Child Development, May/June 2008, Volume 79, Number 3, Pages 788 – 804

The Trouble With Transfer: Insights From Microgenetic Changes in the Representation of Numerical Magnitude

John E. Opfer and Clarissa A. Thompson
The Ohio State University

Spontaneous transfer of learning is often difficult to elicit. This finding may be widespread partly because pretests proactively interfere with transfer. To test this hypothesis, 7-year-olds' transfer was examined across 2 numerical tasks (number line estimation and categorization) in which similar representational changes have been observed. As predicted, children given feedback on numerical estimates learned to use a linear representation of numerical quantity instead of a logarithmic one, but providing practice on a categorization pretest led children to continue using a logarithmic representation on the same task, which they otherwise abandoned with surprising frequency. These findings imply unsupervised practice of inappropriate representations impedes transfer, and studies of learning can greatly underestimate children's potential for transfer if pretest effects are uncontrolled.

Despite the importance of transferring knowledge—whether from domain to domain, from school to everyday life, or from everyday life to school—spontaneous transfer is notoriously difficult to elicit, with learners typically generalizing new approaches to a much narrower set of problems than is optimal (Barnett & Ceci, 2002; Bassok & Holyoak 1989a, 1989b; Gick & Holyoak, 1983; Singley & Anderson, 1989; Spiro, Feltovich, Jacobson, & Coulson, 1991; Thorndike, 1922). This difficulty is not simply because initial learning is incomplete or unstable; even in microgenetic studies of children's learning, which allow the stability of new knowledge to be established by obtaining a dense sampling of children's thinking over time by means of trial-to-trial assessments, children very often underextend novel solutions (see Siegler, 2006, for a review of 105 microgenetic studies examining the issue). For example, once fourth graders have learned to solve addition equations in the form $A + B + C = \_\_\_ + C$, they fail to transfer their new knowledge to multiplication equations in the form $A \times B \times C = \_\_\_ \times C$ (Alibali & Goldin-Meadow, 1993).

Numerous explanations have been proposed to identify why transfer is so difficult for learners to achieve. These explanations have included the degree of overlap in production rules between the base and the transfer domain (Singley & Anderson, 1989), where attention is directed during learning (Anderson, Reder, & Simon, 1996), lack of structural similarity between the base domain and the transfer domain (Gick & Holyoak, 1983; Kaminski, 2006), the stability and organization of initial learning (Opfer & Siegler, 2004), and insufficient prior knowledge in the base domain (Brown & Kane, 1988; Brown, Kane, & Echols, 1986). Indeed, the many possible impediments to transfer have led some investigators to posit the somewhat pessimistic conclusion that learning is universally narrow and ''situation specific'' (Cognition and Technology Group at Vanderbilt, 1997; Lave, 1988).

To add to this already daunting list, we propose a novel (yet somewhat more optimistic) explanation for why transfer is difficult to elicit in microgenetic studies in particular and possibly other studies more generally—proactive interference from previous practice. The basic premise of our account is that when children have practice on a task without any feedback (e.g., when they complete a pretest on a transfer task), children must use *some* representation to complete that task, and the more they use that representation, the greater the strength of the representation and the more likely children will continue using it (Erdelyi & Becker, 1974; Roediger & Payne, 1982; Siegler & Shipley, 1995). Under circumstances in which the representations used on the task are already appropriate, pretests can facilitate transfer (''practice makes perfect''; Gick & Holyoak, 1983; Roediger & Karpicke, 2006). However, when representations are *inappropriate*, practice on a pretest makes *imperfect*

The authors would like to thank the children, parents, and staff at Slate Hill, Wilson Hill, and Wickliffe Alternative schools, as well as Stephanie Heinsons, Jennifer Filicky, and Debbie Lurie for their help in data collection. Vladimir Sloutsky, Robert Siegler, Michael Edwards, and two anonymous reviewers also made many helpful suggestions on a previous draft of the manuscript. Portions of these data were previously presented at the 29th Annual Conference of the Cognitive Science Society.

Correspondence concerning this article should be addressed to John E. Opfer, Department of Psychology, The Ohio State University, 245 Psychology Building, 1835 Neil Avenue, Columbus OH 43210. Electronic mail may be sent to opfer.7@osu.edu.

because practice merely strengthens the inappropriate representation used on the transfer task and thereby blocks transfer of the more optimal representations learned during training (Gick & Holyoak, 1980; Roediger & Marsh, 2005). If true, this explanation is not a trivial one: It suggests that the trouble with transfer is at least partly an experimental artifact of studies that pretest the treatment and control groups.

To test our practice interference hypothesis, we used Solomon and Lessac's (1968) four-group design to assess the independent and interactive effects of treatment (feedback) and pretesting on children's judgments of numerical magnitude. The tasks we chose were ones where previous cross-sectional (Laski & Siegler, 2007; Siegler & Opfer, 2003) and microgenetic (Opfer & Siegler, 2007) studies had shown that young children use an inappropriate (logarithmic) representation before using an appropriate (linear) one (e.g., younger children judge 150 to be closer to 1,000 than to 1, whereas older children judge 150 to be closer to 1 than to 1,000). In the first task, children were asked to make estimates of numerical quantity (e.g., where 150 would fall on a line flanked by 0 and 1,000), and they were given feedback on their estimates so they would learn to use a linear representation (as in Opfer & Siegler, 2007). In the second task, children were asked to categorize numerals by their magnitude (e.g., whether 150 is a small or big number in the context of the 0 – 1,000 range); this is the task where we hoped children would transfer their learning. On the assumption that the two tasks tapped a common representation, we predicted that children's learning of the linear representation on the number line estimation task would transfer to their category judgments (at least when they were not given a categorization pretest). Further, based on our practice interference hypothesis, we predicted that pretesting on the categorization task would also strengthen the logarithmic representation and thus block spontaneous transfer of learning.

In the next three sections, we will briefly (a) examine the literature indicating why transfer might be difficult for children, (b) review previous findings in the domain where we studied transfer (estimation and numerical magnitude judgments), and then (c) describe a microgenetic experiment designed to test our explanation for children's trouble with transfer in this domain and more broadly.

The Trouble With Transfer: Problems Raised by Previous Studies

Ideally, learners generalize freely from the situation where learning occurred to novel but similar situations. For many years, however, research has shown that child and adult learners fall very far from this ideal. Learning to solve symbolic math problems, for example, often generalizes fairly weakly to familiar—but formally identical—real-world problems (Schoenfeld, 1985), and the reverse is equally true: Brazilian street vendors who can solve difficult arithmetic problems in their daily work fail to transfer their solutions to the symbolic problems they face in school (Carraher, Carraher, & Schliemann, 1985). Results such as these have been known since Thorndike's (1922) seminal studies on transfer, which led him to formulate his identical-elements theory of transfer: ''Any disturbance whatsoever in the concrete particulars reasoned about will interfere somewhat with the reasoning, making it less correct, or slower, or both'' (Thorndike, 1922, p. 36).

