Available online at www.sciencedirect.com
ScienceDirect
Journal of Experimental Child Psychology 101 (2008) 20–51
Journal of Experimental Child Psychology
www.elsevier.com/locate/jecp

Costs and benefits of representational change: Effects of context on age and sex differences in symbolic magnitude estimation

Clarissa A. Thompson *, John E. Opfer

Department of Psychology, The Ohio State University, Columbus, OH 43206, USA

Received 3 December 2007; revised 12 February 2008
Available online 1 April 2008

Abstract

Studies have reported high correlations in accuracy across estimation contexts, robust transfer of estimation training to novel numerical contexts, and adults drawing mistaken analogies between numerical and fractional values. We hypothesized that these disparate findings may reflect the benefits and costs of learning linear representations of numerical magnitude. Specifically, children learn that their default logarithmic representations are inappropriate for many numerical tasks, leading them to adopt more appropriate linear representations despite linear representations being inappropriate for estimating fractional magnitude. In Experiment 1, this hypothesis accurately predicted a developmental shift from logarithmic to linear estimates of numerical magnitude and a negative correlation between accuracy of numerical and fractional magnitude estimates ($r = -.80$). In Experiment 2, training that improved numerical estimates also led to poorer fractional magnitude estimates. Finally, both before and after training that eliminated age differences in estimation accuracy, complementary sex differences were observed across the two estimation contexts.

Keywords: Conceptual development; Microgenetic; Transfer of learning; Sex differences; Number concepts; Magnitude estimation; Mathematical cognition; Representation

* Corresponding author. Fax: +1 614 688 8261.
E-mail addresses: thompson.1345@osu.edu (C.A. Thompson), opfer.7@osu.edu (J.E. Opfer).

0022-0965/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jecp.2008.02.003

Introduction

Whether transferring knowledge from one classroom to another classroom, from early in the school year to later in the school year, or from one example to other similar examples, conceptual representations allow learners to generalize over situations that differ merely in place, time, and superficial details (Murphy, 2002). These examples of narrow transfer of learning are unremarkable and easy for learners to achieve because narrow transfer of learning apparently occurs automatically—that is, without learners consciously monitoring the breadth of their generalizations—as learners convert stimulus-specific verbal and visual information into abstract conceptual representations (Kourtzi & Kanwisher, 2001; Naccache & Dehaene, 2001; Potter & Faulconer, 1975; Potter & Kroll, 1987; Potter, Kroll, Yachzel, Carpenter, & Sherman, 1986). In this study, we examined one interesting implication of this analysis: Because narrow transfer is an *automatic* effect of abstract representation, representational changes can impose *costs* as well as benefits, leading to unavoidable setbacks in the course of learning that can persist for many years and across many contexts.

Evidence for the benefits of representational change is widespread in the literature. For example, when children are given corrective feedback on where to place a few numbers on a line flanked by 0 and 1000 and no numbers in between (number line estimation), accuracy improves greatly for numbers in the initial training set, learning transfers to numbers outside of the training set, and there is robust transfer of learning to related numerical tasks (e.g., categorizing numbers as "small" or "large"), with magnitude estimates on the transfer task being nearly identical to estimates on the training task (Opfer & Siegler, 2007; Opfer & Thompson, 2008). Moreover, real-world tasks that involve similar kinds of experiences (e.g., playing board games) also result in transfer to educationally important outcomes such as preschoolers' ability to compare numerical value and perform arithmetic (Griffin, Case, & Siegler, 1994; Ramani & Siegler, in press; Siegler & Ramani, in press).

Evidence for the costs of narrow transfer, however, is rare and indirect. One type of evidence for the costs of narrow transfer comes from research on cognitive illusions in adults (Kahneman & Tversky, 1996), who make grossly mistaken comparisons of risk when framed in a manner that invites inappropriate transfer of numerical representations. As a real-life example, genetic counselors often attempt to simplify the risks reported in epidemiological studies by reporting rates of disease in terms of simple frequencies (e.g., 1 in 333) rather than in scientific format, which reports rates of disease per unit of population exposed to the risk (e.g., 3 per 1000 persons) (Burkell, 2004; Grimes & Snively, 1999; Walker, 1997). Although this simplification is well meaning, research on patients' understanding of medical risks has shown that the simplification has the unfortunate consequence of leading patients to make inaccurate comparisons, for example, judging a disease with a rate of 1 in 384 persons as being higher than a disease with a rate of 1 in 112 persons (Grimes & Snively, 1999). Unlike the beneficial effects of transferring from the spacing of numbers in board games to the spacing of numbers on number lines, transferring from the number line to assessments of risk is costly in this case because, unlike the linear increase in spacing of numbers on number lines, the average risk of disease per unit of population (e.g., 1, .01, or .001 cases per person) increases as a power function of the population base in the simplified frequencies (e.g., 1 in 1, 1 in 100, or 1 in 1000).

To test our idea about the costs of representational change in an experimental setting, we examined short- and long-term changes in children's estimates of numerical and fractional magnitudes and the relation between accuracy across the two tasks. Like the relation between the magnitude of a risk and the population base, estimates of fractional magnitudes are interesting because the value of a fraction also increases as a power function of its denominator. Moreover, previous research has shown that this property of fractions poses tremendous difficulties for adults when estimating the value of salaries (Opfer & DeVries, in press). This led us to hypothesize that inaccurate estimates of fractional value might stem from learners automatically transferring their representations of numerical value to the fractional context regardless of the superficial differences across the two contexts and regardless of the gross inaccuracy of such transfer. To examine this issue directly, we gave children feedback as they placed numbers on a number line and then examined whether their subsequent estimates of fractional value were unchanged, better, or worse.

In the next two sections, we briefly review the evidence leading us to predict that accuracy of children's estimates of fractional magnitude will decline with age and experience on the number line task. In the first section, we review evidence suggesting that numerical representations normally transfer robustly across tasks, thereby playing a central role in mathematical thinking in school. Furthermore, we present evidence that the nature of children's initial numerical representations are better suited for comparing fractional values than for comparing values of whole numbers. In the subsequent section, we briefly review evidence on how children's representations of whole numbers vary as a function of experience, age, and sex, and we present a theoretical analysis that links how these same variables should influence accuracy of fractional magnitude estimates. In the same section, we also describe how we tested these predictions in Experiments 1 and 2.

Development of numerical representations and mathematical thinking

Across a wide range of tasks, children normally improve their expectations about the magnitudes of symbolic numerals. For example, on a number line estimation task, children are presented with a series of lines flanked by two numbers (e.g., 0 and 1000), a third number above the line (e.g., 230), and no other markings. When asked to estimate the position of this third number, children's estimates of the position of the number ideally would increase linearly with the actual value of the third number, thereby reflecting representation of the ratio characteristics of the formal decimal system. In fact, however, children's estimates do not increase linearly—at least not initially. On 0–1000 number lines, sixth graders' estimates increase linearly, whereas second graders' estimates increase logarithmically (Siegler & Opfer, 2003). On 0–100 number lines, second graders' estimates increase linearly (Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; Geary, Hoard, Nugent, & Byrd-Craven, in press; Siegler & Booth, 2004; Siegler & Opfer, 2003), whereas kindergartners' estimates increase logarithmically (Siegler & Booth, 2004). On 0–10 number lines, kindergartners' estimates increase linearly, whereas preschoolers' estimates increase logarithmically (Opfer, Thompson, & Furlong, 2007). Moreover, on the inverse position to number task, where children are asked to assign a number to a position on the number line, children's estimates increase as an inverse of the logarithmic function (Siegler & Opfer, 2003).

Theoretically, these initial expectations that numerical magnitudes increase logarithmically are interesting because they are consistent with Fechner's law, that is, the log Gaussian model of magnitude representations implied by the quantitative performance of time-pressured adults, young preschoolers, human infants, and nonhuman animals (Banks & Hill, 1974; Feigenson, Dehaene, & Spelke, 2004; Gallistel & Gelman, 1992; Moyer & Landauer, 1967; Xu & Spelke, 2000). Also in these groups, the difference between 1 and 10 seems larger (or is detected more quickly) than the difference between 101 and 110, much as these numbers would be spaced on a logarithmic ruler. Moreover, single-cell recordings in the parietal cortex of monkeys show a similar pattern of neural activation, with the numerical tuning functions of neurons that fire to large sets (e.g., 8 or 9 dots) showing greater signal overlap than those of neurons that fire to small sets (e.g., 1 or 2 dots) (Nieder, Freedman, & Miller, 2002). Thus, it appears that the natural mental number line is logarithmically scaled, unlike the decimal system that children must eventually learn in school (Dehaene, Dehaene-Lambertz, & Cohen, 1998).

In school-age children, the change from logarithmic to linear representations of numerical quantity is not unique to either microgenetic studies (e.g., Opfer & Siegler, 2007; Opfer & Thompson, 2008) or the number line estimation task. The timing of development in number line estimation coincides with parallel logarithmic to linear changes in numerosity estimation (i.e., generating a set of approximately *N* objects), in measurement estimation (i.e., drawing a line that is approximately *N* units long), and in number categorization (i.e., categorizing *N* as "very small" [e.g., 0] to "very large" [e.g., 1000]), with consistent individual differences emerging across all four estimation tasks (Booth & Siegler, 2006; Opfer & Thompson, 2008). Linearity of number line estimates (measured by the R^2 value of the best fitting linear regression function) also correlates strongly with other tests of school children's understanding of numerical magnitudes, including speed of magnitude comparison (e.g., deciding whether 4 is greater than 6) (Laski & Siegler, 2007), learning of solutions to unfamiliar addition problems (Booth & Siegler, in press), and overall math achievement on standardized tests (*r*s typically between .50 and .60) (Booth & Siegler, 2006; Siegler & Booth, 2004).

These strong correlations between number line estimation and mathematical proficiency should not be surprising given that they both rely on children's representations of magnitude, which are strongly associated with symbolic numbers and, thus, easily activated by them. Therefore, when numerals appear on either a number line task, an arithmetic problem, or a computer screen, it is difficult for children to inhibit the magnitudes associated with the numerical symbol regardless of whether these magnitudes intrude on accuracy of task performance (Berch, Foley, Hill, & Ryan, 1999; Opfer & DeVries, in press; Opfer et al., 2007). These automatic magnitude activations, of course, are also beneficial in that they aid children in generating approximately correct answers to arithmetic problems and in more swiftly rejecting errors that differ largely in magnitude from the correct answers (Ashcraft, 1992; Siegler, 1988). This positive relation between accurate representations of numerical magnitude and math achievement is also an important reason why the standards published by the National Council of Teachers of Mathematics (NCTM) have consistently recommended that improving estimation skills be made a high educational priority (e.g., National Council of Teachers of Mathematics., 2000).

Although automatic activations of linear magnitude representations are beneficial for estimating the value of whole numbers, such representations may be inappropriate for estimating the value of fractions. Specifically, by automatizing that 150 is closer to 1 than to

1000, adults appear to be subject to a powerful cognitive illusion in which 1/150 seems closer to 1/1 than to 1/1000; in contrast, children's belief that 150 is closer to 1000 than to 1 appears to protect them from this illusion (Opfer & DeVries, in press). The costs of automatic magnitude representations can also be seen in how adults approach the problem of estimating the sum of fractions, where the magnitude of the denominator and numerator—but not the magnitude of the fraction itself—is represented automatically. Thus, for example, when estimating the answer to $12/13 + 7/8$ on a National Assessment of Educational Progress, fewer than a third of 13- and 17-year-olds correctly chose 2 from the options 1, 2, 19, and 21 (Carpenter, Corbitt, Kepner, Lindquist, & Reys, 1981). If students had represented the magnitude of 12/13 and 7/8 as each being approximately equal to 1, the answer would have been easy to solve ($1 + 1 = 2$). Instead, roughly half of students answered 19 or 21, indicating that they focused exclusively on numerators ($12 + 7 = 19$) or denominators ($13 + 8 = 21$).

