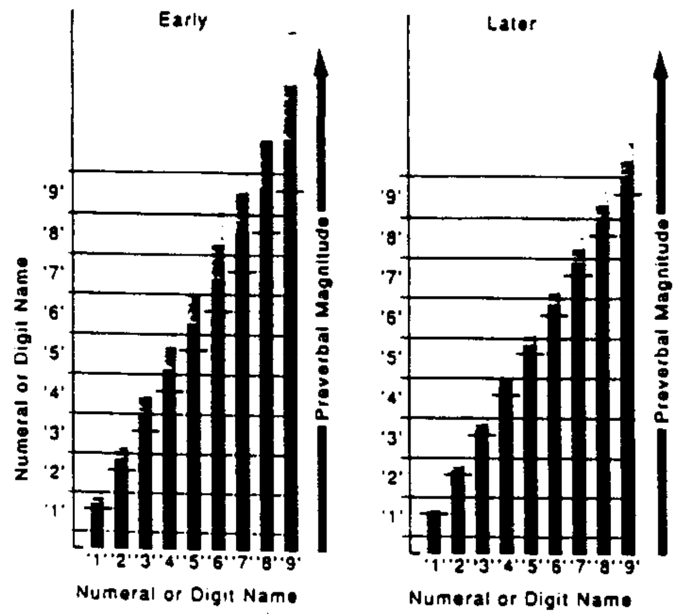

Figure 4. *Schematic illustration of the bidirectional mapping hypothesis. Preverbal magnitudes are represented by the lengths of the black columns. Spoken or written digits are represented by the digits in quotes. The fading of the columns at their tops represents the variability in the magnitudes to which a digit maps. The reduction of the intervals over which this fading occurs between the "Early" and "Later" panels represents the speed-accuracy trade-off. The more time allowed for the mapping, the less variability in the resulting magnitudes. The mapping from magnitudes to digits is mediated by the partition of the mental number line (at left) into segments labeled by digits.*

magnitude will be less than the level at which the column has faded to 50% gray. (This level – the mean of the magnitudes to which a digit maps – is marked by a gray crossbar in Figure 4.) There is negligible probability that the magnitude will exceed the level at which the bar fades to white. Notice that the interval over which a column fades increases in proportion to the mean magnitude. This proportionate increase in variability is presumably what underlies Weber's discrimination law. It is what Gibbon (1977) has termed the "scalar variance" property of the mental magnitudes that represent numerosity, duration, and other physical magnitudes. The term "scalar variability" is perhaps preferable, because it is the standard deviation, not the variance, that scales with the mean magnitude.

The scalar variability assumption about the mapping from number words to the mental number line may be contrasted with the compressive mapping assumption made by Dehaene and Mehler (1992). They assume that for a given difference between two numbers, the greater their mean numerical value, the smaller the difference between the mental magnitudes to which they map (the smaller their subjective difference). This would be the case, for example, if the mapping from numerical value to mental magnitude were logarithmic. In our model, the mapping from numerical value to mental magnitude is linear (indeed, nearly scalar). Hence, for a given difference between two numbers, the difference between the mental magnitudes to which they map (the subjective difference) is independent of their mean numerical value. The discriminability of the two numbers decreases as their mean numerical value increases, not because they are subjectively closer together, but because the variability (noise) in the mapping is scalar. To put the same point from a somewhat different perspective, in the Dehaene and Mehler model, the subjective difference between two equally discriminable numbers is a constant, whereas in our model, the subjective difference between two equally discriminable numbers increases in proportion to their mean numerical value.

We prefer the scalar mapping assumption to the logarithmic (or, more weakly, the compressive) mapping because we assume that the magnitudes to which external number symbols map are the same magnitudes with which the brain performs both preverbal addition and preverbal multiplication, and we assume that these magnitudes are the magnitudes generated by preverbal counting mechanisms in both humans and other animals. If the mapping from numerical value to mental magnitude is scalar, then the concatenation of mental magnitudes is isomorphic to the addition of the corresponding values and the area-generating algorithm described above for histogram arithmetic is isomorphic to their multiplication. If the mapping from numerical value to mental magnitude is logarithmic, then concatenation of the magnitudes becomes isomorphic to multiplication of the numerical values, but it is more cumbersome to construct a mental operation that is isomorphic to addition of the numerical values. Also, as already noted, the psychophysics of number and duration discrimination appear to be identical in animals (Figure 2), and Gibbon and Church (1981) have shown that subjective duration in animals is a linear, not a logarithmic function of objective duration.

A second pivotal assumption about the mapping from digits to preverbal magnitudes is that there is a speed–accuracy trade-off: the longer the time allowed for generating the preverbal magnitude that corresponds to a digit, the less the variability in the magnitudes generated. In Figure 4, the effect of time on the variability in the magnitudes assigned to a given digit is indicated by the contrast between the panel labeled "Early," which shows the distributions if the mapping process is terminated early on, and the panel labeled "Late," which shows the distributions if the mapping process is terminated later on. In the latter panel, the mean magnitudes are the same but the intervals over which the columns fade are narrower by a factor of 3. Alternatively, the Early and Late panels can be understood as referring to different stages in the child's development of numerical competence. We assume that if the time allowed to generate the magnitude that corresponds to a digit is held constant, then the older, more practiced child will make a less variable mapping than the younger, less practiced child.

Subitizing

Gallistel and Gelman (1991) propose that the subitizing process – the rapid preverbal or nonverbal estimation of numerosity – is, in essence, the animal counting mechanism plus the learned mapping from the preverbal magnitudes to the digits. The learned mapping between preverbal magnitudes and digits gives us an alternative way of obtaining the word that represents the numerosity of a set. Instead of, or in addition to, verbally counting the set, we can count it preverbally and use the resulting mental magnitude to generate the corresponding verbal representative.