What has always made such findings interesting, however, is that the general *capacity* for transfer seemingly must exist, even if transfer is not as perfect as hoped. As Meiklejohn (1908) keenly observed:

> What can we say of a theory that the training of the mind is so specific that each particular act gives facility only for the performing again of that same act just as it was before? Think of learning to drive a nail with a yellow hammer, and then realize your helplessness if, in time of need, you should borrow your neighbor's hammer and find it painted red. Nay, further think of learning to use a hammer at all if at each other stroke the nail has gone further into the wood, and the sun has gone lower in the sky, and the temperature of the body has risen from the exercise, and in fact, everything on earth and under the earth has changed so far as to give each new stroke a new particularity all of its own, and thus has cut it off from all possibility of influence upon or influence from its fellows. (Meiklejohn, 1908, p. 126)

Indeed, consistent with Meiklejohn's (1908) conjecture, researchers have also found cases of symbolic learning that are remarkable for the robustness and breadth of the learning obtained. Given only a small hint, people transfer solutions across situations that have *no* identical elements (Gick & Holyoak, 1980). Moreover, transfer from very brief explicit lessons to real-world situations is at least sometimes as great as transfer from many years of real-world experience (Biederman & Shiffrar, 1987). In one famous study, students given an abstract lesson on light refraction immediately transferred the lesson to improve their accuracy in throwing darts at an underwater dartboard (Hendrickson & Schroeder, 1941). As these cases of transfer attest, the chief psychological problem is not whether or how widely transfer does or does not occur, but why—given the capacity for transfer—learners so often have trouble with transfer, such as those reported as being widespread in microgenetic studies of learning and transfer (Siegler, 2006).

Our proposal for why children have trouble with transfer in these situations is inspired by Duncker's (1945) classic study of problem solving. Duncker tested whether participants could apply novel functions to various familiar objects (e.g., a matchbox, tacks, and candles). For example, to solve the problem of placing three small candles on a door at eye level, participants had to conceive of a novel function for the matchbox (i.e., to serve as a platform). Normally, about 86% of such problems were solved; however, if subjects were given a pretest to determine their knowledge of the original functions of the familiar objects, the rate of problem solving dropped to 58%. Duncker's explanation was that previous experience with the objects induced a ''functional fixedness'' that inhibited their novel solutions.

We wondered whether prior experience might broadly induce a kind of ''conceptual fixedness'' that might also prevent children from transferring their knowledge. For example, one reason that children may fail to transfer from school lessons to familiar real-world problems (yet succeed in transferring to novel problems like throwing darts at an underwater dartboard) is that their previous successes on real-world problems interferes with the application of novel school lessons, much like Duncker's (1945) pretest interfered with his subjects' ability to think of novel solutions. That is, previous experiences in school and real-world settings interfere with transfer because they lead children to think that symbolic operations are ''for school'' and not ''for real-world problems,'' much like Duncker's subjects thought of matchboxes as ''for holding matches'' and not ''for supporting candles.''

Although our application of Duncker's (1945) findings to transfer is novel, it is also consistent with previous findings on the formation of associations between strategies and problems (Shrager & Siegler, 1998), the formation of undesirable memory traces through practice (Roediger & Payne, 1982), and with the induction of ''mental sets'' in analogy (Gick & Holyoak, 1980) and problem solving (Luchins, 1942; Luchins & Luchins, 1950). Moreover, pretests can sometimes harm animals' learning as well (Lessac & Solomon, 1969). In Solomon and Lessac's studies (Lessac, 1965; Lessac & Solomon, 1969; Solomon & Lessac, 1968), for example, beagle pups were reared in isolation or a kennel, and half were pretested across a range of tasks (conditioning, perceptual, and motor) to assess the effects of pretesting. When pups were later tested after the isolation/kennel rearing, results indicated that pretesting protected the isolated pups from many—but not all—deleterious effects of social isolation. In one spatial-reasoning task, for example, performance of *pretested* isolated pups decreased 10-fold from pretest to posttest, which was much worse than posttest performance of the isolated but *unpretested* pups.

The potential generality of findings of pretest effects is especially important for microgenetic studies. If true, the hypothesis that pretesting inhibits transfer performance could explain why children's capacity for transfer is so seldom evident in microgenetic studies that repeatedly test children over many trials, sessions, or days. Moreover, this possibility is an important one because it suggests a real-world situation where transfer might be difficult (i.e., for problems where inappropriate approaches are practiced repeatedly), a situation where transfer might be easy (i.e., for problems where children do not practice inappropriate approaches repeatedly), and a very general methodological control for assessing the impact of unsupervised practice (i.e., manipulating the pretesting variable). In the next two sections, we introduce the domain where we studied the potential effect of practice on transfer and how we studied its potential effects.

Development of Estimation Skills and the Representation of Numerical Magnitude

We chose to examine transfer in the domain of estimation for two reasons. First, children's estimation skills play a central role in a wide range of mathematical activities (Dowker, 2003; Siegler & Booth, 2004), leading educators to place a high priority on improving them for at least the past 25 years (e.g., National Council of Teachers of Mathematics, 1980, 1989, 2001). Despite this prolonged effort, however, estimation remains a process that children find difficult. Whether estimating 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), or locations of numbers on number lines (Siegler & Opfer, 2003), 5- to 10-year-olds' estimates are highly inaccurate. Thus, in the domain of estimation, there was significant learning to be had and many tasks to which children might transfer their learning.

Second, recent findings suggest that an important source of children's difficulty with numerical estimation is their use of an inappropriate *representation* of numerical value, thereby making estimation an attractive candidate for studying the influence of practicing inappropriate representations. Specifically, children often use an immature logarithmic representation in situations where accurate estimation requires use of a linear representation (Booth & Siegler, 2006; Opfer & Siegler, 2007; Siegler & Booth, 2004; Siegler & Opfer, 2003). For example, when asked to estimate the locations of numbers on number lines with 0 at one end and 1,000 at the other, the large majority of second graders tested by Siegler and Opfer (2003) generated logarithmic distributions of estimates (e.g., estimating 150 to be closer to 1,000 than to 1). Interestingly, this use of a logarithmic representation is widespread among species, human infants, and time-pressured adults (Dehaene, Dehaene-Lambertz, & Cohen, 1998; Moyer & Landauer, 1967; Nieder & Miller, 2003), and presumably for a good reason: In a great many situations, such representations are useful. For example, to a hungry animal (or baby), the difference between 2 and 3 pieces of food is far more important than the difference between 87 and 88 pieces. In the formal numerical system, however, magnitudes increase linearly rather than logarithmically. Thus, children's initial use of logarithmic representations of numerical magnitudes is understandable, but in school and modern life, it can interfere with accurate estimation.

Although representations of number may start logarithmic, they change with age and experience. Developmental shifts from a logarithmic to a linear representation have been found between kindergarten and second grade for estimates of numerical locations on 0 – 100 number lines (Siegler & Booth, 2004) and between second and sixth grade for estimates of numerical locations on 0 – 1,000 lines (Siegler & Opfer, 2003). Direct evidence for this shift was recently found in a microgenetic study of number line estimation (Opfer & Siegler, 2007). Specifically, second graders who were initially assessed as relying on logarithmic representations were given feedback that varied in the degree of discrepancy between the predictions of the linear and logarithmic representations. The most discrepant feedback (occurring for the number 150 on a 0 – 1,000 number line) produced the greatest representational change. This change was strikingly abrupt, often occurring after a single feedback trial, and impressively broad, affecting estimates over the entire range of numbers from 0 to 1,000. Thus, we expected to be able to induce a representational change fairly quickly, which would allow us to directly examine which aspects of the change process were affected by pretesting performance on transfer problems.