This interpretation of adults' approach to fractions led us to make two predictions about how older children would estimate the value of quantities expressed in fractional notation, for example, when estimating the placement of a salary (e.g., \$1/60 min) on a line that begins with one salary (e.g., \$1/min) and ends with another salary (e.g., \$1/1440 min). The first prediction was that if older children also compare only the value of the denominators in the salary, their linear representation of numbers would lead to radically inaccurate estimates. This inaccuracy is predicted by the fact that the relation between the numeral expressed in the denominator and the magnitude denoted by the whole fraction is provided by a power function rather than by a linear function. For example, although 60 is closer to 1 than to 1440, $k/60$ is closer to $k/1440$ than to $k/1$.

The second prediction was that if younger children also compare denominators, their logarithmic representation of numbers would have a correcting effect and, thereby, lead to more accurate estimates than those of older children. This prediction stems from the fact that the power relation between the value of the denominator and the magnitude of the fraction is somewhat similar to that of a logarithmic function. Thus, for example, the natural logarithm of 60 (4.09) is closer to the natural logarithm of 1440 (7.27) than to the natural logarithm of 1 (0), much as $k/60$ is closer to $k/1440$ than to $k/1$.

Issues examined in current experiments

To test the predicted costs and benefits of representational change, we examined the accuracy of symbolic magnitude estimation across two contexts, whole numbers and fractions, and how these two contexts affected the relations among age, sex, and accuracy. In the next three subsections, we highlight the specific predictions of our hypotheses regarding the process of change in numerical estimation, the relation between numerical and fractional magnitude estimation, and how these two issues can yield insights into sex differences in numerical representations and mathematical proficiency.

Process of change in numerical estimation

How does the logarithmic to linear shift in numerical representations take place with increasing age or experience? An important mechanism implicated in previous work on number line estimation (Opfer & Siegler, 2007; Opfer & Thompson, 2008) is children's use of analogy to structure their generalization of "log discrepant" information.

According to this view, children normally encounter information that does not match their logarithmic representation of numerical magnitudes (e.g., hearing 150 referring to a relatively small part of 1000 items). If children already apply linear representations in some numerical contexts (e.g., for small numerical ranges), such experiences of log discrepancy may lead them to draw analogies between the two contexts and to extend the linear representation to numerical ranges where they previously used logarithmic representations. For example, if a second grader is shown that her estimate of the position of 150 on a 0–1000 number line is too high, and also is shown the correct position of 150 within that range, she may draw the analogy that 150 is to the 0–1000 range as 15 is to the 0–100 range. This analogy may lead her to choose a linear representation for the 0–1000 range on subsequent estimation problems. Moreover, if the analogy is drawn at the level of the entire representation (as opposed to being restricted to numbers near 150), such feedback would lead to more accurate estimates for numbers throughout the 0–1000 range, especially numbers where the log and linear representations differ most dramatically. Such a substitution of representations could occur quite quickly because the linear representation has already been constructed and used in smaller numerical contexts.

What types of experiences would be most likely to stimulate such an analogy? The log discrepancy hypothesis predicts that if children are using a logarithmic representation, the magnitude of change in their estimates in response to feedback should be positively related to the discrepancy between the logarithmic and linear functions for the problems on which the children receive feedback. The discrepancy between logarithmic and linear representations of the values on a 0–1000 number line is illustrated in Fig. 1, with both functions

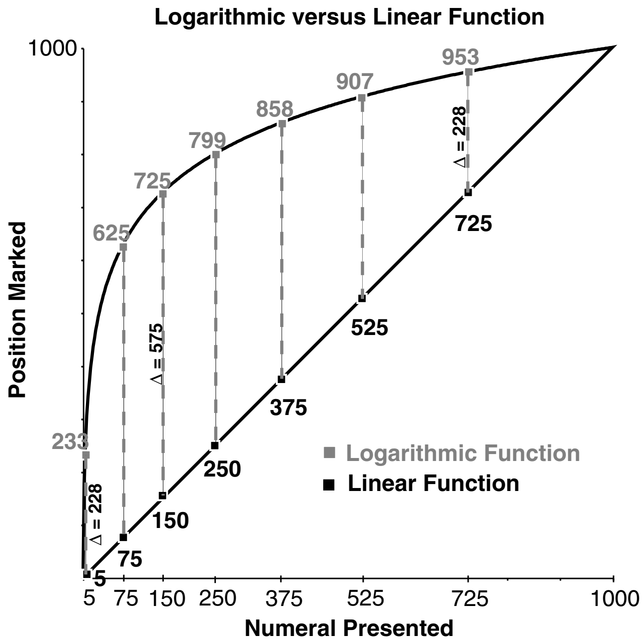

Fig. 1. The discrepancy between logarithmic and linear representations of numerical values on a 0–1000 number line is greatest at 150. The discrepancies for 5 and 725 are equal to each other and approximately half as great as the discrepancy at 150. Linear function: $y = x$; logarithmic function: $y = 144.76 * \ln x$.

constrained to pass through 0 and 1000 (linear function: $y = x$; logarithmic function: $y = 144.76 * \ln x$). As the figure shows, the difference in estimates varies as a function of the number presented. The maximum difference occurs at 150, where the logarithmic representation predicts an estimate of 725 and the linear representation predicts an estimate of 150, resulting in a discrepancy of 575 (57.5% of the line). For purposes of comparison, the absolute numerical discrepancy between the estimates predicted by the linear and logarithmic representations of both 5 and 725 is 228 (22.8% of the line).

Consistent with the log discrepancy hypothesis, Opfer and Siegler (2007) demonstrated that larger discrepancies between children's estimates and the linear function (e.g., feedback on the magnitude of 150) were more likely to provoke representational changes than were smaller discrepancies (e.g., feedback on the magnitudes of 5 and 725). Indeed, after a single feedback trial, the best fitting function switched from logarithmic to linear for 85% of children in the 150–feedback condition and for more than half of all children who received feedback. Moreover, once children's estimates first conformed to the linear function, the linear model continued to provide the best fit on more than 80% of subsequent trial blocks regardless of the problems that led to the apparent switch of representations. This high rate (81%) of children rapidly switching to the linear representation was also found in a subsequent replication by Opfer and Thompson (2008). In addition, learned linear representations were also transferred to a number categorization task in which children were told that 0 is "very small" and 1000 is "very large" and children needed to judge verbally whether numbers were "very small," "small," "medium," "large," or "very large."

The current study offered an opportunity to revisit the log discrepancy hypothesis by looking at both long-term (Experiment 1) and short-term (Experiment 2) changes in numerical estimation. In Experiment 1, we examined age differences in numerical estimation. We were particularly interested in whether age differences in accuracy would be greatest for the placement of numbers around 150, the maximally discrepant point between the logarithmic and linear representations. This was of particular interest to us because the main goal of Experiment 2 was to provide younger children with an experience during the experimental session (e.g., corrective feedback on numbers around 150) that would facilitate learning of a linear representation of numbers during that particular experimental session.

Relation between numerical and fractional magnitude estimation

We next examined whether linear representations of numerical magnitude interfere with estimates of fractional magnitude. To test this hypothesis, we obtained both correlational and causal data.

Correlational evidence for the hypothesis was obtained in Experiment 1 by examining whether accuracy of numerical magnitude estimates was inversely related to accuracy of fractional magnitude estimates. In this experiment, we were interested in (a) whether a *positive* relation between age and accuracy in numerical magnitude estimation would coexist with a *negative* relation between age and accuracy in fractional magnitude estimation and (b) whether individual differences in accuracy of numerical magnitude estimates were negatively correlated with accuracy of fractional magnitude estimates within each age group.

Causal evidence for the hypothesis was obtained in Experiment 2 by examining transfer of learning from the numerical magnitude estimation context to the fractional magnitude estimation context. Transfer of learning from one context to another context is notoriously difficult to elicit (for a review, see Barnett & Ceci, 2002), but previous evidence of broad and robust transfer of numerical representations (Laski & Siegler, 2007; Opfer & Thompson, 2008) lends support to the idea that transfer of numerical representations is automatic. To examine this issue directly, we examined children's transfer of numerical representations to the fractional magnitude context, where linear representations would generate inaccurate task performance. In Experiment 2, we specifically wanted to determine whether children who learned to adopt the linear representation after receiving training on the number line task would transfer the more mature representation to their estimation performance on a fractional units task.

Sex differences in numerical and fractional magnitude estimation

Given the centrality of numerical representations to mathematical thinking, and given widespread public speculation about sex differences in quantitative abilities, we also examined whether boys and girls differed in their estimates of numerical and fractional magnitudes.

Typically, differences between boys' and girls' quantitative performance vary greatly depending on the variable measured and children's ages (Halpern et al., 2007; Hyde, Fennema, & Lamon, 1990). For instance, girls tend to outperform boys during early elementary school years on quantitative tasks related to verbal abilities and curriculum content, whereas boys tend to outperform girls from Grades 4 through 12 when quantitative tasks involve visuospatial concepts, reasoning about real-world problems, and estimating answers to arithmetic problems such as 76×89 (Doolittle & Cleary, 1987; Dowker, Flood, Griffiths, Harriss, & Hook, 1996; Geary, 1996; Hyde et al., 1990; Levine, Huttenlocher, Taylor, & Langrock, 1999; Willingham & Cole, 1997). Given the correlation between these latter measures and numerical estimation performance (Booth & Siegler, 2006; Siegler & Booth, 2004), we wondered whether sex differences might also exist in numerical estimation, possibly due to boys possessing more linear representations of numerical magnitude than girls.

Theoretically, sex differences in numerical representation would be interesting for at least three reasons. First, a sex difference has been predicted on the basis of neurophysiology (Halpern et al., 2007). There is now considerable evidence that the brain represents numerical quantity in at least two regions over the course of development, the prefrontal cortex and the inferior parietal cortex (Ansari, Nicolas, Lucas, Hamon, & Dhital, 2005; Dehaene, Piazza, Pinel, & Cohen, 2005; Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999; Nieder et al., 2002; Pinel, Piazza, LeBihan, & Dehaene, 2004; Rivera, Reiss, Eckert, & Menon, 2005), and sex differences in the function and architecture of these regions have been reported in studies of human and nonhuman animals (Goldstein et al., 2001; Kavaliers, Ossenkopp, Galea, & Kolb, 1998; Knops, Nuerk, Sparing, Foltys, & Wilmes, 2006). For example, in human adults, Goldstein and colleagues (2001) found that, adjusting for overall brain volume, the inferior parietal lobe was 20% larger in males than in females. What has remained unclear is whether these sex differences in the architecture and function of parietal cortex make a difference in *early* numerical representations such as those tapped by the number line task. Interestingly, Halpern and colleagues (2007) noted that, to the extent that these regions are larger in males than in females, a male advantage is predicted for the mental number line. To examine this prediction, we compared the accuracy and linearity of boys' and girls' estimates on a number line.

Second, if sex differences exist at the level of automatic numerical representations, such sex differences should have costs as well as benefits. Specifically, sex differences in the accuracy of symbolic magnitude estimates should be complementary, with the sex that performs more accurately in the numerical magnitude context performing less accurately in the fractional magnitude context. The costs of representational change are well understood in the field of perceptual learning (e.g., Petrov & Anderson, 2005), but to our knowledge this consideration has not figured at all in the discussion of how basic sex differences might contribute to mathematical proficiency. In the extreme, an implication of our theoretical analysis is that robust sex differences in numerical representations might not confer any overall advantage in accuracy but could nevertheless confer some large task-specific advantages to both sexes.

Finally, as Newcombe, Mathason, and Terlecki (2002) observed, although it is scientifically interesting to document sex differences, the more interesting question is how experience affects these sex differences. For example, Opfer and Siegler (2007) and Opfer and Thompson (2008) showed that even a small amount of focused training can virtually eliminate age differences in numerical estimation. We wondered whether the same might be true of sex differences as well. To examine this issue, we looked at sex differences in the accuracy and linearity of estimates prior to, during, and after the course of training.