We suggest that the advantage of this second route is speed. Rapid subvocal verbal counting in adults has a slope of greater than 300 ms per item. The preverbal counting process may be much faster. Pigeons, for example, can count their pecks (up to numerosities at least as great as 50) when pecking at faster than 6 pecks per second (Rilling, 1967; Rilling & McDiarmid, 1965). The disadvantage of preverbal counting, at least when performed very rapidly, is inaccuracy. For sets with numerosities greater than 4 or 5, rapid preverbal counting frequently generates the wrong digit because of the inherent variability of the magnitudes generated by the preverbal counting mechanism. The preverbal counting mechanism may generate a magnitude with its terminus (top) in an adjacent or nearby segment of the number line rather than in the segment labeled the correct digit. The probability of these kinds of errors increases as the numerosity of the counted set increases (scalar variability in the mapping from numerosity to preverbal magnitudes, see Figure 1a and 1b). The increasing probability of generating an erroneous digit limits the usefulness of this second route to the small numerosities. To reduce the probability of an erroneous answer based on the subitizing strategy one has to reduce the variability in the magnitudes generated by the preverbal counting mechanism. The requisite reduction in variability may only be purchased by an increase in preverbal counting time. That is, we assume that the mapping from numerosity to preverbal magnitudes done by the preverbal counting mechanisms shows a speed–accuracy trade-off like that portrayed in Figure 4. Because the requisite increase in preverbal counting time renders the subitizing strategy slower than the verbal counting strategy, skilled performers might be expected to limit their use of preverbal counting. We argue that more often than not, when adults are in timed tasks, they use the subitizing strategy for numerosities of 4 or fewer, because the probability that rapid subitizing will produce an erroneous number word is sufficiently small to favor its use over a slower verbal counting mechanism.

Judgments of digit order

The hypothesis that mental operations with verbally or visually presented digits depends on a mapping to mental magnitudes goes back to the seminal experiments of Moyer and Landauer (1967) and Restle (1970). If humans use mental magnitudes to represent numerosities when they judge which of two digits represents the greater numerosity, then our judgments of numerical inequality should obey Weber's law. The more nearly equal the numerosities specified by the two digits, the harder it should be to determine which is larger (or which is smaller). And, for a given difference in the specified numerosities, the larger they both are, the harder it should be to determine which is larger. It comes as a distinct surprise to most people to learn that both of these results have repeatedly been experimentally demonstrated since they were first reported by Moyer and Landauer (1967). When subjects of any age and any degree of mathematical education are shown a pair of digits and asked to press one of two buttons as quickly as they reasonably can to indicate which is larger (or which smaller), they react more quickly and make fewer errors as the difference between the specified numerosities increases (Figure 1c and 1d, right panel). Also, the greater the specified numerosities both are, the longer it takes to say which is larger (or smaller), and the more likely a subject is to err (Figure 1d, left panel).

When humans are asked to judge which of two physical magnitudes (line segments, pitches, etc.) is greater (longer, higher, etc.), the reaction time data are well represented by the empirical equation

$$RT = a + k \log [L/(L - S)]$$

with RT the reaction time, L the larger physical magnitude, S the smaller physical magnitude, and a and k constants (Welford, 1960). The same equation accounts for more than 80% of the variance in the reaction time data shown in Figure 1c and 1d (Moyer and Landauer, 1973), suggesting that "the decision process . . . is one in which the displayed numerals are converted to analogue magnitudes, and a comparison is then made between these magnitudes in much the same way that comparisons are made between physical stimuli such as loudness or length of line" (Moyer & Landauer, 1967, p. 1520).

The hypothesis that ordination is mediated by a mapping from verbal or written numbers to preverbal magnitudes does not, by itself, deal adequately with the behavioral data on the ordination of double-digit numbers. Supplementary hypotheses or perhaps entirely new explanatory frameworks are required. The extensive literature on the so-called split and magnitude effects in the judgment of numerical order with double-digit numbers are treated at length by Dehaene (1989), Dehaene, Dupoux, and Mehler (1990) and Link (1990).

Retrieving the number facts

Given that the verbally mediated computation of sums and products depends on verbally taught number facts and algorithms, one might assume that the verbal combinatorial operations were not dependent on a mapping from the verbal to the preverbal representation of numerosity. We wish to suggest, however, that the reaction time data suggest that the retrieval of the single-digit addition and multiplication number facts, which is central to these verbal algorithms, is mediated by the same mapping to preverbal magnitudes that makes the ordination of the digits psychologically possible.

The extensive experimental literature on the chronometry of number fact retrieval in adults is reviewed by Ashcraft (this issue) and McCloskey (this issue). The most salient finding is that there are similar magnitude effects (problem size effects) for both addition and multiplication. The bigger the numerosities represented by a pair of digits, the longer it takes to recall their sum or product and the greater the likelihood of an erroneous recall. The same is true in children (Campbell & Graham, 1985). For both sets of number facts, there is a glaring exception to this generalization (and some less striking exceptions). The glaring exception is that the sums and products of number twins (for example, 4 + 4 or 9 × 9) are recalled much faster than is predicted by the regressions for non-twins (although ties, too, show a magnitude effect: Miller *et al.*, 1984). A third finding of theoretical importance is the striking similarity in the effect of problem size on the reaction times for both addition and multiplication. The slopes of the regression lines (reaction time vs. the sum or product of the numbers involved) are not statistically different (Geary, Widman, & Little, 1986). More importantly, Miller, Perlmutter, and Keating (1984) found that the best predictor of reaction times for digit multiplication problems was reaction times for digit addition problems, and vice versa. In other words, the reaction time data for these two different sets of facts, which are mastered at different ages, show very similar microstructure.

The following hypothesis about the mechanism underlying the retrieval of the digit addition and multiplication facts explains these three aspects of the chronometric data. Following Restle (1970), we assume that the results of digit addition are recalled by mapping the addends to positions on the mental number line, preverbally adding the magnitudes thus demarcated to obtain a new magnitude that is their (preverbal) sum, then mapping from this new magnitude back to the verbal domain. The magnitude effect in addition is explained in the same way as the magnitude effect in judging numerical order. It is a consequence of the speed–accuracy trade-off in the mapping from the verbal domain to the preverbal number line. The greater the variability in the magnitudes assigned to each addend, the greater the likelihood that the magnitude that is their sum will map back to the wrong number word. Since the variability in the inverse mapping shows the Weber characteristic – it increases in proportion to the numerosity represented – the greater the numerosities represented, the longer the process must wait in order to obtain a mapping of acceptable reliability. The reason twins are so much faster is that the mapping to the number line is only done once: the same magnitude represents both addends.