Finally, findings regarding individual and task differences are also consistent with the idea that children's difficulties in numerical estimation stem in large part from their use of logarithmic representations. First, children who are generally skillful at one type of numerical estimation tend to be skillful at others, and from second to fourth grade, the linearity of children's estimates improve to a similar degree across many different estimation tasks for which the linear representation is more appropriate (Booth & Siegler, 2006; Laski & Siegler, 2007). Second, evidence that these age-related improvements are caused by representational changes (rather than by generally improving mathematical skills) is provided by a very telling exception to the general developmental trend, a case in which the logarithmic representation is actually more beneficial than the linear representation. Specifically, when estimating fractional magnitudes (e.g., placing 1/150 on a line from 1/1 to 1/1,000), children's logarithmic representation leads them to generate surprisingly accurate estimates, better in fact than those of more mathematically savvy adults, whose knowledge that 150 is closer 1 than to 1,000 causes the illusion that 1/150 is closer to 1/1 than to 1/1,000 (Opfer & DeVries, in press; Thompson & Opfer, in press). Thus, we expected there to be a high potential for transfer of newly acquired representations across contexts.

In summary, children's estimates and numerical category judgments appear to tap a common representation that changes in a coherent fashion with age and experience. This interpretation led us to expect a high potential for children transferring representations learned in an estimation context to performance in a categorization context. However, based on our practice interference hypothesis, we were also concerned that the pretest commonly used in studies of children's learning might impede this transfer by strengthening the logarithmic representation.

The Present Study

To test our practice interference hypothesis, we conducted a microgenetic study in which we examined the process of change in number line estimation, the transfer of representations from an estimation context to a categorization context, and the impact of pretesting on the categorization task for later transfer. To examine the change process in number line estimation, we employed the basic design previously used by Opfer and Siegler (2007). To examine the transfer of representations from estimation to categorization, we additionally examined performance on a number categorization task that we adapted from

Laski and Siegler (2007). Finally, to test the hypothesis that using representations interferes with the ability to change them, we structured the microgenetic study using the Solomon and Lessac (1968) four-group design rather than the traditional two-group design used in the 105 microgenetic studies reviewed by Siegler (2006). Specifically, the Solomon and Lessac design allowed us to examine the effect of pretesting the transfer (categorization) task on subsequent performance on that task.

Although the purpose and design of microgenetic studies have been widely discussed, the purpose and features of the Solomon and Lessac (1968) design require some explanation. The purpose of the design is to assess change from pretest to posttest while controlling for the effect of the pretest itself, which is often important in studies of isolation (e.g., Lessac & Solomon, 1969), sensory deprivation (Conlee & Parks, 1981; Rubel, 1984), and attachment (van den Boom, 1994, 1995). To accomplish this goal, the design calls for an administration of posttest to four groups that vary orthogonally in the administration of pretest and treatment: Group I receives both pretest and treatment, Group II receives no pretest but does receive treatment, Group III receives a pretest but no treatment, and Group IV receives neither pretest nor treatment. This design provides two major benefits. First, strategic comparisons among the four groups allow one to anchor interpretations of developmental change to absolute as well as to relative data. Second, it allows for the test of pretest effects and the interaction between pretreatment testing and treatment itself.

Although there are many possible interactions between pretest and treatment, we illustrate two of the most interesting ones for our study in Figure 1: the case where pretesting the treatment group on the transfer task *inhibits* the effect of treatment (i.e., transfer of learning from the training task to the transfer task), and the case where pretesting the treatment group on the transfer task *enhances* the effect of treatment. The first case directly corresponds to the predictions of our pretest interference hypothesis, whereas the second case corresponds to findings that would falsify our hypothesis.

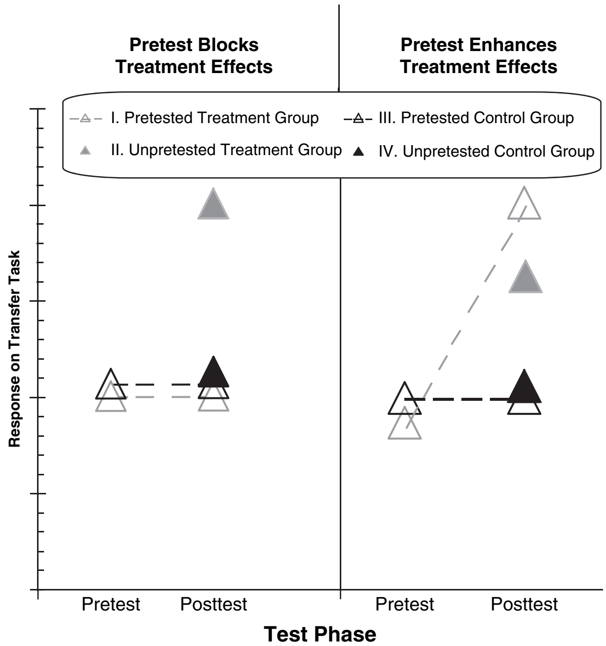

Figure 1. Illustration of two interactions detectable by the Solomon–Lessac four-group design.
Note. In the first case (Panel a), pretesting inhibits treatment effects. In second case (Panel b), pretesting enhances treatment effects.

Method

Participants

Participants were 56 first and second graders ($M = 7.85$, $SD = 0.65$; 29 girls, 27 boys). The children attended neighborhood schools in largely European American, middle-class suburbs surrounding a large metropolitan city in Midwestern United States. One of two female research assistants served as experimenter.

Tasks

Number line task. All number line problems during number line pretest, training, and number line posttest consisted of a 20-cm line with the left endpoint labeled 0, the right endpoint labeled 1,000, and with the number to be estimated appearing 2 cm above the midpoint of the number line. Over the course of the experiment, participants were asked to place the following numbers on a number line by making a hatch mark: 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. These numbers were chosen to maximize the discriminability of logarithmic and linear functions by oversampling the low end of the range, to minimize the influence of specific knowledge (such as that 500 is halfway between 0 and 1,000), and to test predictions about the range of numbers where estimates would most differ between number line pretest and posttest.

Categorization task. In the categorization task, children were asked to say how large the numbers were when compared to 0 (*really small*) and 1,000 (*really big*). To do this, children were given the following instructions (adapted from Laski & Siegler, 2007): "I'm going to ask you what you think about some of the numbers between 0 and 1,000. Some of these numbers are *really small*, some are *small*, some are *medium*, some are *big*, and some are *really big*. I'm going to say a number, and you need to tell me if you think the number is a 'really small' number, a 'small' number, a 'medium' number, a 'big' number, or a 'really big' number." The experimenter then told the children that they could refer to boxes to help them remember all their choices, and the experimenter set out five labeled, identically sized boxes, one at a time and from left to right, and read each label in turn, "really small," "small," "medium," "big," and "really big."