Experiment 1: Long-term changes in numerical and fractional magnitude estimation

Experiment 1 had three major purposes. One was to replicate the finding that children initially generate estimates that increase logarithmically with numerical value. This goal was important because these children subsequently participated in a microgenetic study of the transition from use of a logarithmic representation to use of a linear representation in numerical estimation (Experiment 2). The second goal was to test whether the greatest improvement between first and third grade occurred for numbers around 150. This goal was important both because of the theoretical prediction that the greatest improvement with age should come in this area, where the logarithmic and linear functions are most discrepant, and because we later provided feedback for estimates in this region. The third purpose of Experiment 1 was to examine whether increasing accuracy in numerical estimation was accompanied by decreasing accuracy in fractional estimation. This test was the most important because it would disconfirm our hypotheses about the relation between numerical and fractional magnitude estimation.

Method

Participants

Participants were 64 first through third graders (mean age = 8.41 years, $SD = 0.75$, 37 girls and 27 boys) who attended neighborhood schools in largely European American, middle-class suburbs surrounding a large metropolitan city in the midwestern United States. One of two female research assistants served as the experimenter.

Tasks

Number line estimation task

All number line problems (Fig. 2A) consisted of a 20-cm line with the left endpoint labeled 0, the right endpoint labeled 1000, and the number to be estimated appearing 2 cm above the midpoint of the number line. The experimenter gave the follow instructions:

> Today we're going to play a game with number lines. We use number lines to help us with math. It looks just like a line with numbers at each end. It shows us where all the numbers in between go. Different numbers go in different places on a number line. In this game, there will be a number in a circle up here. Your job is to show me where that number goes on a number line like this one. Each number line will have a 0 at one end and 1000 at the other end. When you decide where the number goes, I want you to make a mark through the line like this.

Before the first item, the experimenter said, "This number line goes from 0 at this end to 1000 at this end. If this is 0 and this is 1000, where would you put *N*?" Participants were then asked to place numbers (2, 5, 18, 27, 34, 42, 56, 78, 100, 111, 122, 133, 147, 150, 156, 162, 163, 172, 179, 187, 246, 306, 366, 426, 486, 546, 606, 666, 722, 725, 738, 754, 818, 878, and 938) on a number line by making a hatch mark. Each number line problem appeared on its own page. These numbers maximized the discriminability of logarithmic and linear functions by oversampling the low end of the range, minimized the influence of specific knowledge (e.g., that 500 is halfway between 0 and 1000), and tested predictions about the range of numbers where estimates would differ most with age.

Fraction line estimation task

Participants were asked to estimate the total amount of money a person would make at a given salary (e.g., \$1/60 min) by placing a mark on a 20-cm "money line" (Fig. 2B). Both

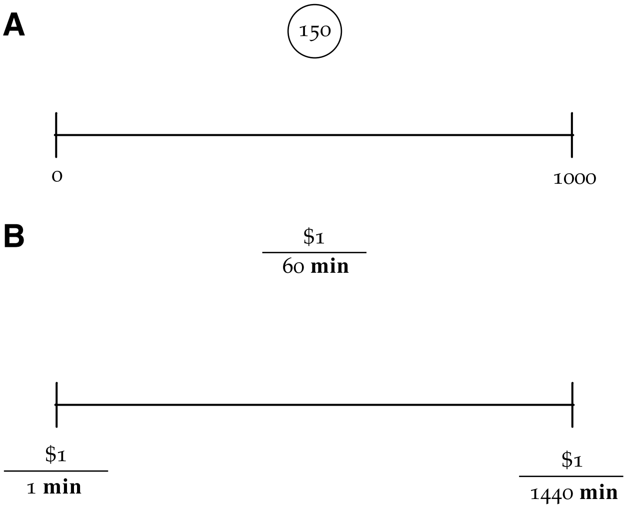

Fig. 2. Experimental stimuli: sample number line estimation problem (A) and sample fraction line estimation problem (B).

the value of the salaries to be estimated (e.g., $ 1/60 min) and the endpoints of the money line ($1/1 min and $1/1440 min) were expressed in fractional units. For this task, the experimenter gave the following instructions:

> Today we're going to play a game with money lines. We use money lines to tell us how much money a person makes. It looks just like a line with amounts of money at each end. It shows us where all the amounts of money in between go. Different amounts of money go in different places on a money line. In this game, there will be an amount of money that a person might make up here [pointing to the top of the blank money line data sheet where the amount of money should be]. Your job is to show me where that amount of money goes on a money line like this one. Each money line will have $1 every minute at one end and $1 every 1440 minutes at the other end. When you decide where the amount of money goes, I want you to make a mark through the money line like this.

Children were asked to estimate the location of the following salaries on the money line: $1/2 min, $1/8 min, $1/9 min, $1/60 min, $1/120 min, $1/240 min, $1/360 min, $1/480 min, $1/540 min, and $1/720 min. Again, every problem appeared on its own page. All fractions were read aloud to children (e.g., "where would you put 1 dollar every 2 minutes?").

Design and procedure

To ensure that participation in the fractional units task did not affect subsequent performance on number line problems or the fractional units posttest in Experiment 2 (cf. Solomon & Lessac, 1968), children were randomly assigned to two groups. One group ($n = 35$) received the fractional units task first, and the other group ($n = 29$) did not receive the task. All participants then received 22 number line estimation problems. (Because we found no effect of giving the fractional units task on subsequent performance, we combined these two groups for all subsequent analyses.) Participants were tested during a single session. The items within each scale were randomly ordered, separately for each child, and presented in small workbooks, one problem per page.

Results

Age group differences in numerical estimation

We first examined age differences in the accuracy of numerical estimates. To measure accuracy, we converted the magnitude estimate for each number (the child's hatch mark) to a numerical value (the linear distance from the 0 mark to the child's hatch mark), divided the result by the total length of the line, and then multiplied the result by 1000. The magnitude of each child's error was calculated by taking the mean absolute difference between each of the child's estimated values and the actual values. As expected, the mean absolute error declined with age, $r(63) = -.56$, $p < .001$, decreasing from 25% for the younger half of the sample (7- to 8.49-year-olds, $n = 30$, 21 girls and 9 boys) to 11% for the older half of the sample (8.5- to 9.5-year-olds, $n = 34$, 16 girls and 18 boys), $F(1, 63) = 41.75$, $p < .0001$, $d = 1.64$.

To determine whether this improvement in accuracy was associated with the hypothesized logarithmic to linear shift, we next compared the fit of the best fitting linear and logarithmic functions with the median numerical estimates of the younger children and older children. As in previous studies of this age range (Opfer & Siegler, 2007; Siegler & Opfer, 2003), the fit of the linear function to children's estimates increased, whereas the fit of the logarithmic function decreased. The median estimates of the younger group were better fit by the logarithmic function ($R^2 = .95$) than by the linear function ($R^2 = .63$), whereas the median estimates of the older children were better fit by the linear function ($R^2 = .98$) than by the logarithmic function ($R^2 = .74$) (Fig. 3).

We next tested whether the logarithmic/linear characterization of median estimates reflected individual children's performance by comparing individuals' estimates on each task against the predictions of the best fitting linear and logarithmic functions. We assigned a 1 to participants when the linear model provided the best fitting function to their estimates and a 0 to participants when the logarithmic model provided the best fitting function. (Because the degrees of freedom were identical for these two models, the simple comparison of R^2 values was appropriate.) The linear function provided the better fit for 17% of the younger group and 79% of the older group (Fig. 4), whereas the logarithmic function provided the better fit for 83% of the younger children and 21% of the older children. To test the association of age with generation of linear estimates more precisely, we used a logistic regression model to test for the effect of age on the odds of generating linear estimates, where age was entered as a continuous variable (range: 7.15–9.64 years). The test indicated that there was a significant positive effect of age, with children in this sample being 5.46 times more likely to generate linear estimates with each year of age, $\hat{\beta} = 1.87$, $z = 3.99$, Wald(1, $N = 64$) = 22.40, $p < .0001$. The strength of the age effect differed somewhat for boys and girls. Boys were 3.47 times more likely to generate linear estimates with each year of age, $\hat{\beta} = 1.50$, $z = 2.25$, Wald(1, $N = 27$) = 6.07, $p < .05$, whereas girls, due to their lower starting point, were 7.53 times more likely to generate linear estimates with

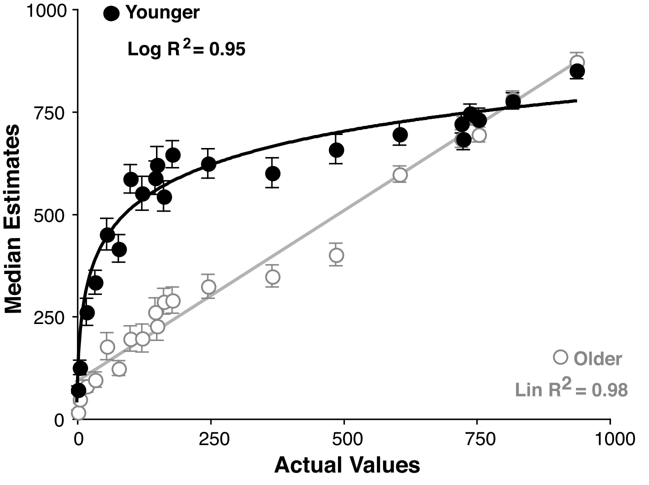

Fig. 3. Experiment 1: Long-term changes in numerical magnitude estimation. The estimates of 7- to 8.49-year-olds were better fit by a logarithmic function than by a linear function, whereas the estimates of 8.5- to 9.5-year-olds were better fit by a linear function than by a logarithmic function.

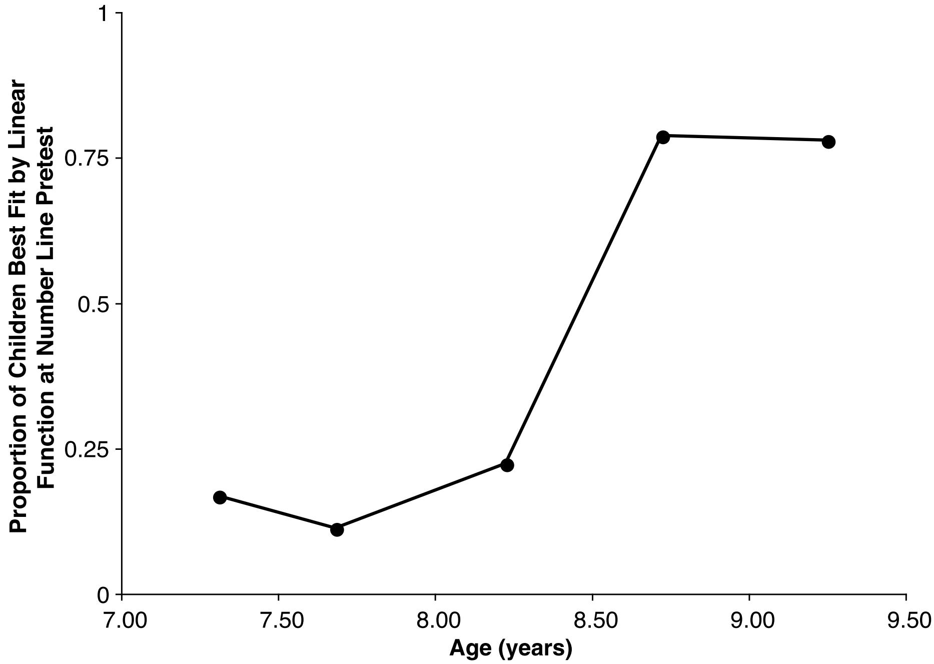

Fig. 4. Experiment 1: The proportions of children whose estimates were best fit by a linear function increased substantially between 7.5 and 8.5 years of age.

each year of age, $\hat{\beta} = 2.14$, $z = 2.97$, Wald(1, $N = 37$) = 13.70, $p < .005$. Thus, data from multiple levels of analysis—individual children, boys and girls, younger and older—indicated a logarithmic to linear shift in the 0–1000 numerical context.