This model may be extended to the retrieval of multiplication facts by assuming that mapping from the preverbal to the verbal domain is a mapping from a two-dimensional mental field (rather than a one-dimensional mental line, as in addition). The field is divided into non-overlapping subregions that fill the field (tile the plane). Each product in the multiplication table is associated with one and only one subregion in this mental field, and vice versa. In short, the mental field is isomorphic to the multiplication table. The number word that is retrieved as the product of two digits is determined by the locus of activity in this mental field. The locus of activity is determined by the mental magnitudes to which the digits to be multiplied map. The distance of this activity from one boundary of the field is determined by the magnitude to which the multiplier has been mapped and the distance of this activity from the orthogonal boundary is determined by the magnitude to which the multiplicand has mapped. Therefore, the two-dimensional probability density function for the locus of activity in this field at a given moment after the retrieval process been initiated is determined by the time-dependent probability density functions for the mappings from the digits to be multiplied to their corresponding preverbal magnitudes. At a fixed time after the initiation of number-fact retrieval, the greater the product of the two numbers, the more likely it is that the locus of activity in the multiplication field will fall within an erroneous subregion. However, the longer the retrieval process is allowed to work before a decision is made, the less the likelihood that the locus of activity will be in an erroneous subregion at the moment of decision. The hypothesis is illustrated in Figure 5.

The magnitude effect in the retrieval of the multiplication facts is explained in the same way as the magnitude effect in numerical comparison and in the retrieval of the addition facts. All three are consequences of the scalar variability in the mapping from verbal and written digits to preverbal magnitudes and the trade-off between speed and accuracy in the determination of these magnitudes. The quicker reaction to number twins is explained in the same way as it is in the case of addition: only one magnitude needs to be determined. The similarity between the reaction time function for addition and the reaction time function for multiplication follows from the assumption that both functions are determined primarily by systematic variations in the temporal intervals required to achieve an acceptably accurate mapping to preverbal magnitudes. Miller et al. (1984) argued that this similarity in the reaction time patterns for addition and multiplication was reason for rejecting analog models of these processes, but this similarity is a necessary property of our model for number fact retrieval, as are the magnitude

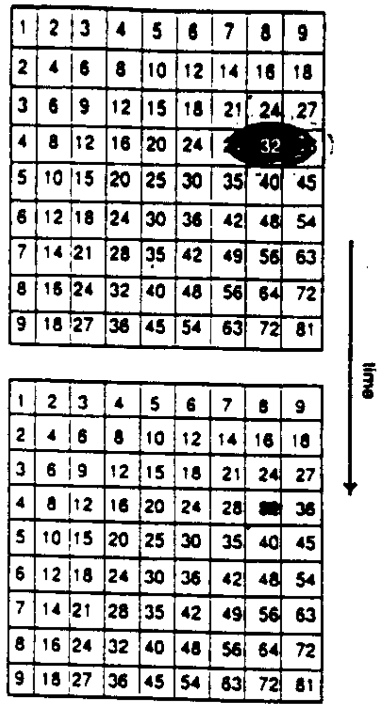

Figure 5. *Schematic illustration of the model for the retrieval of the multiplication number facts. Each product is assumed to be associated with a subregion of the number field. When the product of a pair of digits is to be retrieved, each digit evokes a corresponding preverbal magnitude. There is a time-dependent decrease in the variability in the magnitude assigned to a digit and the standard deviation of the probability density function for this assignment is a scalar function of the numerosity represented by the digit (as illustrated in Figure 4). The magnitudes to which the two digits map determine the coordinates of the locus of activity in this field. The resulting time-dependent two-dimensional probability density function for the locus of activity is indicated by the shaded oval blur; the darker the shading, the greater the probability density. The number in any square that is at least partly shaded has some probability of being retrieved as the product.*

effect and the shorter latencies for number twins. In associative network models of number fact retrieval, all three predictions depend on ad hoc, not very plausible assumptions about the frequencies with which people have experienced correct and incorrect pairings of the operands and their sums or products.

Our model of number fact retrieval also explains salient aspects of the error patterns. When adults respond with an erroneous product, the great majority of these errors are numbers from the multiplication table, rather than numbers like 11, 13, and 20, which are not in the multiplication table (Campbell, 1987; Miller et al., 1984), even though there are slightly more of the latter numbers in the range from 1 to 81. Moreover, table errors are usually products located close to the correct product in the table, most often a neighboring product along a row or column (so-called operand errors, because the product retrieved shares an operand with the product sought). As is evident in Figure 5, our model predicts this error pattern. It predicts that when the magnitudes used to retrieve a product diverge by too much from their correct values, the coordinates of the resulting activity will fall most often in the subregions for products that are adjacent along a row or column of the table.

Another, suprisingly common error is the retrieval of the sum in place of the product and vice versa. This is part of the evidence that the retrieval of the product and sum of two numbers, like the reading of a word, is psychologically obligatory in adults (LeFevre, Bisanz, & Mrkonjic, 1988). Thus, in our model, one would imagine that after the two magnitudes have been generated, they are laid end to end on the mental number line to determine their sum and they are used to mark off orthogonal coordinates on the multiplication field to determine their product. The answer from the wrong operation is retrieved if a subsequent decision process fails to filter the obligatorily produced answers properly.

While our model explains both the effect of ties and the strong correlation between the speeds of retrieving the addition and multiplication facts for a given digit pair – both of which have been seen as incompatible with "structural" models (Miller et al., 1984) – there remain aspects of the reaction time data that it does not explain. The average reaction times across all combinations involving a given operand (ties excluded) do not increase monotonically with the magnitude of the numerosity specified by the operand, as our model predicts they should. For example, combinations involving 5 are on average considerably faster than problems involving 4 and 6 (Figure 6). One is tempted to explain this with an assumption that the magnitude for 10 is a fixed referent with low or no variability and that the magnitude for 5 may be generated rapidly from it by bisection. Problems involving 7 are also somewhat easier than one would predict. Why this should be so is obscure.