To orient the participants to the endpoints and to ensure that they understood the task, children were first asked to categorize 0 and 1,000. On these two practice trials (and no others), the experimenter provided feedback if the participant did not categorize 0 as "really small" and 1,000 as "really big." The other 10 numbers children were asked about comprised a subset of the numbers (2, 5, 78, 100, 150, 246, 486, 606, 725, and 938) used in the number line task, and these numbers were randomized for each participant.

Design and Procedure

Children were randomly assigned to one of four groups (see Table 1): a pretested treatment group (Group I, $n = 13$), an unpretested treatment group (Group II, $n = 11$), a pretested control group (Group III, $n = 18$), or an unpretested control group (Group IV, $n = 14$). The four groups differed in whether they were pretested on the transfer task (number categorization) and in whether they received treatment (i.e., feedback) in the training task (number line estimation). Participants in Group I received the categorization task (where we hoped for transfer) immediately prior to and following their completion of number line problems, and they were also given feedback on number line problems. Participants in Group II received the categorization task only following their completion of number line problems, and they were also given feedback on these number line problems. Participants in Group III received the categorization task immediately prior to and following their completion of number line problems, but they were not given feedback on number line problems. Participants in Group IV received the categorization task only after their completion of number line problems, but they too were not given feedback on number line problems during training.

Table 1
Experimental Design: Items (e.g., 10 Numbers Between 0 and 1,000) Presented in Each Test Phase by Experimental Group

	Pretest		Training phase						Posttest	
Experimental groups	Categorization	Number line	Feedback	Test	Feedback	Test	Feedback	Test	Number line	Categorization
I. Pretested treatment group	10 (0–1,000)	22 (0–1,000)	150	10 (0–1,000)	3 (147–187)	10 (0–1,000)	3 (147–187)	10 (0–1,000)	22 (0–1,000)	10 (0–1,000)
II. Unpretested treatment group	None	22 (0–1,000)	150	10 (0–1,000)	3 (147–187)	10 (0–1,000)	3 (147–187)	10 (0–1,000)	22 (0–1,000)	10 (0–1,000)
III. Pretested control group	10 (0–1,000)	22 (0–1,000)	None	10 (0–1,000)	None	10 (0–1,000)	None	10 (0–1,000)	22 (0–1,000)	10 (0–1,000)
IV. Unpretested control group	None	22 (0–1,000)	None	10 (0–1,000)	None	10 (0–1,000)	None	10 (0–1,000)	22 (0–1,000)	10 (0–1,000)

As shown in the outline of the procedure in Table 1, children in all four 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 ensure that learning had occurred prior to the transfer task (i.e., on the number line posttest) and to examine the course of learning prior to posttest (i.e., to examine changes from the number line pretest through posttest). On the number line pretest and posttest, children in all four groups were presented the same 22 problems without feedback (i.e., without treatment). For children in the two treatment groups, 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 included either one item on which children received feedback (Trial Block 1) or three items on which they received feedback (Trial Blocks 2 and 3). The test phase in all three training trial blocks included 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 two control groups received the same number of estimation problems, but they never received treatment.

Feedback was administered to the two treatment groups (Group I and Group II) following the same procedure used in Opfer and Siegler (2007). 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 the 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 1,000. 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 a 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, she said, ''You told me that N would go here. Actually, this is where N goes (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 said, ''That's quite a bit too high/too low. You can see these two lines [the child's and experimenter's hatch marks] are really quite far from each other.'' Finally, for both high and low deviation estimates, the experimenter asked children to explain the feedback given (a feedback strategy that often has large effects on changes in accuracy; see Chi et al., 1989; Siegler, 2002).

Results

We organized our results into two sections: (a) results concerning the process of change in numerical estimation and (b) results concerning transfer of learning to numerical categorization. Within the first section, we report on the conditions that led to changes in numerical estimation (''source of change''), how quickly those changes occurred (''rate of change''), approaches that children used up to and following the use of mature approaches (''path of change''), and individual differences in learning (''variability of change''). Within the second section, we report results testing our hypotheses about transfer (''breadth of change''). These dimensions of change have proven useful in prior microgenetic studies and in characterizing cognitive change more broadly (Siegler, 1996, 2006).

Process of Change in Numerical 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 and influenced the degree to which their estimates came to follow a linear function. To find out, we compared number line pretest and posttest estimates of the treatment groups to the control groups. (Throughout this section, we collapsed the two treatment groups and two control groups because there was never an effect of testing numerical categorization on children's subsequent numerical estimates, $F\text{s} < 1.25$.)

To examine the effect of treatment on estimation accuracy, we first calculated the mean absolute error for each child ($|\text{actual} - \text{estimate}|/\text{range of scale}$) and then performed a 2 (condition: treatment vs. control) × 2 (test phase: number line pretest vs. number line posttest) repeated measures analysis of variance (ANOVA) on the error scores. As expected, test phase interacted with condition, $F(1, 54) = 4.31$, $p < .05$. For the children in the treatment groups, the mean absolute error declined from 22% (SD = 0.14) to 17% (SD = 0.1), $t(23) = 2.2$, $p < .05$, $d = .41$, whereas for the children who were in the control groups, the mean absolute error did not differ by test phase (number line pretest, M = 19%, SD = 0.06; number line posttest, M = 20%, SD = 0.08).

We next examined whether these changes in estimation accuracy were also accompanied by the hypothesized logarithmic-to-linear shift. As shown in Figure 2, on the number line pretest, children's mean estimates for each number were in fact fit better by the logarithmic regression function than by the linear one, regardless of experimental condition. The precision of the fit of the logarithmic function, and the degree of superiority of that function to the linear function, was similar across the treatment (log R^2 = .95, lin R^2 = .80) and control groups (log R^2 = .93, lin R^2 = .82). In contrast, the groups differed considerably in their posttest estimation patterns (see Figure 2). Children in the control groups continued to generate estimates that fit the logarithmic function better than the linear one (log R^2 = .92, lin R^2 = .85), with their median estimates being almost identical from pretest to posttest (hence the large overlap between the pretest and posttest series in Figure 2). In contrast, children in the treatment groups generated posttest estimates that fit the linear function substantially better than the logarithmic one (lin R^2 = .96, log R^2 = .69).

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 (66%) provided estimates that were better fit by the logarithmic function than the linear one, regardless of whether they later received treatment, 63% of children generated logarithmic estimates in the treatment groups and 69% of children generated logarithmic estimates in the control groups, $\chi^2(1) = 0.24$, *ns*. Furthermore, posttest estimates also indicated that receiving the treatment led to more children providing linear estimates: Eighty-three percent of children who were in the treatment groups provided more linear than logarithmic estimates, whereas only 34% of children who were in the control groups provided more linear than logarithmic series of estimates, $\chi^2(1) = 13.30$, $p < .001$.