Breadth of age group differences in numerical estimation

We next examined whether the greatest improvements in estimates occurred on numbers around 150, where the discrepancy between the logarithmic and linear functions was greatest (Fig. 1) and where children would later receive feedback in Experiment 2. This analysis was important because it is possible, for example, for children to have a linear representation with a very high or low slope, thereby affecting where the maximum discrepancy would actually occur. To examine improvement with age over the numbers tested, we used the absolute age differences between each age group's median estimate for each number and the correct value for the number.

From these, we correlated the absolute numerical distance of each to-be-estimated number from 150 with the absolute difference in estimates between younger and older children on that number. Improvement in estimation accuracy proved to be highly correlated with distance from 150: $r(21) = -.71$, $p < .001$; the closer the number to 150, the greater the improvement with age. We then examined estimates for a fixed numerical range (32) around three anchors of interest: 150 (where the discrepancy in estimates is greatest between the logarithmic and linear functions), 725 (where the discrepancy is 40% of the discrepancy at 150), and 5 (where the discrepancy is also 40% of the discrepancy at

150). The stimulus set included four numbers in each of these three numerical ranges. As anticipated by the log discrepancy hypothesis, a one-way analysis of variance (ANOVA) indicated differences among these three ranges, $F(2, 11) = 25.55$, $p < .001$. Post hoc analyses indicated that age-related improvements in estimation accuracy were greater for the four numbers around 150 than for either the four numbers around 725, $t(3) = 10.72$, $p < .01$, $d = 7.06$, or the four numbers around 5, $t(3) = 3.35$, $p < .05$, $d = 2.48$. In contrast, younger and older students' estimates around 5 and 725 did not differ significantly from each other, $t(3) = 2.39$, *ns.*

Relation between numerical and fractional magnitude estimation

We next examined whether younger children provided more accurate estimates of fractional magnitude than did older children and, if so, whether this change was associated with performance on the numerical estimation task.

To measure accuracy on the fraction line task, we converted each estimate to a numerical value by measuring the distance between the child's estimate and the origin of the scale (0–20 cm) divided by the length of the scale (20 cm). The magnitude of error for each estimate (0–1) was obtained by taking its absolute difference from the correct placement of the fraction on the scale (0–1), and accuracy was obtained by subtracting the error from 1. As predicted by the representational change hypothesis, age was negatively related to accuracy, $r(63) = -.43$, $p < .001$, with younger children's accuracy (48%) being significantly higher than older children's accuracy (38%), $F(1, 63) = 9.81$, $p < .01$, $d = 0.73$.

We next performed a linear regression to examine whether each child's accuracy on the number line task (0–100%) predicted accuracy on the fraction line task (0–100%). The relation between the two tasks was very strong and negative ($r = -.80$), $F(1, 34) = 57.93$, $p < .0001$, indicating that 63.7% of variation in accuracy of fractional magnitude estimation was accounted for by inaccuracy in numerical magnitude estimation (Fig. 5). Because age alone accounted for a marginally significant amount of variation in accuracy of fractional magnitude estimates ($R^2 = .10$), $F(1, 33) = 3.80$, $p = .06$, we next entered the age variable in the regression model, and we found that the combination of age and accuracy on the number line task accounted for only an additional 1.3% of variance ($R^2 = .65$), which was not a significant addition to the model, F change $= 1.19$, *ns.* The full regression model with age, accuracy of numerical magnitude estimation, and interaction between the other two variables accounted for 69.5% of variance in accuracy of fractional magnitude estimation.

Why might inaccurate numerical magnitude estimation reliably predict accurate fractional magnitude estimation? According to the representational change hypothesis, this relation stems from the particular pattern of errors that children are likely to make when estimating numerical value; that is, children's errors in numerical estimation are not random but rather generated by their reliance on a logarithmic representation, which generates estimates that are somewhat similar to the power function relating the value of a fraction to its denominator. The similarity of the two functions is apparent in Fig. 6, which depicts children's fractional magnitude estimates against the denominator. Ideally, estimates of the value of the fraction (y) should initially decrease dramatically as the denominator increases in magnitude (i.e., $y = 1/x$). Younger children's representation of the denominator also leads them to generate estimates in a way that is somewhat similar to this pattern, with their estimates of fractional value decreasing logarithmically with the

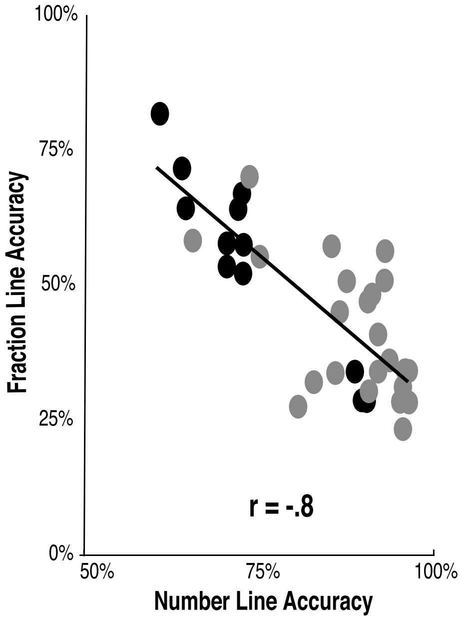

Fig. 5. Experiment 1: Relation between accuracy on number line task and fraction line task. The darkness of circles depicts age group of participants (dark = 7–8.49 years, light = 8.5–9.5 years).

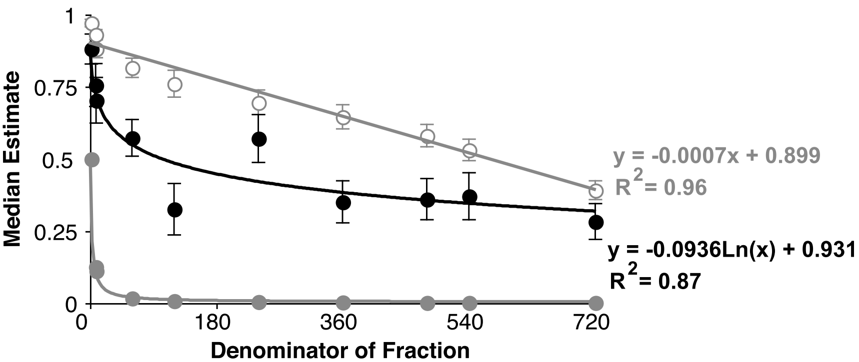

Fig. 6. Experiment 1: Long-term changes in fractional magnitude estimation. The estimates of 7- to 8.49-year-olds (dark circles) were better fit by a logarithmic function than by a linear function, whereas the estimates of 8.5- to 9.5-year-olds (light circles) were better fit by a linear function than by a logarithmic function. The ideal pattern of performance ($y = 1/x$) is depicted in gray.

value of the denominator (log $R^2 = .87$). In contrast, older children's representation leads them to generate a less accurate pattern, with their estimates of fractional value decreasing linearly with the value of the denominator (lin $R^2 = .96$).

To test this explanation more directly, we next examined accuracy in fractional magnitude estimation as a function of the logarithmicity and linearity of children's numerical magnitude estimation. The first relevant evidence came from the fit of the linear and logarithmic R^2 values associated with each child's numerical magnitude estimates. As hypothesized by the representational change hypothesis, each of these variables accounted for a significant amount of variation in children's fractional magnitude estimates. As children's numerical estimates grew more logarithmic, their fractional estimates increased in accuracy ($r = .45$), $F(1, 33) = 8.43$, $p < .0001$. As children's numerical estimates grew more linear, their fractional estimates decreased in accuracy ($r = .76$), $F(1, 33) = 44.25$, $p < .0001$.

Sex differences in numerical versus fractional magnitude estimation

We then examined whether boys and girls differed in the accuracy of their numerical and fractional magnitude estimates (Fig. 7). Using the same measures reported above, boys' numerical magnitude estimates were more accurate on average ($M = 88\%$, $SD = 10\%$) than girls' estimates ($M = 78\%$, $SD = 10\%$), $F(1, 63) = 14.78$, $p < .0005$, $d = 1$. In contrast, girls' fractional magnitude estimates tended to be more accurate on average ($M = 50\%$, $SD = 16\%$) than boys' estimates ($M = 40\%$, $SD = 13\%$), $F(1, 34) = 3.66$, $p = .06$, $d = 0.69$. Again, differences in accuracy reflected a logarithmic to linear shift in the 0–1000 context. Boys' numerical estimates (average lin $R^2 = .81$, $SD = .24$) were more linear than girls' estimates (average lin $R^2 = .60$, $SD = .22$), $F(1, 63) = 13.80$, $p < .0005$, $d = 0.91$, and boys were less likely than girls to provide estimates best fit by the logarithmic function (30% vs. 65%), $\chi^2 = 7.75$, $p < .01$.

Sex differences were chiefly evident for younger children; among younger children, a greater proportion of boys generated linear estimates than did girls, 44.4% versus 4.8%,

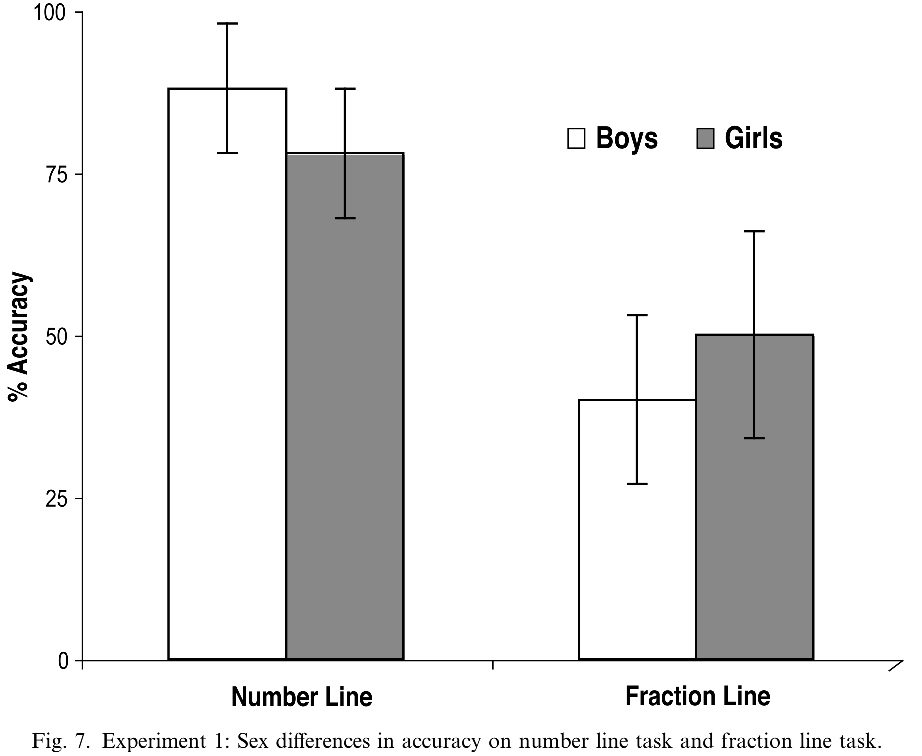

Fig. 7. Experiment 1: Sex differences in accuracy on number line task and fraction line task.

Fisher's $p = .02$, whereas among older children, boys and girls generated linear estimates at similar rates, 83.3% versus 75.0%, *ns.* Whether differences between boys and girls in this sample indicated a developmental delay for girls' numerical and fractional magnitude estimation or a stable sex difference is an issue that we explored further in Experiment 2.