Our model is at least compatible with the evidence from cognitive neuropsychology (see McCloskey, this issue) in that it is clear how one could get selective impairment in the retrieval of specific number facts. The association between a magnitude on the number line or a position in the number field and the corresponding number or word could be weakened by a pathological process, resulting in a diminished ability to retrieve a particular sum or product. In other words, patterns of fact-weakening that are patchily distributed within the multiplication table are explained on this model by patchy damage to the number line or the number field.

Dehaene and Cohen (1991) report a study of an acalculic patient that seems to support the idea that we map from verbal and written number symbols to magnitudes and back again, although the details are puzzling. The patient was almost normal in both reaction time and error rate when judging whether a visually presented number was greater or smaller than a fixed reference value (5 for single-digit sessions, 55 for 2-digit sessions). He was very poor at producing

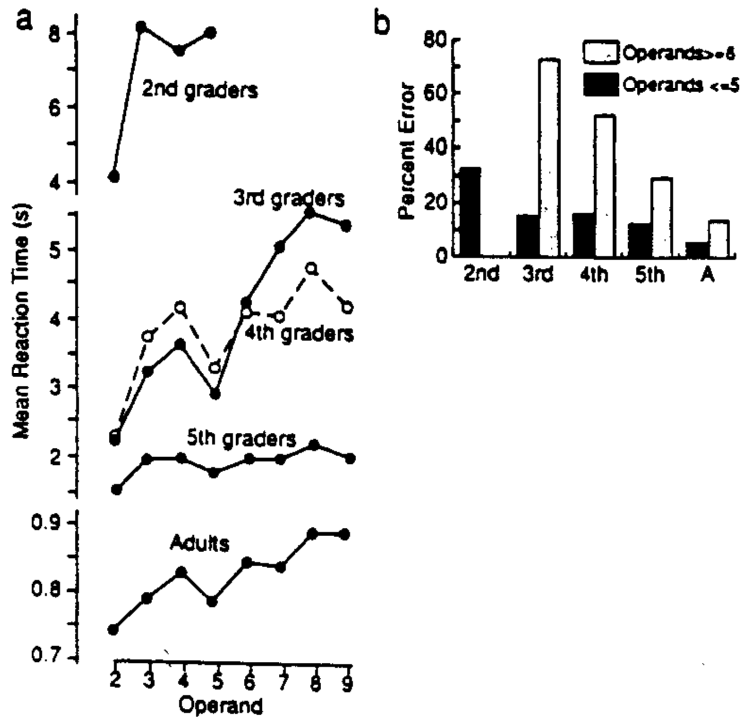

Figure 6. *(a) Mean voice-onset reaction time to give the answer to a digit multiplication problem, averaging across all the problems involving a given operand (ties excluded), as a function of the operand, for operands from 2 to 9. (Redrawn from Campbell and Graham, 1985, p. 348, by permission of the authors and the publisher.) (b) Percent errors in giving (verbally) the products of digit multiplication problems presented on a video monitor for 2nd, 3rd, 4th and 5th graders, and for adults. (Data from Table 1, p. 347, of Campbell & Graham, 1985.)*

the correct answer to single and double-digit addition, subtraction and multiplication problems and at verifying a proffered answer (2 + 2 = 3?), but, at least for addition, he clearly knew the ball park of the correct answer. He was not at all sure whether 2 + 2 equaled, 3, 4, or 5 and he was apt to produce any one of these numbers as an answer, but he was sure the sum did not equal 9 and never produced a number that large as an answer. There are many puzzling details of this complex case, one of which is that in verifying addition problems he showed a strong split effect but no problem-size effect. The bigger the difference between the correct answer and the proffered answer, the more readily he rejected the latter, but his ability to reject an error of a given size was not a function of the size of the two operands or their sum. Nonetheless, the case clearly suggests that operations with numbers are aided and abetted, if not mediated, by a mapping to mental magnitudes.

Developmental implications

The acquisition of verbal counting

We (Gelman, 1990; Gelman & Gallistel, 1978), have argued that the acquisition of verbal counting is made possible by implicit principles, which define what constitute acceptable instances of counting and direct or organize the learning of the conventional verbal counting sequence. When we first put forward this hypothesis, we were not aware of the experimental evidence for counting processes in animals; hence, we had little to say about the provenance of the implicit principles. We now suggest that the preverbal counting process is the source of these principles. In particular, we argue that what guides the acquisition of verbal counting is the isomorphism between the preverbal counting process and the verbal counting process – the similar formal structure of the two processes.

The preverbal process provides a framework that makes the verbal counting process intelligible, hence learnable. Children assimilate verbal counting because it maps onto the unconscious preverbal counting process. The count words map to the preverbal magnitudes. The one–one constraint on the use of count words corresponds to the fact that in the preverbal process the pulse former gates a burst of impulses to the accumulator once and only once for each item in the to-be-counted set. The constraint on the order of the count words – that they should always be used in the same sequence – replicates the ordering of the preverbal magnitudes. The accumulation process passes through the intervening magnitudes en route to the cardinal magnitude just as the verbal counting process passes through the intermediate count words en route to the cardinal count word. The fact that the last count word used represents a property of the set corresponds to the fact that the final magnitude in the accumulator is read out into long-term memory, where it represents the numerosity of the set that was counted.

Implicit in our suggestion is, of course, a more general theory about verbal learning, namely, that verbal learning is possible insofar as there are nonverbal models or mental representations that mediate the interpretation of verbal reference (see also Carey, 1991).

The development of verbal arithmetic reasoning

Gelman and Gallistel (1978) argued that the development of the child's ability to reason verbally about numerosity, numerical relations, and operations that affect numerosity was made possible by implicit domain-specific principles. We now suggest that the preverbal system for reasoning about numerosity provides the framework – the underlying conceptual scheme – that makes it possible for the young child to understand and assimilate verbal numerical reasoning. We suggest further that the mapping from preverbal magnitudes to digits and the inverse mapping from digits to preverbal magnitudes, together with the primitive ability to carry out arithmetic operations with these mental magnitudes, play a fundamental role in the development of mastery over the verbally based algorithms that every school child must learn. We suggest that the preverbal arithmetic computations go on in parallel with the verbally mediated computational algorithms in adult arithmetic, providing a check on whether the results arrived at verbally are in the right ball park. We suggest also that the ability to get approximate results nonverbally via the mapping from verbal numbers to nonverbal magnitudes and the nonverbal arithmetic operations performed with these magnitudes makes it possible for the child to assimilate the verbal system of arithmetic reasoning.