Thus, as in Opfer and Siegler (2007), feedback on a very small (but strategic) set of estimation problems led to large changes in estimation accuracy and in an overall pattern of estimates that is more characteristic of adults than children. In the next section, we examined just how small this set of feedback problems could be to produce the large changes observed in the experimental session.

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 feedback (i.e., pretest) and thus 0 trials of feedback. We also included experimental group (treatment vs. control, as above) and the interaction of experimental group and trial

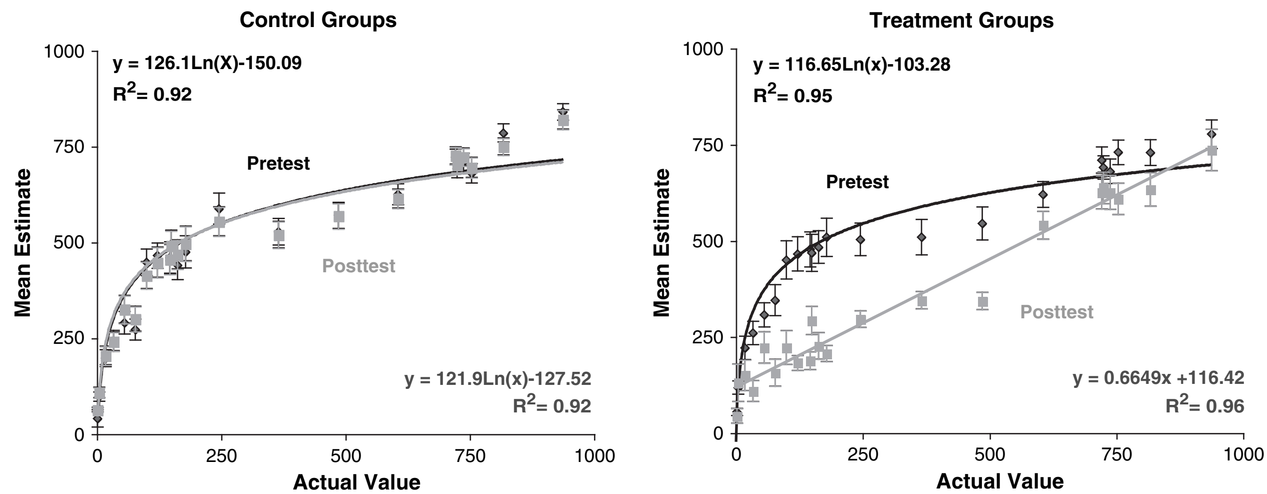

Figure 2. Source of change: Best fitting functions for mean estimates on number line (training) task at pretest (indicated in black) and posttest (indicated in gray) for treatment and control groups.

Note. Solid function lines indicate that the function fit the data significantly better than the alternative model did. Error bars indicate standard errors.

block as predictor variables in the logistic regression model. Our key prediction was that training group and trial block would interact, with the interaction occurring due to children in the treatment group generating linear estimates after fewer trial blocks than those in the control group.

As predicted, there was a significant interaction between number of trial blocks (0–4) and experimental group (treatment or control) on likelihood of generating linear patterns of estimates, when controlling for the other predictors in the model, $\hat{\beta} = .48$, $z = 2.48$, Wald $(3, N = 280) = 38.11$, $p < .05$. This interaction between experimental group and trial block (Figure 3) reflected different rates of learning in the treatment and control groups. For the treatment group, there was a significant positive effect of trial block on the likelihood of generating linear estimates, $\hat{\beta} = .52$, $z = 3.34$, Wald $(1, N = 120) = 12.49$, $p < .001$, indicating that with each trial block children were 1.62 times as likely to generate linear estimates. For the control group, however, there was no effect of trial block on the likelihood of generating linear estimates, $\hat{\beta} = .04$, $z = 0.35$, Wald $(1, N = 160) = 0.12$, *ns*.

We next looked at how quickly children responded to treatment by using logistic regression to examine the effect of treatment on generating linear estimates in each trial block (see Table 2). On Trial Block 0, before feedback was given to the treatment group, the treatment and control groups did not differ in how often they generated linear estimates, $\hat{\beta} = .10$, $z = .17$, Wald $(1, N = 56) = 0.03$, *ns*. On Trial Block 1, after children in the treatment group were given feedback

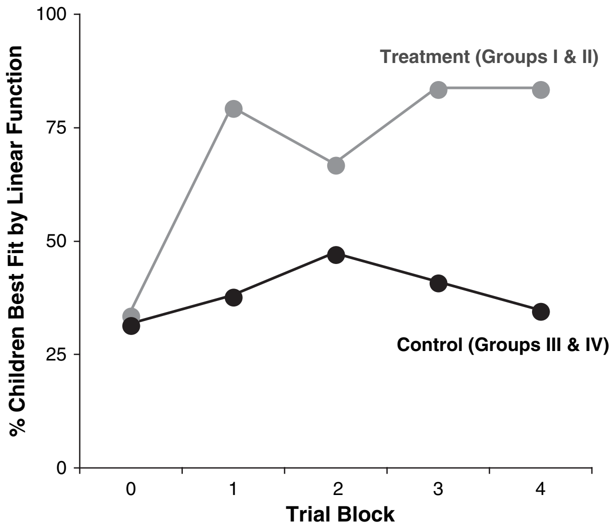

Figure 3. Rate of change: Trial block-to-trial block changes in percentage of children whose estimates were best fit by the linear function on number line (training) task.

Table 2
Rate of Change in Training Phase

	Trial blocks of treatment									
	0		1		2		3		4	
	Frequency of linear best	Odds of linear best	Frequency of linear best	Odds of linear best	Frequency of linear best	Odds of linear best	Frequency of linear best	Odds of linear best	Frequency of linear best	Odds of linear best
Treatment groups	0.33	0.50	0.79	3.80	0.67	2.00	0.83	5.00	0.83	5.00
Control groups	0.31	0.45	0.38	0.60	0.46	0.88	0.40	0.68	0.34	0.52

only on the magnitude of 150, the effect of treatment was immediately evident, $\hat{\beta} = 1.85$, $z = 2.97$, Wald (1, $N = 56$) = 10.08, $p < .005$, with children in the treatment group being 6.33 times as likely to generate linear estimates as children in the control group. Further, on Trial Blocks 2–4, children in the treatment groups continued to generate linear patterns of estimates more frequently than children in the control groups, $\hat{\beta} = 1.55$, $z = 4.48$, Wald (1, $N = 168$) = 22.12, $p < .0001$, with children in the treatment group being 4.7 times as likely to generate linear estimates on each trial block as children in the control group. What this meant was that treatment effects manifested and persisted after *feedback on a single estimate*.