Experiment 2: Short-term changes in numerical and fractional magnitude estimation

Experiment 2 was designed to provide a more direct test of the representational change hypothesis regarding accuracy of numerical and fractional magnitude estimation. In the first section, we examined changes in children's numerical magnitude estimates in response to feedback on numbers around 150 or in response to answering the same problems without feedback. We predicted that feedback on numbers around 150 would elicit a large change in the proportion of children best fit by the logarithmic function (because this is the area of maximum discrepancy between logarithmic and linear representations) and that the change would involve a broad range of numbers and would occur abruptly rather than gradually (because the change involved a choice of a different representation rather than a local repair to the original representation). In the second section, we tested the hypothesis that feedback on accuracy of numerical magnitude estimates would lead to increasing accuracy of numerical magnitude estimates but decreasing accuracy of fractional magnitude estimates (again because change was hypothesized to involve substituting linear representations of numbers for logarithmic ones). In the subsequent section, we revisited sex differences in numerical and fractional magnitude estimation after children had received feedback. We were particularly interested in whether conditions that substantially reduced age differences in estimation would also reduce sex differences.

Method

Participants

Participants were the same children who participated in Experiment 1. For convenience, we divided them into younger children (7- to 8.49-year-olds, $n = 30$, 21 girls and 9 boys) and older children (8.5- to 9.5-year-olds, $n = 34$, 16 girls and 18 boys), as we had done in Experiment 1. One of two female research assistants served as the experimenter.

Tasks

The number line and fraction line estimation tasks described in Experiment 1 were also used in Experiment 2.

Design and procedure

Immediately after Experiment 1, children were randomly assigned to one of two groups; one group received feedback during the training phase (treatment group, $n = 32$, 16 younger and 16 older children), whereas the other group did not receive feedback (control group, $n = 32$, 14 younger and 18 older children). Boys and girls were equally divided between the two conditions (treatment: 18 girls and 14 boys; control: 19 girls and 13 boys).

As shown in the outline of the procedure in Table 1, children in both groups completed the number line estimation task for a pretest, three training trial blocks, and a posttest. The purpose of these three phases (pretest, training trial blocks, and posttest) was to examine the course of learning prior to posttest (i.e., to examine changes from the number line pretest through posttest) and to ensure that learning had occurred prior to the transfer task (i.e., the fractional estimation task). On the number line pretest and posttest, children in the treatment and control groups were presented with the same 22 problems without feedback (i.e., without treatment). For children in the treatment group, each training trial block included a feedback phase and a test phase. As shown in Table 1, the feedback phase of each training trial block consisted of either 1 item on which children received feedback (Trial Block 1) or 3 items on which they received feedback (Trial Blocks 2 and 3). The test phase in all three training trial blocks consisted of 10 items on which children did not receive feedback; this test phase occurred immediately after the feedback phase in each training trial block. Children in the control group received the same number of estimation problems, but they never received treatment. On the posttest, children in all four groups were presented with the same 22 problems without feedback as in Experiment 1. The children's estimates in Experiment 1 provided pretest data that was used as a point of comparison for their subsequent performance and was elicited during the same session.

Feedback was administered to the treatment group following the same procedure used in Opfer and Siegler (2007) and Opfer and Thompson (2008). The treatment procedure was as follows. On the first feedback problem, children were told, "After you mark where you think the number goes, I'll show you where it really goes so you can see how close you were." After each child answered, the experimenter took the page from the child and superimposed on the number line a 20-cm ruler (hidden from the child) that indicated the location of every 10th number from 0 to 1000. Then the experimenter wrote the number corresponding to the child's mark ($N_{estimate}$) above the mark and indicated the correct location of the number that had been presented (N) with a hatch mark. For example, if the child was asked to mark the location for 150 (i.e., N) and his estimate corresponded to the actual location of 600 (i.e., $N_{estimate}$), the experimenter would write the number 600 above the child's mark and mark where 150 would go on the number line. After this, the experimenter showed the corrected number line to the child. Pointing to the child's mark, the experimenter said, "You told me that N would go here. Actually, this is where N goes

Table 1
Outline of procedure in Experiment 2

Experimental group	Training phase						Posttest	
	Number line task						Number line task	Fraction line task
	Feedback	Test	Feedback	Test	Feedback	Test		
Treatment group	150	10 items (0–1000)	3 items (147-187)	10 items (0–1000)	3 items (147-187)	10 items (0–1000)	22 items (0–1000)	10 items (0–1000)
Control group	None	10 items (0–1000)	None	10 items (0–1000)	None	10 items (0–1000)	22 items (0–1000)	10 items (0–1000)

[pointing]. The line that you marked is where N_{estimate} actually goes." When children's answers deviated from the correct answer by no more than 10%, the experimenter said, "You can see these two lines are really quite close." When children's answers deviated from the correct answer by more than 10%, the experimenter asked children to explain the feedback given, "That's quite a bit too high (or too low). You can see these two lines [the child's and experimenter's hatch marks] are really quite far from each other." Regardless of whether answers were close or far from accurate, explanations for correct answers were elicited (see also Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Siegler, 2002).

Results

We organized our results into three sections. In the first section, we report on the conditions that led to changes in numerical estimation (i.e., source of change), how quickly those changes occurred (i.e., rate of change), and approaches that children used up to and following the use of mature approaches (i.e., path of change). In the second section, we report on results testing our hypotheses about transfer of learning to fractional estimation. In the third section, we report on results concerning sex differences in estimation after training.

Process of change in numerical magnitude estimation

Source of change

We first examined the source of change in estimation performance on the number line task. Specifically, we wanted to test whether the experiences that children received during the training phase of the experiment improved their estimation accuracy on posttest and influenced the degree to which those estimates came to follow a linear function. To find out, we first examined posttest accuracy (0–1) as a function of age and condition (treatment or control). As expected, a linear regression indicated a main effect of age, $F(1, 60) = 29.40$, $p < .0001$, a main effect of treatment, $F(1, 60) = 5.14$, $p < .05$, and an interaction between age and treatment, $F(3, 60) = 3.85$, $p = .05$. To examine this interaction more closely, we looked at changes in younger and older children separately.

Among younger children, we first examined the effect of treatment on estimation accuracy by calculating the mean absolute error for each child and then performing a 2 (Condition: treatment or control) $\times$ 2 (Test Phase: number line pretest or number line posttest) repeated measures ANOVA on the error scores. As expected, there was a main effect of test phase, $F(1, 28) = 19.73$, $p < .0001$, a trend toward a main effect of feedback, $F(1, 28) = 3.52$, $p = .07$, and a significant interaction between test phase and feedback, $F(1, 28) = 6.99$, $p < .01$. For the children in the treatment group, the mean absolute error declined from 24% ($SD = 9\%$) to 15% ($SD = 6\%$), $t(31) = 4.45$, $p < .001$, $d = 1.18$, whereas for the children who were in the control group, the mean absolute error did not differ by test phase (number line pretest: $M = 27\%$, $SD = 10\%$; number line posttest, $M = 24\%$, $SD = 11\%$). Finally, the treatment group's posttest mean absolute error was significantly lower than the control group's posttest mean absolute error ($M = 24\%$, $SD = 9\%$), $F(1, 29) = 8.35$, $p < .01$, $d = 1.18$. Thus, among younger children, feedback had a large effect on short-term changes in the accuracy of numerical magnitude estimates.

We next examined whether younger children's short-term changes in estimation accuracy were also accompanied by the hypothesized logarithmic to linear shift. On pretest, young children in both the treatment group and the control group initially provided median estimates for each number that were in fact fit better by the logarithmic regression function than by the linear one (Fig. 8). The precision of the fit of the logarithmic function and the degree of superiority of that function to the linear function were similar across the treatment group (log $R^2 = .92$, lin $R^2 = .58$) and control group (log $R^2 = .93$, lin $R^2 = .65$). In contrast, the treatment and control groups differed considerably in their number line posttest estimation patterns (Fig. 8). Children in the control group continued to generate estimates that fit the logarithmic function better than the linear one (log $R^2 = .91$, lin $R^2 = .78$). In contrast, children in the treatment group generated posttest estimates that fit the linear function substantially better than the logarithmic one (lin $R^2 = .91$, log $R^2 = .75$).

To determine whether the fit of the two functions merely arose from aggregating data over individual estimates, we also performed the same analyses for each individual participant's set of estimates. As expected, before children received any training, the majority of children (83%) provided estimates that were better fit by the logarithmic function than by the linear one regardless of whether they later received treatment; here 88% of children generated logarithmic estimates in the treatment group, whereas 79% of children generated logarithmic estimates in the control group, Fisher's exact probability test, $p = .62$, *ns.* Furthermore, posttest estimates also indicated that treatment led to more children providing linear estimates; here 69% of children who were in the treatment groups provided more linear than logarithmic estimates, whereas only 29% of children who were in the control groups provided more linear than logarithmic estimates, $\chi^2(1) = 4.82$, $p < .05$. Thus, as in Opfer and Siegler (2007) and Opfer and Thompson (2008), feedback on a very small (but strategic) set of estimation problems led to large changes in estimation accuracy.

We next examined changes in older children's estimation performance by performing a 2 (Test Phase: pretest or posttest) × 2 (Condition: treatment or control) repeated measures ANOVA on the error scores. We found a main effect of test phase, $F(1, 32) = 7.86$, $p < .01$, and a significant interaction between test phase and condition, $F(1, 32) = 6.50$, $p < .05$. For children in the control group, a paired t test indicated that error scores did not differ from pretest ($M = 11\%$, $SD = 8\%$) to posttest ($M = 11\%$, $SD = 9\%$), $t(17) = 0.35$, *ns.* For

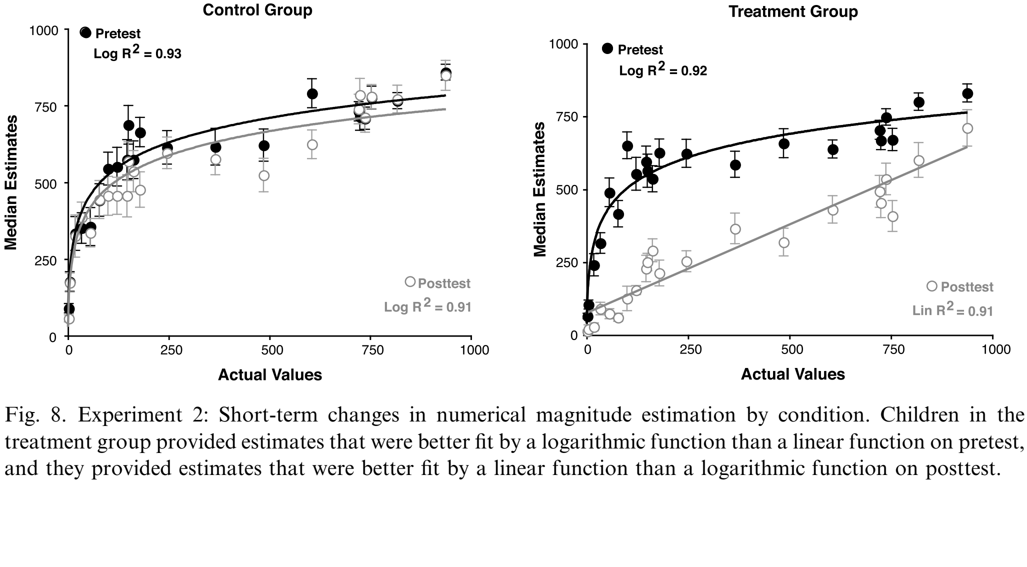

Fig. 8. Experiment 2: Short-term changes in numerical magnitude estimation by condition. Children in the treatment group provided estimates that were better fit by a logarithmic function than a linear function on pretest, and they provided estimates that were better fit by a linear function than a logarithmic function on posttest.

children in the treatment group, however, error scores declined significantly from pretest ($M = 12\%$, $SD = 8\%$) to posttest ($M = 7\%$, $SD = 5\%$), $t(15) = 2.73$, $p < .05$, $d = 0.75$. Unlike younger children's estimates, older children's improvement in accuracy did not stem from their estimates becoming better fit by the linear function. For the treatment group, results of a paired *t* test indicated that linearity of individual children's pretest estimates ($M = .83$) did not differ from posttest estimates ($M = .90$), $t(15) = 1.42$, *ns.* For the control group, linearity of pretest ($M = .83$) and posttest ($M = .83$) estimates also did not differ.