Verbal and preverbal addition, subtraction, and ordination in young children

For preschool children (5-year-olds), as for adults, the closer the two numbers are together, the longer the reaction time in judging which is bigger (Schaeffer, Eggleston, & Scott, 1974; Sekuler & Mierkiewicz, 1977; Siegler & Robinson, 1982). As Resnick (1983, p. 113) points out, this result seems to imply that we can attribute to children entering school "the ability to directly enter the positional representation for a number upon hearing its name (i.e., without counting up to it)." In other words, it implies the preschoolers have already learned the mapping from the digits to the corresponding preverbal magnitudes, although this mapping is in all likelihood slower and more variable in younger, less practiced children.

Preschool children generate verbal answers to verbally posed addition and subtraction problems using vocal or subvocal counting algorithms of their own devising (Gelman & Gallistel, 1978; Ginsburg, 1977; Groen & Resnick, 1977). Reaction time studies show that the subvocal use of verbal counting algorithms continues through the early school years (Groen & Parkman, 1972; Groen & Poll, 1973; Groen & Resnick, 1977; Resnik, 1983; Svenson & Broquist, 1975; Svenson & Sjoberg, 1983; Woods, Resnick, & Groen, 1975), when students are required to "do it in the head" without overt counting.

However, one remarkable finding from research on subtraction algorithms in elementary school children is that the most common way of doing subtraction is by the "choice" algorithm (Resnik, 1983; Woods et al., 1975). In this algorithm, the child obtains the correct number word either by counting the number of steps required to get up to the minuend from the subtrahend or by counting down from the minuend a number of steps equal to the subtrahend. The child computes the difference 7 – 5, by saying (overtly or covertly), "Six, seven – two." [It takes two counts to get up to "seven" from "five."]. The same child computes the difference

7 − 2 by saying, "Six, five – five". [Counting down two steps from "seven" brings you to "five".] The evidence from reaction times that children routinely use this either/or choice strategy is confirmed by the results of interviews (children say that is what they are doing) and by the occasional overt use of this algorithm (Resnick, 1983 and citations therein). The choice algorithm minimizes the number of counting steps that must be counted. If the subtrahend is smaller than the answer sought (as in $7-2=5$, $2<5$), then the child counts down from the minuend by the number of steps specified by the subtrahend. If, however, the answer sought is smaller than the subtrahend (as in $7-5=2$, $5>2$), then the child counts up from the subtrahend.

What is remarkable about the choice algorithm is that it presupposes that the child computes the relative magnitudes of the subtrahend and the answer sought, before it chooses how to compute the verbal form of the answer (the requisite number word). Thus, before it computes the verbal answer, it must compute (preverbally) the magnitude of the difference between the minuend and the subtrahend and compare this magnitude to the magnitude of the subtrahend. The outcome of this preverbal computation determines which verbal count will be used to obtain the verbal representative of the difference. The recursive use of verbal counting (counting the steps in a count) is apparently difficult and error prone, so the child uses preverbal subtraction and comparison to minimize the number of steps that will have to be counted in its verbal computation.

We take the evidence that children use the choice algorithm in subtraction as support for our hypothesis that the acquisition and performance of verbal arithmetic is mediated by the preverbal system for represented numerosity and doing arithmetic computation, the system that we share with the nonverbal animals.

Children in elementary school compute some products by repeated addition, but this is painfully slow and error prone. The algorithms for doing verbal multiplication and division in the absence of useable knowledge of the digit multiplication facts are unsatisfactory, at least without long practice. More commonly, in the extended period before they can reliably retrieve all the products of single-digit numbers, children rely on counting up from or adding to products that they can retrieve (Siegler, 1988). The teaching of the algorithms for multidigit multiplication and of the multiplication number facts on which they rest is an important part of the mathematics curriculum in elementary school. The development of the ability to retrieve these multiplication facts rapidly and accurately is surprisingly prolonged. The average voice-onset reaction time for a fourth grader verbalizing the answer to a digit multiplication problem presented on a video monitor is more than 4 s when one of the operands is 6 or greater, and about 50% of the responses given are erroneous (Figure 6). For adults (college students), the average reaction time for problems with an operand greater than or equal to 6 is 0.8–0.9 s and about 14% of the responses are erroneous.

We think that our model renders some commonly known facts about the pedagogy of mathematics at the elementary level more intelligible than other models of the process for learning the number facts. If learning the number facts depends on the same general-purpose associative process that other kinds of learning are supposed to depend on, then it is not clear how it is possible for some otherwise intelligent children to have so much difficulty mastering the number facts, particularly the multiplication table. Why should it take several years for children to master the 81 multiplication facts (Figure 6), when those same children learn the meanings of hundreds of words every month? And why should learning the multiplication table be so much harder for one child than for another child of comparable intelligence? And why should the speed with which number facts can be retrieved increase throughout the elementary school grades, only reaching an asymptote after puberty (Figure 6)? Our proposal that number fact retrieval depends on a domain-specific mechanism – the mapping from digits to the preverbal magnitudes that represent numerosity and the mapping from the preverbal magnitudes back to digits (via the multiplication field) – makes these selective learning difficulties, wide individual differences, and prolonged development more intelligible. These may all be explained in part by assuming that there are wide individual differences in the speed–accuracy functions for these mappings at a given age and that the trade-off between speed and accuracy improves steadily with age, either from constant practice, or from maturation of the neural mechanisms, or for both reasons.