Two types of evidence were consistent with this portrait of fast and persistent learning. The first type of evidence came from the trial of discovery, that is, the number of trials before children used the linear representation for the first time. For this analysis, we necessarily excluded children ($n = 13$) whose estimates were never better fit by the linear function (indicating no discovery) and children whose estimates were already better fit by the linear function on pretest ($n = 18$) (indicating that the discovery had already occurred). Of the remaining 25 children (10 from control group and 15 from treatment group), the fastest learners, children whose estimates were better fit by the linear model on Trial Block 1, were assigned a score of 1; the slowest learners, children whose estimates were better fit by the linear model for the first time on the posttest, were assigned a score of 4. An ANOVA indicated a trend toward a difference in the rate of discovery, $F(1, 24) = 3.48$, $p < .10$, $d = .70$, with the first trial block on which the linear function provided a better fit occurring slightly earlier in the treatment groups than in the control groups ($M = 1.47$ trial blocks, $SD = 0.70$ vs. $M = 2.20$ trial blocks, $SD = 1.29$). Thus, even when we excluded children who *never* used the linear representation (principally from the control group), treatment still had the effect of leading children to use the representation earlier (typically after feedback on a single numerosity) than they would have had they not received treatment.

The second type of evidence for faster learning in the treatment groups came from children's trial block of last error, which was the trial block on which children's estimates were last best fit by the logarithmic function. By this measure, too, treatment led to earlier adoption of the linear representation. For children in the control groups, the average trial block of last error ($M = 3.6$, $SD = .15$ of 4 trial blocks) occurred later than for children in the treatment groups ($M = 1.67$, $SD = .54$ of 4 trial blocks), $F(1, 24) = 15.29$, $p < .001$, $d = 4.87$. Thus, once children discovered the linear representation following treatment ($M = 1.47$ trial blocks), they generally continued to use it throughout the experimental session.

In summary, both the rapidity of the change in estimates and its stability once it was made suggest that the change was made at the level of the entire representation, rather than as a local repair, a conclusion that the data on the path and breadth of change (in the next sections) also supported.

Path of change. Children could have moved from a logarithmic to a linear representation 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 because the linear function provided a better fit than the logarithmic (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 to a linear representation. 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 gradually increased, 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 a shift from a logarithmic to linear representation.

A second hypothesis was that the path of change involved a discontinuous switch from a logarithmic to a linear representation, with no intermediate state (see also, Opfer & Siegler, 2007). 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 Figure 4, from Trial Block −3 to Trial Block −1, a one-way ANOVA on the linear regression functions indicated that there was no change in the fit of the linear function ($F < 1$, *ns*).

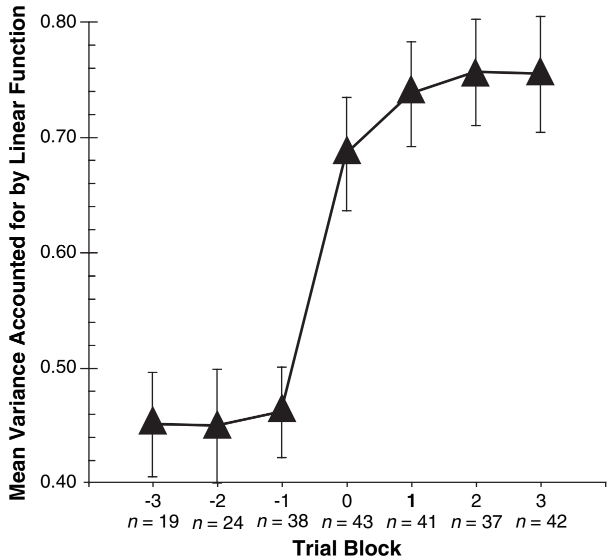

Figure 4. Path of change: Backward trials graph of fit of linear model to children's estimates on number line (training) task.
Note. The 0 trial block is the block on which the linear function first provided a better fit to each child's estimates; the −1 trial block is the block before that, and so on. The *N*s indicate the number of children who contributed data at each trial block; thus, 43 children used the linear representation on a least one trial block and therefore contributed data to Trial Block 0, 41 of these children had at least one trial block after this point and therefore contributed data to Trial Block 1, and so on.

There also was no change from Trial Block 0 to Trial Block 3 in the fit of the linear function ($F < 1$, *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, from an average $R^2 = .46$ to an average $R^2 = .68$, $F(1, 80) = 12.20$, $p < .001$, $d = 2.67$. 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.

Variability of change. We last examined individual differences in responses to treatment. Previous work (Opfer & Siegler, 2007) found that children whose initial representations were consistently *logarithmic* responded to treatment by adopting representations that were more consistently *linear* than did children whose initial representations were less consistently logarithmic; in contrast, the fit of the linear function to children's number line pretest estimates did not predict the fit of the linear function to their number line posttest estimates. The explanation previously offered for this finding was that the difference between the children's estimates and the feedback they received was more dramatic, and thus more likely to motivate a shift to the alternative (linear) representation, among children whose initial estimates were most strongly logarithmic.

In an attempt to replicate this finding, we regressed the percentage of variance in number line pretest estimates accounted for by the logarithmic function against percent variance in number line posttest estimates accounted for by the linear function for the children in the treatment groups. As previously found, the fit of the logarithmic model to each child's number line pretest estimates (mean log $R^2 = .72$, $SD = 0.20$) predicted the fit of the linear model to the child's number line posttest estimates (mean lin $R^2 = .71$, $SD = 0.34$), $r = .5$, $F(1, 23) = 7.36$, $p < .05$. That is, the better the logarithmic model fit the children's number line pretest estimates, the better the linear model fit their number line posttest estimates. Surprisingly, the fit of the linear function to children's number line pretest estimates (mean lin $R^2 = .62$, $SD = 0.29$) also predicted the fit of the linear function to their number line posttest estimates, $r = .8$, $F(1, 22) = 38.29$, $p < .01$, whereas it did not in Opfer and Siegler (2007). Thus, it seems that having consistent representations of numeric value (whether accurate or not) produces the greatest learning.

Transfer of Learning to Numerical Categorization

Finally, to test our practice interference hypothesis, we examined the *breadth of changes* in children's numerical judgments by examining whether children transferred learning on number line problems to their performance on the number categorization task and whether this transfer was impeded by previous practice on the categorization task.

To examine these issues, we first analyzed the relation between numerical value and categorization judgments at pretest. To do so, category labels were converted to a numeric code (i.e., *very small* = 0, *small* = 1, *medium* = 2, *big* = 3, and *very big* = 4), and then we examined the fit of the linear and logarithmic regression functions to the mean judgments. As on the number line task, children's pretest magnitude judgments were again better fit by a logarithmic (log $R^2 = .95$) than by a linear function (lin $R^2 = .69$). These fits of the logarithmic function did not result from aggregating over subjects: Of all the children who received a categorization pretest, 90% of individual children provided patterns of judgments that were better fit by the logarithmic than linear function, with the average fit of the logarithmic function to each child's judgments (mean log $R^2 = .78$, $SD = 0.03$) being significantly better than the average fit of the linear function (mean lin $R^2 = .59$, $SD = 0.05$), $t(30) = 5.71$, $p < .001$,

d = 4.61. Furthermore, pretest performance on the two tasks was generally correlated: The more linear were the estimates on the number line task, the more linear were the judgments on the number categorization task, r = .52, $F(1, 30)$ = 10.92, $p < .01$, and the more logarithmic were the estimates, the more logarithmic were the judgments on the number categorization task, r = .72, $F(1, 30)$ = 30.41, $p < .001$. Thus, it appeared that a common, logarithmic representation of numeric value influenced children's categorization as well as estimation performance.