In summary, short-term gains in accuracy of children's numerical estimates came from feedback on a small set of estimates, which had a larger effect on younger children's (logarithmic) estimates than on older children's (linear) estimates. The short-term gains of younger children in response to treatment were also quite impressive; by posttest, treatment had boosted younger children's accuracy to levels comparable to the accuracy of older children who received no treatment, 15% versus 11%, $F(1, 33) = 2.80$, *ns.*

Rate of change

To address the rate of change in numerical estimation, we used logistic regression to examine the relation between generation of more linear than logarithmic patterns of estimates (linear model fitting best or not) and number of trial blocks of treatment (0–4), where 0 corresponded to the trial block prior to the administration of the treatment and, thus, 0 trials of treatment (feedback). First, we examined the effect of trial block for the treatment group of younger children. There was a significant positive effect of trial block for the treatment group, indicating that with each additional trial block the likelihood of generating linear estimates was 1.44 times greater than the previous one, $\hat{\beta} = .37$, $z = 2.18$, Wald$(1, N = 80) = 5.02$, $p < .05$. A similar analysis found no significant effect of trial block for the younger children in the control group, indicating that time on task did not elicit change, nor did we find significant effects for older children in either group (largely because they were already very likely to generate linear estimates).

To put this rate of change into context, it is useful to compare the average incremental change in a single trial block of training ($\times$ 1.44) with the average incremental change found with a year of real-life experience in Experiment 1 ($\times$ 6.46). Taken literally, the comparison suggests that four trial blocks of training accomplished as much as nearly 11 months of real-world experience. Of course, the caveat that must be raised is that it is not at all clear whether the change occurred at a constant rate. Suggesting that the rate was not constant (at least over trial blocks), we observed a sixfold increase in the proportion of younger children in the treatment group best fit by the linear function from Trial Block 0 (13%) to Trial Block 1 (81%), consistent with very rapid and abrupt learning. In the next subsection, we examined the abruptness of change more directly.

Path of change

Younger children could have moved from a logarithmic representation to a linear one via several paths. To examine which path(s) they actually took, we examined the fit of the linear regression function to each individual child's numerical estimates as a function of the number of trials that elapsed since the linear function provided a better fit than did the logarithmic one (i.e., when the logarithmic to linear shift was thought to occur). To measure this, we identified the first trial block on which the linear function provided the best fit to a given child's estimates, and we labeled it Trial Block 0. The trial block immediately before each child's Trial Block 0 was that child's Trial Block –1, the trial block before that was the child's Trial Block –2, and so on.

These assessments of the trial block on which children's estimates first fit the linear function made possible a backward trials analysis that allowed us to test alternative hypotheses about the path of change from a logarithmic representation to a linear one. One hypothesis, suggested by incremental theories of representational change (Brainerd, 1983), was that the path of change entailed gradual continuous improvements in the linearity of estimates (and, thus, the fit of the linear regression function to their estimates). According to this hypothesis, the fit of the linear model would have increased gradually from Trial Block –3 to Trial Block 3. In this scenario, Trial Block 0 (the first trial block in which the linear model provided the better fit) would simply mark an arbitrary point along a continuum of gradual trial block to trial block improvement rather than the point at which children first chose a different representation.

A second hypothesis was that the path of change involved a discontinuous switch from a logarithmic representation to a linear one with no intermediate state. This would have entailed no change in the fit of the linear model from Trial Block –3 to Trial Block –1, a large change from Trial Block –1 to Trial Block 0, and no further change after Trial Block 0. This second hypothesis clearly fit the data. As illustrated in Fig. 9, from Trial Block –3 to Trial Block –1, a one-way ANOVA on the linear regression function, *F*(4,

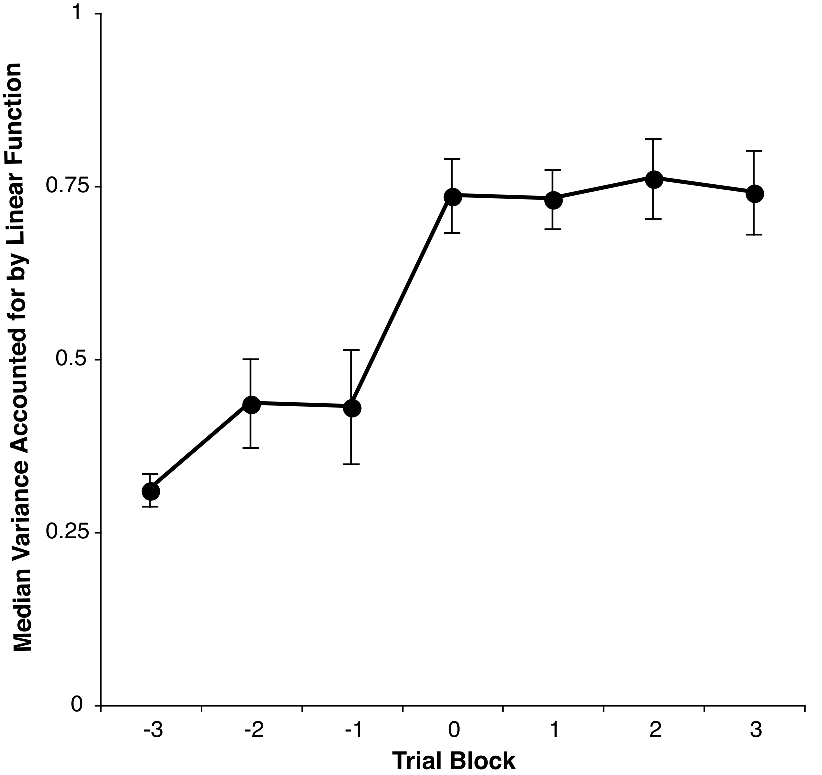

Fig. 9. Experiment 2: Trial block to trial block changes in numerical magnitude estimation.

56) = 0.74, $p > .05$, *ns,* indicated that there was no change in the fit of the linear function across these trial blocks. There also was no change from Trial Block 0 to Trial Block 3 in the fit of the linear function, $F(4, 92) = 0.80$, $p > .05$, *ns.* However, from Trial Block –1 to Trial Block 0, there was a large increase in the fit of the linear function to individual children's estimates (median $R^2 = .43$ to median $R^2 = .74$), $F(1, 48) = 8.85$, $p < .01$, $d = 0.85$. Thus, rather than Trial Block 0 reflecting an arbitrary point along a continuous path of improvement, it seemed to mark the point at which children switched from a logarithmic representation to a linear one.

Transfer of learning to fractional magnitude estimation

In the previous section, we observed that feedback on a small set of numerosities around 150 induced a large and broad change in both the accuracy and linearity of younger children's numerical estimates. We reasoned that if this improved accuracy resulted from a representational change, learning should transfer to a superficially different context—the fraction line task—regardless of its costs in accuracy.

To examine transfer to the fractional magnitude context, we first calculated the mean absolute error for each child on the fractional magnitude estimation task (|actual–estimate| / range of scale) and regressed condition (1 = treatment, 0 = control) and age (7.15–9.64) against the mean absolute error scores. As expected, there was a significant interaction between age and condition, $F(3, 60) = 12.58$, $p < .001$ (Table 2). For younger children (who had mostly generated logarithmic estimates in the numerical context), treatment had a large and negative effect on accuracy of fractional magnitude estimates, with the mean absolute error on the fraction line task being larger for the treatment group ($M = 61\%$, $SD = 11\%$) than for the control group ($M = 41\%$, $SD = 16\%$), $F(1, 29) = 15.75$, $p < .001$, $d = 1.46$. In contrast, for older children (who had mostly generated linear estimates in the numerical context), treatment did not affect accuracy of fractional magnitude estimates (treatment: $M = 62\%$, $SD = 12\%$; control: $M = 61\%$, $SD = 11\%$), $F(1, 33) = 0.06$, *ns.* Thus, correcting younger children's numerical estimates imposed a large cost on the accuracy of their fractional magnitude estimates.

To check whether this differing pattern of performance might be related to how younger children interpreted the meaning of the denominator, we next regressed their median estimate for each fractional value against the denominator of the fraction (Fig. 10). Specifi-

Table 2
Analysis of age differences at posttest

	Number line accuracy	Fraction line accuracy
Model	$F(3, 60) = 12.80$, $p = .000001$	$F(3, 60) = 12.58$, $p < .0001$
Age	$t = 4.98$, $p < .0001$ Older ($M = 91\%$) > Younger ($M = 80\%$)	$t = 5.36$, $p < .0001$ Younger ($M = 48\%$) > Older ($M = 38\%$)
Treatment	$t = 2.16$, $p = .04$ Treatment ($M = 89\%$) > Control ($M = 83\%$)	$t = 3.82$, $p = .0003$ Control ($M = 47\%$) > Treatment ($M = 38\%$)
Age × Treatment	$t = 1.96$, $p = .05$ Younger: Treatment ($M = 85\%$) > Control ($M = 76\%$), $p < .007$ Older: Treatment ($M = 93\%$), Control ($M = 89\%$), *ns*	$t = 3.61$, $p = .0006$ Younger: Control ($M = 59\%$) > Treatment ($M = 39\%$), $p < .0005$ Older: Control ($M = 38\%$), Treatment ($M = 37\%$), *ns*

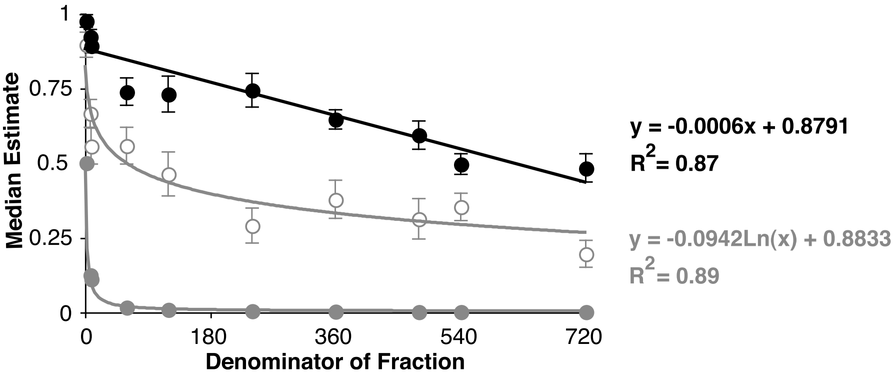

Fig. 10. Experiment 2: Short-term changes in fractional magnitude estimation. The estimates of children in the control group (light circles) were better fit by a logarithmic function than by a linear function, whereas the estimates of children in the treatment group (dark circles) were better fit by a linear function than by a logarithmic function. The ideal pattern of performance ($y = 1/x$) is depicted in gray.

cally, we wanted to test our assumption that children's estimates across the two tasks were quite similar in spite of the superficial differences between the tasks and the costs this entailed for accuracy. As predicted, children in the control group, who were typically better fit by the logarithmic function at number line posttest, also provided a series of fractional estimates that were better fit by the logarithmic function ($R^2 = .89$) than by the linear function ($R^2 = .67$). In contrast, children in the treatment group, whose number line posttest estimates typically were better fit by the linear function, provided a series of fractional magnitude estimates that were better fit by the linear function ($R^2 = .87$) than were the estimates made by participants in the control group. Thus, children did appear to transfer their understanding of numerical value to the unfamiliar fractional context regardless of its cost in accuracy.

Finally, to determine whether linearity of performance on the number line posttest was correlated with performance on the fraction line posttest, we next regressed individual participants' linear R^2 values for each task. As expected, there was a high correlation between the two variables ($r = .62$), $F(1, 29) = 17.78$, $p < .001$. When we analyzed the two groups' R^2 values on these tasks separately, we found that linear performance on the number line estimation task predicted linear performance on the fractional units task for the control group ($r = .67$), $F(1, 13) = 9.91$, $p < .01$, as well as for the treatment group ($r = .60$), $F(1, 15) = 7.98$, $p < .01$. Consistent with this finding, the degree of inaccuracy in the fractional magnitude context was strongly predicted by the degree of accuracy in the numerical magnitude context ($r = -.77$), $F(1, 29) = 42.44$, $p < .001$ (Fig. 11).