The difficulty in learning fractions

Part of the evidence that it is the availability of an isomorphic preverbal counting model that makes it possible to assimilate the verbal system of number is the difficulty children have in learning those parts of the modern number system that are not modeled by the preverbal system. Although the representatives of numerosity in the preverbal system are magnitudes and hence continuous variables, the preverbal system for representing numerosity is rendered discrete by the discrete gating of bursts of impulses to the accumulator, one burst for each item in a set of discrete items. There is no provision in this system for generating representatives for fractional numerosities, despite the fact that the representation of numerosity by magnitudes makes it in principle possible to represent intermediate, that is, fractional, numerosities. The concept of a fractional numerosity is counter-intuitive, because there is no provision for it in the preverbal scheme for generating representatives of numerosity.

The principles implicit in the structure of the preverbal counting mechanism enable the initial mastery of the verbal system of number. Most Down's syndrome children do not seem to have access to this system and they find learning to count and add all but impossible, even with the aid of highly structured input and much opportunity for rote learning practice (Gelman & Cohen, 1988). However, there are reasons to expect that these initial principles hinder later progress almost as much as they promote initial progress. Much of mathematics involves operations such as multiplication and division for which algorithms based on counting are extremely cumbersome and error prone, and it involves the manipulation of numbers that cannot be generated by counting processes. The first such numbers the child must come to terms with are the fractions. The fractions, of course, are generated by the unconstrained use of division, an operation for which algorithms based on counting are singularly cumbersome. The positive and negative fractions, together with the positive and negative integers and zero, constitute the so-called rational numbers.

Teachers in the elementary schools have always known that the teaching of fractions is a major pedagogical challenge (e.g., Carpenter, Corbitt, Kepner, Lindquist, & Reys, 1980). Gelman (1990) argues that this is to be expected because fractions cannot readily be assimilated into a system in which numbers are defined as "what one gets when one counts". Within this conceptual framework, common classroom inputs for learning about fractions cannot be interpreted correctly because counting algorithms are useless, both for generating the proper number words and for ordering, adding and subtracting the numbers. One cannot count things to answer "Which is more, 1/2 or 1/4? or 1.5 or 1.0?" But if children cannot answer these questions, then they should not be able to place fractions properly in relation to the whole numbers on the number line – and, indeed, they cannot (Gelman, Cohen, & Hartnett, 1989). Young children may know how to divide a circle or a rectangle and call the pieces "halves" and still not appreciate the numerical meaning of "half". They need to understand that a fraction is the number one gets when one number is divided by another and that all such numbers (the numbers that result from arbitrary divisions) may be ordered, added, subtracted, multiplied and divided right along with the numbers that one gets by counting. Without this kind of understanding they cannot make sense of the claim that "two halves" means the same as "three thirds", "four fourths", "one hundred one hundredths", and so on.

As we would expect, young learners have a robust tendency to "overgeneralize" their counting principles in assimilating the instructional data on fractions (Gelman, 1991; Gelman et al., 1989). For example, they "read" fractions, or non-integer numergraphs, as if these are novel representations for the counting numbers. Most 6- and 7-year old children misread $\frac{1}{4}$ and $\frac{1}{2}$ as "one and four" and "one and two", although some preferred other interpretations, including turning the task into an addition problem and answering "one plus four; one plus two" or "five, three". Further, although these children had learned the correspondences between a few verbal and written expressions for fractions (they knew that "$\frac{1}{2}$" corresponds to "one half" and "$\frac{1}{4}$" to "one fourth"), they had not learned to order these same fractions; they choose $\frac{1}{4}$ as more than $\frac{1}{3}$. When asked to place $1\frac{1}{2}$ circles on a number line, they placed the display at the position for 2 on the line (because the display had two parts, a whole circle and a half circle). Attempts to verbally guide the interpretation of displays intended to instantiate fractions may make performance worse. Children told to place a circle divided into three wedges on a number line along which were arrayed patterns of circles instantiating whole numbers (one circle, two circles, and three circles) often placed the three-wedge display correctly at the same locus as the one circle, but when they were told that the item had "three thirds" on it they most often placed the three-wedge display at the same locus as the three circles. Better performance in the condition where the stimulus was not labeled as a fractional entity was due to children's tendency to apply the perceptual principle of closure and therefore to treat the display of three wedges forming a circle as one thing. Describing the stimulus as three thirds called attention to the three wedges rather than to the one circle they formed.

Thus, as predicted children misinterpret inputs designed to support learning about fractions. They do so, we suggest, because their counting principles do not provide the conditions for accurate uptake of the data; there is no isomorphism between the preverbal system and the relevant characteristics of the data. As a result, they find a way to assimilate these inputs to the counting principles implicit in the preverbal system for generating representatives of numerosity. As creative as these are, they are nevertheless wrong; they do not lead to the growth of an understanding of why fractions are numbers.

Although many children have potent tendencies to misinterpret fraction data, not all do. Some students do acquire an understanding of what it means to say that fractions are numbers. Some clues on how they might do this come from Gelman's (1991) work on the problem. She reports that, with development, children start to behave as if they think the count list has more entries than they knew, for example that one can count "1, $1\frac{1}{2}$, 2, $2\frac{1}{2}$, 3, $3\frac{1}{2}$", etc. Still, at this stage, the children seem not to recognize the pseudodensity of the rational numbers, the fact that the gap between any two numbers, such as 1 and $1\frac{1}{2}$, is occupied by other fractions. And, curiously, they do not recognize the elementary or noncompound fractional numbers as numbers at this stage. In the words of one 7-year-old: "You can count one, one and half, two, two and half, but you can't count zero, zero and a half".

The fact that children come to talk of numbers between "one" and "two" is significant, even if they do not know the mathematical meaning of the terms they use. At least the talk is consistent with the induction that more "numbers" could come between each of those in the extended list. As children continue to add names of this kind, they also create a database that could support relevant inductions, for example that "numbers" could even come between each of those in their extended count list, and so on. This in turn could lead on to the realization that there is no referent for "the next number after one and a half", in other words to the realization that the property of the natural (counting) numbers that every number has a specifiable next number does not hold when one extends the number system to include fractions. Another way to put these developments is to note that they resemble developments in the history of science and mathematics. Although, as just indicated, the meanings of the term "number" in the system of natural numbers and in the system of rational numbers are incommensurate, the meanings of the count words themselves are not incommensurate in the two systems; the integers have the same properties in the system of rational numbers that they have in the system of natural numbers. If one takes advantage of these local commensurabilities, one gets a foot in the door that opens to the new concept. Remarkably, some children do just this. They find places where what they know and what they have to learn in school are at least locally commensurate and try to build from there.