As an overall measure of transfer, we next examined whether individual differences in learning to generate linear estimates on the number line estimation task (as measured by the fit of the linear regression function to estimates on the number line posttest) were associated with individual differences in the linearity of judgments on the number categorization task (also measured by the fit of the linear regression function). If children who had learned to generate linear estimates on the number line task failed to generalize their learning to the categorization task, we would expect no correlation between the linearity of judgments on the two tasks. This was not the case, however. Rather, the linearity of judgments across the two tasks were highly correlated, r = .67, $F(1, 55)$ = 44.42, $p < .0001$. Could this correlation have arisen simply because the two tasks tapped a third factor unrelated to learning? If so, the correlations would be expected to be equally high in both the training and the control groups. This was not the case, either. Rather, the correlation in performance across the two tasks was very high for the treatment groups (r = .91), as would be expected by transfer of learning, but very low for the control groups (r = .20), as would be expected with no transfer of learning to the categorization context.

We next examined the effect of categorization pretesting on transfer. As expected, category judgments on posttest varied substantially with the administration of treatment and categorization pretest (see Figure 5). At the group level, the linear function provided a better fit for the mean judgments of children in the unpretested treatment group (lin R^2 = .84) than in the pretested treatment (lin R^2 = .72), pretested control (lin R^2 = .75), and unpretested control (lin R^2 = .68) groups. The same pattern emerged when looking at the proportion of children who were best fit by each function, with 46% of children's judgments in the unpretested treatment group being best fit by the linear function versus 23% of children in the pretested treatment group, 21% of children in the unpretested control group, and 17% of children in the pretested control group.

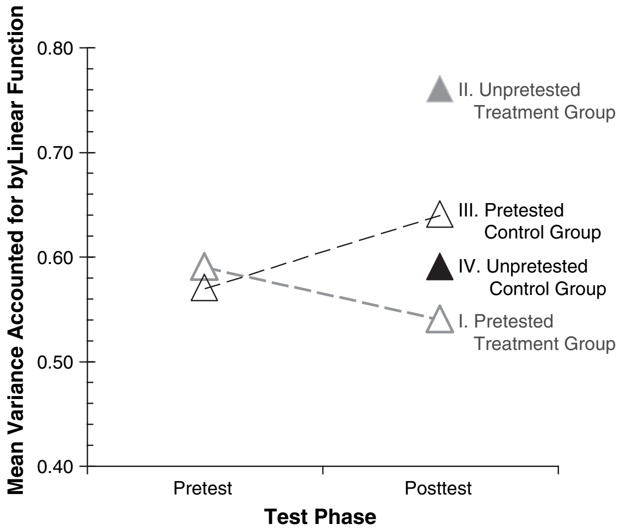

Figure 5. Breadth of change.
Note. Fit of linear model to children's category judgments (transfer task) across four groups: Pretested treatment group (Group I), unpretested treatment group (Group II), pretested control group (Group III), and unpretested control group (Group IV).

Finally, to test for the predicted interaction between categorization pretesting and treatment, we conducted a 2 (categorization pretesting: yes vs. no) × 2 (treatment: yes vs. no) factorial ANOVA on the fit of the linear function for each child's judgments. Categorization pretesting and treatment produced no main effects, but there was a substantial interaction between the two variables, $F(1, 56)$ = 5.32, $p < .05$. The unpretested treatment group provided significantly more linear judgments (mean R^2 = .76, SD = 0.02) than did the pretested treatment group (mean R^2 = .54, SD = 0.07), $t(22)$ = 2.43, $p < .05$, d = 4.27, and they also provided slightly more linear judgments than did the control groups (pretested: mean R^2 = .64, SD = 0.03; unpretested: mean R^2 = .59, SD = 0.07; ps < .07), which did not differ from each other. To appreciate just how powerful transfer was when children received treatment but no categorization pretest, it is useful to compare the unpretested treatment group's performance on the categorization task (shown in Figure 5) to their last trial blocks of performance on the estimation task (shown in Figure 4): The variance accounted for by the linear function in the two tasks is nearly identical (estimation: mean lin R^2 = .79; categorization: mean lin R^2 = .76), which is consistent with an almost perfect transfer of a linear representation across the two contexts.

In summary, the results regarding transfer supported two major conclusions. First, changes in estimation again seemed to occur at the level of the whole representation rather than as a series of local (i.e., task- and treatment-specific) repairs, thereby leading children to generalize a little treatment both to new numerical ranges within the same task (the number line task) and to the whole numerical range on a new task (the number categorization task). Second, as predicted by the pretest interference hypothesis, the initial *use* of the logarithmic representation on the number categorization pretest made it more difficult to change performance on this task, thereby leading to the finding that quite robust transfer was almost completely blocked by the administration of a categorization pretest.

Discussion

Spontaneous transfer of learning is notoriously difficult to elicit (Barnett & Ceci, 2002; Gick & Holyoak, 1983; Thorndike, 1922), even in microgenetic studies that allow one to control for the amount of learning that occurs prior to administration of the transfer task (Siegler, 2006). We hypothesized that one source of this difficulty comes from previous experience with the transfer task, such as that provided on a pretest. To test our hypothesis about this difficulty with transfer, we used Solomon–Lessac's four-group design to examine how children acquired linear representations of numerical magnitude on an estimation task (see the Process of Change in Numerical Estimation section), whether they transferred these representations to a new context in which they were asked to categorize numbers, and whether pretesting interfered with this transfer (see the Transfer of Learning to Numerical Categorization section).

Regarding the process of change in numerical estimation, our study lent further support to the representational change hypothesis about development of estimation (Opfer & Siegler, 2007; Opfer & DeVries, in press; Siegler & Booth, 2004; Siegler & Opfer, 2003; Thompson & Opfer, in press). replicating results from the only previous microgenetic study of numerical estimation (Opfer & Siegler, 2007) in fine detail. As in Opfer and Siegler (2007), we found that feedback in a single session yielded a large increase in accuracy of estimates, and this change was associated with a shift from generating logarithmic to linear estimates (see the Source of Change section). Further, this change occurred after feedback on the magnitude of just one numeral—150—and was remarkably stable thereafter (see the Rate of Change section), with the linearity of an individual child's estimates typically increasing greatly in one trial rather than gradually over many trials (see the Path of Change section). These three findings were important because they supported the idea that children's estimates improved due to a shift from a logarithmic to a linear representation of numerical quantity.