Effect of learning on sex differences in estimation accuracy

Having eliminated age group differences through training, we then examined the effect of training on sex differences. Specifically, we wanted to know whether boys continued to generate more accurate numerical magnitude estimates than girls and whether girls continued to generate more accurate fractional magnitude estimates than boys. To find out, we examined three measures of estimation performance for each task: accuracy (0–100%), fit

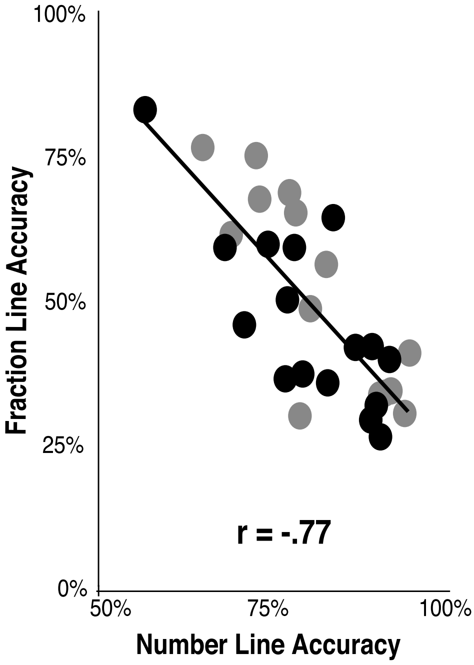

Fig. 11. Experiment 2: Relation between accuracy on number line task and fraction line task on posttest. Dark circles depict the treatment group, and light circles depict the control group.

of the function ($R^2 = .00–1.00$) associated with greater accuracy (i.e., linear for numerical estimation or logarithmic for fractional magnitude estimation), and percentage of children best fit by the function providing better accuracy (Table 3).

On each of the six dependent variables, boys and girls continued to differ in their estimation performance, and we observed no interaction between sex (0 = boys, 1 = girls) and condition (0 = control, 1 = treatment) on posttest scores, *F*s < 1.25, *ns.* As at pretest, boys' accuracy on the number line task ($M = 90\%$, $SD = 8\%$) continued to be greater than girls' accuracy ($M = 83\%$, $SD = 10\%$), $F(1, 62) = 9.34$, $p = .003$, $d = 0.77$, whereas girls' accuracy on the fraction line task ($M = 46\%$, $SD = 16\%$) continued to be greater than

Table 3
Analysis of sex differences on posttest

Number line			Fraction line		
Accuracy	Fit of linear model	% Best fit by lin	Accuracy	Fit of log model	% Best fit by log
Sex					
Boys: $M = 90\%$	Boys: lin $R^2 = .99$	Boys: 81% lin best	Boys: $M = 38\%$	Boys: log $R^2 = .78$	Boys: 33% log best
Girls: $M = 83\%$	Girls: lin $R^2 = .97$	Girls: 59% lin best	Girls: $M = 46\%$	Girls: log $R^2 = .91$	Girls: 59% log best
$F(1, 62) = 9.34$, $p = .003$		$\chi(1) = 3.04$, $p = .08$	$F(1, 62) = 5.69$, $p = .004$		$\chi(1) = 4.27$, $p = .03$

boys' accuracy ($M = 38\%$, $SD = 12\%$), $F(1, 62) = 5.69$, $p = .003$, $d = 0.57$. On the number line task (which was favored by a linear representation), the fit of the linear model to boys' median estimates ($R^2 = .99$) was slightly greater than the fit of the linear model to girls' median estimates ($R^2 = .97$); on the fraction line task (which was favored by a logarithmic representation), the fit of the logarithmic model to girls' median estimates ($R^2 = .91$) was slightly greater than the fit of the logarithmic model to boys' median estimates ($R^2 = .78$). Finally, on the number line task, a higher proportion of boys (81%) generated estimates best fit by the linear model than did girls (59%), whereas on the fraction line task, a higher proportion of girls (59%) generated estimates best fit the logarithmic function than did boys (33%). Thus, sex differences in estimation performance appeared to be stable over short-term training, whereas age differences in estimation performance were not stable over short-term training.

Discussion

Representational changes were hypothesized to lead to automatic transfer of learning, leading to gains in accuracy in one context coming at the expense of accuracy in another context. To test this hypothesis, we examined the accuracy of symbolic magnitude estimation across two contexts, whole numbers and fractions, and how these two contexts affected the relations among age, sex, and accuracy. In the next two sections, we highlight the specific findings of our tests, and in the subsequent section, we remark on the broader implications of findings for future research on age and sex differences in mathematical thinking as well as implications of findings for prevention of "cognitive illusions."

Effect of context on age differences in accuracy of magnitude estimation

Across a broad range of contexts, including estimation of distance (Cohen, Weatherford, Lomenick, & Koeller, 1979), amount of money (Sowder & Wheeler, 1989), number of discrete objects (Hecox & Hagen, 1971), answers to arithmetic problems (LeFevre, Greenham, & Naheed, 1993), and locations of numbers on number lines (Siegler & Opfer, 2003), children's estimates are highly inaccurate (for a review, see Siegler & Booth, 2005). According to the representational change hypothesis, children's inaccurate estimation across these many contexts stems from initial reliance on logarithmic representations of numerical value (Siegler & Opfer, 2003), a representation of numerical value that is consistent with Fechner's law and that is widespread among species, human infants, and time-pressured adults (for a review, see Dehaene et al., 1998). In contrast, older children and adults are thought to have learned a linear representation of numerical value from encountering experiences in school and daily life that provide log discrepant information, which typically induces rapid switching from logarithmic to linear estimation patterns in experimental studies (Opfer & Siegler, 2007; Opfer & Thompson, 2008). Also in these studies, learned linear representations appear to generalize robustly to novel contexts (Laski & Siegler, 2007; Opfer & Thompson, 2008; Ramani & Siegler, in press).

Although evidence of young children's initially poor estimation skills and logarithmic representation of numerical value and evidence of older children's good estimation skills and linear representations of numerical value have been drawn from a wide range of contexts, these contexts share an important property in that accuracy could be attained either from representational changes (Joram, Subrahmanyam, & Gelman, 1998; Siegler & Opfer,

2003) or from improving mathematical skills (Dowker et al., 1996; Hiebert & Wearne, 1986). To address this issue, we sought to provide a particularly strong test for the representational change hypothesis by examining short- and long-term changes on an estimation task—fractional magnitude estimation—that favors the logarithmic representation at the expense of the linear one. Our reasoning was that if children learn to make better estimates by automatizing use of a linear representation, estimation of fractional magnitudes should suffer with age and experience.

The results of the current study provided both correlational and experimental evidence supporting the hypothesis that numerical representations have a powerful effect on estimates of fractional magnitude. Correlational evidence was provided in Experiment 1, where older children provided more accurate estimates of numerical magnitude than did younger children, but younger children provided more accurate estimates of fractional magnitude than did older children. Overall, inaccuracy of numerical estimates accounted for 64% of the variance in accuracy of fractional magnitude estimates even when controlling for age. Experimental evidence from Experiment 2 indicated that this correlation reflected a causal link between numerical and fractional magnitude estimation. Specifically, children who were given training in numerical estimation subsequently provided more accurate estimates of numerical magnitude than did a control group whose members had received no training, but children in the treatment group also provided less accurate estimates of fractional magnitude than did children in the control group.

We believe that these findings of context effects provide particularly strong evidence for the hypothesis that numerical representations are automatic and, thereby, can impose costs as well as benefits for accuracy of estimation. This idea is an important one because it simultaneously explains (a) high correlations among individual estimation tasks, (b) the breadth of transfer that typically is observed in training studies of numerical estimation, and (c) why adults are subject to certain cognitive illusions involving fractions, including incorrect comparisons of salaries (Opfer & DeVries, in press) and incorrect evaluation of medical risks (Burkell, 2004).

Effect of context on sex differences in accuracy of magnitude estimation

Although the number line task has been used extensively in prior research on numerical representations (e.g., Booth & Siegler, 2006; Opfer & DeVries, 2008; Opfer & Siegler, 2007; Opfer & Thompson, 2008; Siegler & Booth, 2004; Siegler & Opfer, 2003), sex differences on the task have not been reported previously, leading some researchers (e.g., Spelke, 2005) to infer quite reasonably that sex differences on our number line task do not exist. At least in our own research, our failure to report sex differences previously (e.g., Opfer & Siegler, 2007; Opfer & Thompson, 2008; Siegler & Opfer, 2003) has been one of pure neglect; we simply assumed that boys' and girls' numerical representations did not differ, and we never tested that assumption. If we had done so, our findings would have been fairly similar to our findings of sex differences in Experiments 1 and 2. For example, Opfer and Thompson (2008) examined numerical estimation in first and second graders (mean age = 7.85 years) and then used a training procedure identical to that in Experiment 2, thereby allowing for comparison of sex differences prior to and following training. In our reanalysis of the data, we found that, prior to training, the 27 boys tended to generate more linear estimates on average than did the 29 girls (boys: mean lin $R^2 = .69$; girls: mean lin $R^2 = .58$), $F(1, 55) = 3.13$, $p = .08$; after training, boys also generated more linear estimates on average than did girls (boys: mean lin $R^2 = .77$; girls: mean lin $R^2 = .58$), $F(1, 55) = 6.19$, $p = .02$. An interesting question is whether similar sex differences are also evident in the number line data collected by Siegler and colleagues (e.g., Booth & Siegler, 2006; Laski & Siegler, 2007; Siegler & Booth, 2004; Siegler & Ramani, in press).

Do sex differences on number line tasks generalize to other tests of children's numerical representations? One reasonable concern is that our magnitude estimation tasks might not be representative of more widespread sex differences in numerical representations because the number line task imposes a spatial performance demand (i.e., marking a position on a line), and other tests of spatial visualization (e.g., Levine et al., 1999) have found sex differences in this ability despite the inherently nonnumerical nature of the spatial visualization tasks. To check this idea, we reexamined sex differences on a test of numerical concepts—numerical categorization (Laski & Siegler, 2007; Opfer & Thompson, 2008)—that simply imposed a verbal performance demand. On this task, children were told that 0 is a "really small number" and that 1000 is a "really big number," and then they were asked to say whether a novel number (e.g., 150) was "really small," "small," "medium," "big," or "really big." With each category assigned an arbitrary ordinal number (i.e., 1 for "really small," 5 for "really big"), one can again regress the number given against the categorization judgment to assess the linearity of boys' and girls' verbal category judgments. Although the number of children who participated in this task in Opfer and Thompson's (2008) study was fairly small (14 boys and 8 girls) and the sex differences were not always statistically significant, absolute performance and magnitude of sex differences were nevertheless quite similar on the verbal number categorization task (boys: mean lin $R^2 = .62$; girls: mean lin $R^2 = .54$) and on the spatial number line task (boys: mean lin $R^2 = .69$; girls: mean lin $R^2 = .58$). We believe this similarity suggests that early sex differences in numerical representation do not simply reflect a performance demand imposed by the number line, although clearly more tests are needed to test the claim.

Even if early sex differences in numerical representation do not reflect a performance demand, sex differences in accuracy clearly reflect a task demand. As Simon (1996) noted, accuracy is simply a measure of the fit between the approach that participants use on a task and what the task demands for accurate performance. In the case of our number line, a logarithmic representation was less adequate for accuracy than was a linear representation; thus, we observed boys to outperform girls on the number line task. In the case of our fraction line, the reverse was true; thus, we observed girls to outperform boys on the fraction line task. Combining the two tasks led to there being no overall advantage for boys or girls (boys' accuracy: $M = 64\%$; girls' accuracy: $M = 65\%$; *ns*). Thus, the sex difference in accuracy was not absolute but depended on the numerical context.