Conclusions

We propose that the development of the ability to deal with numbers at the verbal level depends on a preverbal system for representing numerosity and for carrying out simple computations with numerons (our generic term for representatives of numerosity). The preverbal system uses magnitudes to represent numerosities. These magnitudes are continuous variables. By means of a slightly different mapping mechanism, these same magnitudes may be used to represent a continuous variable like the duration of temporal intervals. However, the mapping mechanism that assigns magnitudes to numerosities – the preverbal counting mechanism – is a discrete process rather than a continuous process.

The preverbal system for deriving a representation of numerosity – the preverbal counting mechanism – renders the verbal system intelligible by providing a domain-specific isomorphic system to which the verbal system may be mapped. The principles that govern verbal counting are the principles implicit in the structure of the preverbal counting mechanism. The extent to which the principles of counting implicit in the structure and functioning of the preverbal counting mechanism determine what can readily be assimilated at the verbal level is shown by the remarkable difficulty children have in learning the numbers that cannot be generated by counting, namely the fractions. We believe that the essence of their difficulty is that in the absence of a suitable counting algorithm they cannot readily learn to map the fractions to the appropriate preverbal magnitudes and hence to their appropriate positions along the mental number line. As a result, they cannot order them or add them.

Similarly, the preverbal system for computing with magnitudes renders the verbal system of arithmetic reasoning intelligible to the child. The preverbal process for comparing magnitudes (preverbal ordination) renders the ordering of the verbal numbers intelligible. We suggest also that the preverbal processes for combining mental magnitudes to obtain the magnitudes equal to their sum or difference provides the foundation for the assimilation of verbal mastery of addition, which begins before children enter school and begin receiving instruction in mathematics. The case for multiplication is less clear, because it is not clear how well children understand multiplication in the absence of lengthy instruction.

The data on reaction times and error patterns in the judgment of numerical order and in the retrieval of the addition and multiplication facts implies, we believe, that a learned mapping from the verbal and written representatives of numerosity to the preverbal magnitudes mediates order judgments and the retrieval of sums and products. The verbal system and the mechanisms that mediate basic operations in the verbal system, such as the retrieval of the number facts, are erected on a foundation provided by the preverbal system, which, we believe, is one of the foundations of animal mentation.

The system of number is remarkable both for its simplicity and its representational power. It is, on the one hand, a system whose rudiments appear to be present in the mental functioning of a wide range of animals, while, on the other hand, it is a vehicle for the most profound and abstruse aspects of human thought. We believe that by studying the ontogeny of the number concept, we may begin to discern the means by which language permits human thought to transcend some of the limitations imposed by the preverbal representations (conceptions) that make language intelligible in the first place.

References

Boysen, S.T., & Berntson, G.G. (1989). Numerical competence in a chimpanzee (*Pan troglodytes*). *Journal of Comparative Psychology*, *103*, 23–31.

Campbell, J.I.D. (1987). Production, verification, and priming of multiplication facts. *Memory and Cognition*, *15*, 349–364.

Campbell, J.I.D., & Graham, D.J. (1985). Mental multiplication skill: Structure, process, and acquisition. *Canandian Journal of Psychology*, *39*, 338–366.

Capaldi, E.J., & Miller, D.J. (1988). Counting in rats: Its functional significance and the independent cognitive processes which comprise it. *Journal of Experimental Psychology: Animal Behavior Processes*, *14*, 3–17.

Carey, S. (1991). Knowledge acquisition: Enrichment or conceptual change? In S. Carey & R. Gelman (Eds.), *The epigenesis of mind: Essays on biology and cognition* (pp. 257–292). Hillsdale, NJ: Erlbaum

Carpenter, T.P., Corbitt, M.K., Kepner, H.S., Jr., Lindquist, M., & Reys, R.E. (1980). National assessment: Prospective of students' mastery of basic skills. In M. Lindquist (Ed.), *Selected issues in mathematics education*. Berkeley, CA: McCutchan.

Church, R.M., & Meck, W.H. (1984). The numerical attribute of stimuli. In H.L. Roitblatt, T.G. Bever, & H.S. Terrace (Eds.), *Animal cognition* (pp. 445–464). Hillsdale, NJ: Erlbaum.

Dehaene, S. (1989). The psychophysics of numerical comparison: A reexamination of apparently incompatible data. *Perception and Psychophysics*, *45*, 557–566.

Dehaene, S., & Cohen, L. (1991). Two mental calculation systems: A case study of severe acalculia with preserved approximation. *Neuropsychologia, 29*, 1045–1074.

Dehaene, S., Dupoux, E., & Mehler, J. (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.

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

Diamond, A., & Goldman-Rakic, P.S. (1989). Comparison of human infants and rhesus monkeys with lesions of the hippocampal formation on AB and object retrieval, two tasks that mark developmental changes in human infants. *Behavioral Neuroscience, 103*, 526–537.

Fernandes, D.M., & Church, R.M. (1982). Discrimination of the number of sequential events by rats. *Animal Learning and Behavior, 10*, 171–176.

Gallistel, C.R. (1990). *The organization of learning*. Cambridge, MA: Bradford Books/MIT Press.

Gallistel, C.R. (in press). A conceptual framework for the study of numerical estimation and arithmetic reasoning in animals. In S. Boysen, & J. Capaldi (Eds.), *The development of numerical abilities: Animal and human models* Hillsdale, NJ: Erlbaum.

Gallistel, C.R., & Gelman, R. (1991). The preverbal counting process. In W.E. Kessen, A. Ortony, & F.I.M. Craik (Eds.), *Thoughts, memories and emotions: Essays in honor of George Mandler* (pp 65–81). Hillsdale, NJ: Erlbaum.

Geary, D.C., Widman, K.F., & Little, T.D. (1986). Cognitive addition and multiplication: Evidence for a single memory network. *Memory and Cognition, 14*, 478–487.

Gelman, R. (1990). First principles organize attention to and learning about relevant data: Number and the animate-inanimate distinction as examples. *Cognitive Science, 14*, 79–106.

Gelman, R. (1991). Epigenetic foundations of knowledge structures: Initial and transcendent constructions. In S. Carey & R. Gelman (Eds.), *The epigenesis of mind: Essays on biology and cognition* (pp. 293–322). Hillsdale, NJ: Erlbaum.

Gelman, R. and Cohen, M. (1988). Qualitative differences in the way Down's syndrome and normal children solve a novel counting problem. In L. Nadel (Ed.), *The psychobiology of Down's syndrome* (pp. 51–99). Cambridge, MA: MIT Press.

Gelman, R., Cohen, M., & Hartnett, P. (1989). To know mathematics is to go beyond the belief that "Fractions are not numbers". *Proceedings of Psychology of Mathematics Education. Volume 11 of the North American Chapter of the International Group of Psychology*.

Gelman, R., & Gallistel, C.R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard University Press.

Gibbon, J. (1981). On the form and location of the psychometric bisection function for time. *Journal of Mathematical Psychology, 24*, 58–87.

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

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

Gibbon, J., Church, R.M., & Meck, W.H. (1984). Scalar timing in memory. In J. Gibbon & L. Allan (Eds.), *Timing and time perception* (pp. 52–77). New York: New York Academy of Sciences.

Ginsburg, H. (1977). *Children's arithmetic: The learning process*. New York: Van Nostrand.

Godin, J.-G.J., & Keenleyside, M.H.A. (1984). Foraging on patchily distributed prey by a cichlid fish (Teleosti, Cichlidae): A test of the ideal free distribution theory. *Animal Behaviour, 32*, 120–131.

Groen, G.J., & Parkman, J.M. (1972). A chronometric analysis of simple addition. *Psychological Review, 79*, 329–343.

Groen, G.J., & Poll, M. (1973). Subtraction and the solution of open sentence problems. *Journal of Experimental Child Psychology, 16*, 292–302.

Groen, G.J., & Resnick, L.B. (1977). Can preschool children invent addition algorithms? *Journal of Educational Psychology, 69*, 645–652.

Harper, D.G.C. (1982). Competitive foraging in mallards: Ideal free ducks. *Animal Behaviour, 30*, 575–584.

Herrnstein, R.J. (1970). On the law of effect. *Journal of the Experimental Analysis of Behavior, 13*, 243–266.

Lea, S.E.G., & Dow, S.M. (1984). The integration of reinforcements over time. In J. Gibbon & L. Allan (Eds.), *Timing and time perception* (pp. 269–277). New York: Annals of the New York Academy of Sciences.

LeFevre, J.-A., Bisanz, J., & Mrkonjic, L. (1988). Cognitive arithmetic: Evidence for obligatory activation of arithmetic facts. *Memory and Cognition, 16*, 45–53.

Link, S. (1990). Modeling imageless thought: The relative judgment theory of numerical comparisons. *Journal of Mathemetical Psychology, 34*, 2–41.

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.

Mechner, F., & Guevrekian, L. (1962). Effects of deprivation upon counting and timing in rats. *Journal of the Experimental Analysis of Behavior, 5*, 463–466.

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.

Miller, K.F., Perlmutter, M., & Keating, D. (1984). Cognitive [illegible]hmetic: Comparison of operations. *Journal of Experimental Psychology: Learning, Memo[illegible]d Cognition, 10*, 46–60.

Moyer, R.S., & Landauer, T.K. (1967). Time required for ju[illegible]s of numerical inequality. *Nature, 215*, 1519–1520.

Moyer, R.S., & Landauer, T.K. (1973). Determinants of reaction time for digit inequality judgments. *Bulletin of the Psychonomic Society, 1*, 167–168.

Parkman, J.M. (1971). Temporal aspects of digit and letter inequality judgments. *Journal of Experimental Psychology, 91*, 191–205.

Platt, J.R., & Johnson, D.M. (1971). Localization of position within a homogeneous behavior chain: effects of error contingencies. *Learning and Motivation, 2*, 386–414.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 4*, 515–526.

Resnick, L.B. (1983). A developmental theory of number understanding. In H. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 109–151). New York: Academic Press.

Restle, F. (1970). Speed of adding and comparing numbers. *Journal of Expeirmental Psychology, 83*, 274–278.

Rilling, M. (1967). Number of responses as a stimulus in fixed interval and fixed ratio schedules. *Journal of Comparative and Physiological Psychology, 63*, 60–65.

Rilling, M., & McDiarmid, C. (1965). Signal detection in fixed ratio schedules. *Science, 148*, 526–527.

Schaeffer, B., Eggleston, V.H., & Scott, J.L. (1974). Number development in young children. *Cognitive Psychology, 6*, 357–379.

Sekuler, R., & Mierkiewicz, D. (1977). Children's judgment of numerical inequality. *Child Development, 48*, 630–633.

Siegler, R.S. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General, 3*, 258–275.

Siegler, R.S., & Robinson, M. (1982). The development of numerical understanding. In H.W. Reese, & L.P. Lipsitt (Eds.), *Advances in child development and behavior* (Vol. 16, pp. 241–312). New York: Academic Press.

Smith, J.N.M., & Dawkins, R. (1971). The hunting behavior of individual great tits in relation to spatial variations in their food density. *Animal Behaviour, 19*, 695–706.

Svenson, O., & Broquist, S. (1975). Strategies for solving simple addition problems: A comparison of normal and subnormal children. *Scandinavian Journal of Psychology, 16*, 143–151.

Svenson, O., & Sjoberg, K. (1983). Evolution of cognitive processes for solving simple subtractions during the first three school years. *Scandinavian Journal of Psychology, 24*, 117–124.

Welford, A.T. (1960). The measurement of sensory-motor performance: Survey and reappraisal of twelve years' progress. *Ergonomics, 3*, 189–230.

Woods, S.S., Resnick, L.B., & Groen, G.J. (1975). An experimental test of five process models for subtraction. *Journal of Educational Psychology, 67*, 17–21.