Broader implications of context effects on age and sex differences in magnitude estimation

Our findings suggest that there are real age and sex differences in the representations used to solve magnitude estimation problems but that age and sex differences in accuracy simply reflect the match of a particular numerical context to those representations. This point is an important one; if a difference between boys and girls (or between younger and older children) is at the level of the representation, the advantage of either sex (or either age) reflects a kind of "dumb accuracy," that is, representations that are broadly generalizable across contexts (leading to positive transfer of accuracy) but also inflexible to changing task demands (leading to negative transfer of accuracy).

This conclusion is also important in that it provides guidance for making broader predictions about age and sex differences in mathematical cognition. Specifically, the analysis suggests that there is no necessary positive relation in accuracy among various mathematical tasks. Rather, positive relations in accuracy among specific mathematical tasks depend on the relative similarity of the representations used to solve the specific tasks and the ability of those representations to accurately encode relevant mathematical properties. For example, for fractions expressed in decimal format or with common denominators, a logarithmic representation would be inappropriate because it fails to encode both the linear nature of the decimal system and the linear relation between the value of a fraction and its numerator. In contrast, for fractions expressed in simpler familiar units, such as expressing salaries in terms of "$1 per hour" versus "$1 per day," the linear representation would be inappropriate due to the power relations between the implicit denominator in hours and days (60 and 1440 min, respectively) and the value of the salary. A testable prediction of the current analysis is that girls and younger children should outperform boys and older children when estimating the magnitude of salaries expressed in familiar units, whereas boys and older children should outperform girls and younger children when estimating the magnitude of salaries expressed in decimals and in fractions with common denominators.

Finally, we believe that our results have a number of important educational implications. The most important implication is that efforts at improving estimation should not stop at improving numerical estimation. Rather, efforts at improving fractional magnitude estimation are likely to help the development of representations of fractional value, and these—following the literature on development of numerical representations (Booth & Siegler, 2006; Laski & Siegler, 2007; Opfer & Thompson, 2008)—are likely to transfer to comparison of fractional values, to categorization of fractions as small or large, and to accurate similarity judgments of fractional magnitudes. Discoveries about educational interventions that lead to improvement of children's numerical estimation (e.g., Opfer & Siegler, 2007; Opfer & Thompson, 2008; Ramani & Siegler, in press; Siegler & Ramani, in press) suggest a number of interventions that could improve children's representations of fractional magnitude. We believe that these interventions are important given the inappropriateness of relying exclusively on linear representations of symbolic magnitude, and we believe that the failure to provide children with effective intuitions about fractional value is lasting and may contribute to adults' known difficulties with the fractions encountered in everyday life, including understanding of statistics (Evans, Handley, Perham, Over, & Thompson, 2000), addition of discounts (Chen & Rao, 2007), comparison of salaries (Opfer & DeVries, in press), and evaluation of medical risks (Burkell, 2004).

Acknowledgments

Thanks are due to Debbie Lurie and Jennifer Filicky, who made important contributions to the data collection, and to the children and teachers of Slate Hill and Wilson Hill Elementary Schools.

References

Ansari, D., Nicolas, G., Lucas, E., Hamon, K., & Dhital, B. (2005). Neural correlates of symbolic number processing in children and adults. *Brain Imaging, 16*, 1769–1773.

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition, 44*, 75–106.

Banks, W. P., & Hill, D. K. (1974). The apparent magnitude of number scaled by random production. *Journal of Experimental Psychology Monograph, 102*, 353–376.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*, 612–637.

Berch, D. B., Foley, E. J., Hill, R. J., & Ryan, P. M. (1999). Extracting parity and magnitude from Arabic numerals: Developmental changes in number processing and mental representation. *Journal of Experimental Child Psychology, 74*, 286–308.

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology, 41*, 189–201.

Booth, J. L., & Siegler, R. S. (in press). Numerical magnitude representations influence arithmetic learning. *Child Development*.

Brainerd, C. J. (1983). Young children's mental arithmetic errors: A working-memory analysis. *Child Development, 54*, 812–830.

Burkell, J. (2004). What are the chances? Evaluating risk and benefit information in consumer health materials. *Journal of the Medical Library Association, 92*, 200–208.

Carpenter, T. P., Corbitt, M. K., Kepner, H. S., Lindquist, M. M., & Reys, R. E. (1981). *Results from the second mathematics assessment of the National Assessment of Educational Progress*. Washington, DC: National Council of Teachers of Mathematics.

Chen, H., & Rao, A. R. (2007). When two and two is not equal to four: Errors in processing multiple percentage changes. *Journal of Consumer Research, 34*, 327–340.

Chi, M. T. H., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145–182.

Cohen, R., Weatherford, D. L., Lomenick, T., & Koeller, K. (1979). Development of spatial representations: Role of task demands and familiarity with the environment. *Child Development, 50*, 1257–1260.

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

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2005). parietal circuits for number processing. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 433–453). New York: Taylor & Francis.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science, 284*, 970–974.

Doolittle, A. E., & Cleary, T. A. (1987). Gender-based differential item performance in mathematics achievement items. *Journal of Educational Measurement, 24*, 157–166.

Dowker, A., Flood, A., Griffiths, H., Harriss, L., & Hook, L. (1996). Estimation strategies of four groups. *Mathematical Cognition, 2*, 113–135.

Evans, J. S. B. T., Handley, S. J., Perham, N., Over, D. E., & Thompson, V. A. (2000). Frequency versus probability formats in statistical word problems. *Cognition, 77*, 197–213.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*, 307–314.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44*, 43–74.

Geary, D. C. (1996). Sexual selection and sex differences in mathematical abilities. *Behavioral and Brain Sciences, 19*, 229–284.

Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development, 78*, 1343–1359.

Geary, D.C., Hoard, M.K., Nugent, L., & Byrd-Craven, L. (in press). Development of number line representations in children with mathematical learning disability. *Developmental Neuropsychology*.

Goldstein, J. M., Seidman, L. J., Horton, N. J., Makris, M., Kennedy, D. N., Caviness, V. S., Jr., Faraone, S. V., & Tsuang, M. T. (2001). Normal sexual dimorphism of the adult human brain assessed by in vivo magnetic resonance imaging. *Cerebral Cortex, 11*, 490–497.

Griffin, S., Case, R., & Siegler, R. S. (1994). Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 25–49). Cambridge, MA: MIT Press/Bradford Books.

Grimes, D. A., & Snively, G. (1999). Patients' understanding of medical risks: Implications for genetic counseling. *Obstetrics & Gynecology, 93*, 910–914.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest, 8*, 1–51.

Hecox, K. E., & Hagen, J. W. (1971). Estimates and estimate-based inferences in young children. *Journal of Experimental Child Psychology, 11*, 106–123.

Hiebert, J., & Wearne, D. (1986). Procedures over concepts: The acquisition of decimal number knowledge. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 199–223). Hillsdale, NJ: Erlbaum.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*, 139–155.

Joram, E., Subrahmanyam, K., & Gelman, R. (1998). Measurement estimation: Learning to map the route from number to quantity and back. *Review of Educational Research, 68*, 413–449.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review, 103*, 582–591.

Kavaliers, M., Ossenkopp, K. P., Galea, L. A., & Kolb, B. (1998). Sex differences in spatial learning and prefrontal and parietal cortical dendritic morphology in the meadow vole, Microtus pennsylvanicus. *Brain Research, 810*, 41–47.

Knops, A., Nuerk, H. C., Sparing, R., Foltys, H., & Wilmes, K. (2006). On the functional role of human parietal cortex in number processing: How gender mediates the impact of a "virtual lesion" induced by rTMS. *Neuropsychologia, 44*, 2270–2283.

Kourtzi, Z., & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science, 293*, 1506–1509.

Laski, E. V., & Siegler, R. S. (2007). Is 27 a big number? Correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. *Child Development, 76*, 1723–1743.

LeFevre, J. A., Greenham, S. L., & Naheed, N. (1993). The development of procedural and conceptual knowledge in computational estimation. *Cognition and Instruction, 11*, 95–132.

Levine, S. C., Huttenlocher, J., Taylor, A., & Langrock, A. (1999). Early sex differences in spatial skill. *Developmental Psychology, 35*, 940–949.

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

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

Naccache, L., & Dehaene, S. (2001). The priming method: Imaging unconscious repetition priming reveals an abstract representation of number in the parietal lobes. *Cerebral Cortex, 11*, 966–974.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics: Higher standards for our students . . . higher standards for ourselves*. Reston, VA: Author.

Newcombe, N. S., Mathason, L., & Terlecki, M. (2002). Maximization of spatial competence: More important than finding the cause of sex differences. In A. McGillicuddy-De Lisi & R. De Lisi (Eds.), *Biology, society, and behavior: The development of sex differences in cognition* (pp. 183–206). Westport, CT: Ablex.

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

Opfer, J.E., & DeVries, J. (in press). Representational change and magnitude estimation: Why young children can make more accurate salary comparisons than adults. *Cognition*.

Opfer, J. E., & Siegler, R. S. (2007). Representational change and children's numerical estimation. *Cognitive Psychology, 55*, 169–195.

Opfer, J. E., & Thompson, C. A. (2008). The trouble with transfer: Insights from microgenetic changes in the representation of numerical magnitude. *Child Development, 79*, 790–806.

Opfer, J.E., Thompson, C.A., & Furlong, E. (2007). *Early development of spatial–numeric associations: Evidence from spatial and quantitative performance of pre-reading preschoolers*. Unpublished manuscript, The Ohio State University.

Petrov, A., & Anderson, J. R. (2005). The dynamics of scaling: A memory-based anchor model of category rating and absolute identification. *Psychological Review, 112*, 383–416.

Pinel, P., Piazza, M., LeBihan, D., & Dehaene, S. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. *Neuron, 41*, 983–993.

Potter, M. C., & Faulconer, B. A. (1975). Time to understand pictures and words. *Nature, 253*, 437–438.

Potter, M. C., & Kroll, J. F. (1987). Conceptual representation of pictures and words: Reply to Clark. *Journal of Experimental Psychology: General, 116*, 310–311.

Potter, M. C., Kroll, J. F., Yachzel, B., Carpenter, E., & Sherman, J. (1986). Pictures in sentences: Understanding without words. *Journal of Experimental Psychology: General, 115*, 281–294.

Ramani, G.B., & Siegler, R.S. (in press). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Development*.

Rivera, S. M., Reiss, A. L., Eckert, M. A., & Menon, V. (2005). Developmental changes in mental arithmetic: Evidence for increased functional specialization in the left inferior parietal cortex. *Cerebral Cortex, 15*, 1779–1790.

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General, 117*, 258–275.

Siegler, R. S. (2002). Microgenetic studies of self-explanations. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 31–58). New York: Cambridge.

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development, 75*, 428–444.

Siegler, R. S., & Booth, J. L. (2005). Development of numerical estimation: A review. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 197–212). New York: Taylor & Francis.

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science, 14*, 237–243.

Siegler, R.S., & Ramani, G.B. (in press). Playing board games promotes low-income children's numerical development. *Developmental Science*.

Simon, H. A. (1996). *The sciences of the artificial*. Cambridge, MA: MIT Press.

Solomon, R. L., & Lessac, M. S. (1968). A control group design for experimental studies of developmental processes. *Psychological Bulletin, 70*, 145–150.

Sowder, J. T., & Wheeler, M. M. (1989). The development of concepts and strategies used in computational estimation. *Journal for Research in Mathematics Education, 20*, 130–146.

Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science: A critical review. *American Psychologist, 60*, 950–958.

Walker, A. P. (1997). Genetic counseling. In D. L. Rimoin, J. M. Connor, & & R. E. Pyeritz (Eds.), *Emery and Rimoin's principles and practice of medical genetics* (pp. 595–618). New York: Churchill Livingstone.

Willingham, W. W., & Cole, N. S. (1997). *Gender and fair assessment*. Mahwah, NJ: Lawrence Erlbaum.

Xu, F., & Spelke, E. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74*, B1–B11.