Regarding the transfer of learning to numerical categorization, our study of transfer (see the Transfer of Learning to Numerical Categorization section) revealed surprisingly robust transfer of learning from the estimation context to the categorization context—at least when no categorization pretest was administered. That is, children who were given feedback and no pretest (i.e., the unpretested treatment group) categorized 1 as "very small," the numbers 78 and 150 as "small," and 1,000 as "very large," which was consistent with the linear representation they learned to use in the estimation context. Indeed, the most striking evidence for transfer of representations across the two contexts came from performance of the unpretested treatment group shown in Figure 5 (which depicts linearity of number categorization) to the last trial blocks shown in Figure 4 (which depicts linearity of number line estimates after learning to generate linear representations): The variance accounted for by the linear function in the two tasks is nearly identical (estimation: mean lin R^2 = .79; categorization: mean lin R^2 = .76). This performance is consistent with an almost perfect transfer of a linear representation across the two contexts. In contrast, when children were given neither treatment nor categorization pretest, they categorized the number 1 as "very small," the number 78 as "medium," and the numbers 150 and 1,000 as "very large," which was consistent with use of a logarithmic representation. Indeed, as Figure 5 depicts, children who had received no feedback (i.e., children in the control groups) and children who had been pretested on the transfer task (i.e., children in the pretested groups) generated category judgments that were just as nonlinear before training as after training. Thus, without the administration of a categorization pretest, we observed nearly perfect transfer to the categorization task; with the administration of a categorization pretest, we observed virtually no transfer at all.

Why Did the Pretest Interfere With Transfer?

One mechanism that might account for the pretest interfering with transfer is simple fatigue. Within this hypothesis, our giving pretested children so many problems to solve reduced their performance relative to the unpretested children because the pretest caused children to become tired of the experiment. As a direct test of whether fatigue caused the pretested and unpretested groups to differ, we ran an additional condition consisting of 20 children who were given an equally long, but nonnumeric pretest (identifying the category labels of 10 animals) before estimation training. For these 20 children who received the animal pretest, pretesting did nothing to block transfer of estimation training: Their numeric categorizations were quite linear after estimation training (R^2 = .73), almost identical to those of the unpretested treatment group in Figure 5 (R^2 = .76), and more linear than both children pretested for numeric categorization and given the treatment (R^2 = .54) and children who received no estimation training (R^2 = .47). Thus, our follow-up study indicated that *the numeric content of the categorization pretest—and not its length—blocked transfer*.

Another mechanism that might account for the categorization pretest interfering with transfer is children avoiding inconsistent answers. The strongest version of this hypothesis would maintain that children remembered their categorization pretest answers and repeated them on categorization posttest to appear consistent; a weaker (though perhaps less plausible) version of the hypothesis would be that children remembered which representation they used on categorization pretest and continued to use that representation to appear consistent. Although we cannot directly rule out either hypothesis (because we did not include a condition in which children answered questions secretly, which would allow them to change their answers while saving face), at least their literal answers did differ between categorization pretest and posttest. For example, among the children in the two groups that received a categorization pretest, most changed their category judgments to some degree between pretest and posttest (e.g., on categorization posttest, children tended to provide judgments that were generally smaller than on pretest), though almost all continued to provide category judgments consistent with a logarithmic representation. Thus, unless children had some remarkable insight into the distribution of estimates they were providing (which the experimenters themselves could not discern without performing statistics), it seems unlikely that the categorization pretest interferes with transfer because children avoid giving inconsistent answers.

In our view, categorization pretests interfered with transfer for an altogether different reason: Practice on the number categorization pretest strengthened the association between the logarithmic representation and the categorization pretest, and thereby blocked transfer of the more optimal (but underpracticed) linear representation learned during training on the estimation task. This explanation is consistent not only with the results presented and our follow-up condition but also consistent with previous findings of multiple representations of numeric magnitudes (Holyoak, 1978; Siegler & Opfer, 2003), with interference of the linear representation in adults' performance on a fractional magnitude task (Opfer & DeVries, in press; Thompson & Opfer, in press), with findings of set effects in analogy and transfer (Duncker, 1945; Gick & Holyoak, 1980; Luchins, 1942, Luchins & Luchins, 1950), and with the formation of memory traces through practice (Roediger & Payne, 1982; Shrager & Siegler, 1998). To our knowledge, however, the implications of this research for the possibly harmful effects of repeated testing on transfer has not been examined previously, despite its importance in interpreting the narrow transfer of learning sometimes observed in microgenetic studies.

If correct, our explanation for children's trouble with transfer in this experiment suggests a quite widespread obstacle to children's transfer of learning. Specifically, our depiction of pretests being associated with the representations used to solve them is clearly analogous to Duncker's (1945) observation of functional fixedness. In both cases, previous practice leads to something (in our case, a numerical representation; in Duncker's, memory for functions) becoming more resistant to change with increasing use. A broader conceptual fixedness is also evident in the mental set (*Einstellung*) phenomena studied by Luchins and Luchins (1950), who used many different strategies (all unsuccessful) to induce students to abandon the use of an equation that yielded correct solutions for one problem type but that was less than ideal for other types of problems. And a similar conceptual fixedness also seems evident in the ''change resistance'' phenomena described by McNeil and Alibali (2004, 2005), in which children's concept of what follows the equals sign (''= __'') becomes strongly associated with the instructions to complete arithmetic operations, leading to their difficulty in learning to solve problems (e.g., 7 + 4 + 5 = 7 + __) that deviate from the practiced form (e.g., 7 + 4 + 5 = __). In all of these cases, performance on transfer tasks that had been well practiced were difficult to improve by training on different but related tasks. On transfer tasks that are not well practiced, however, surprisingly robust transfer of learning can be obtained. For example, throwing darts at an underwater target is probably not something that students perform everyday, yet in Hendrickson and Schroeder's (1941) classic study of transfer, a brief classroom lesson on light refraction was sufficient to improve students' aim. Thus, this handful of examples suggest that the variable of previous practice can explain some of the differences in transfer that researchers have already observed.

Implications for Microgenetic Studies

The finding that pretesting inhibited transfer in our microgenetic study has a number of implications for future microgenetic studies. First, the primary implication is *not* to abandon pretests or trial-to-trial assessments. The rationale for including pretests—allowing each learner to serve as his or her own control—remains valid and serves the critical function of also allowing researchers to characterize the rate and path of change. Without trial-to-trial assessments, it would have been impossible for us to test the representational change hypothesis about estimation, which predicted rapid and abrupt changes in the linearity of estimates. Moreover, these features of the change process were not affected by categorization pretesting in our study, nor would the practice interference hypothesis predict them to be given that children received feedback (treatment) on that task.

Rather, we take the primary implication to be that research on the breadth of change would be greatly enhanced by the use of the Solomon and Lessac (1968) four-group design. Pretesting on the categorization task strongly interacted with treatment in this study, with the result that transfer appeared almost perfect when no categorization pretest was given and almost nonexistent when a categorization pretest was given, a fact that is even more striking given that our transfer task required "near transfer" (Barnett & Ceci, 2002). Thus, previous reports that transfer is difficult to achieve may depend more on the administration of pretest than other variables, such as the particular type of knowledge being transferred (e.g., whether transfer is more difficult for novel categories than problem-solving strategies).

In summary, this study suggests that children may be much more likely to transfer their learning than previously supposed. A number of researchers have made this argument on both theoretical and empirical grounds (e.g., Anderson et al., 1996). In addition to these arguments, we believe we have identified an important mechanism for the trouble that children do have with transfer (viz. conceptual fixedness), and we have provided a demonstration of a potentially systematic methodological bias that has contributed to the impression that trouble with transfer is a necessary feature of learning. Assessing the magnitude of this bias will presumably require many more studies that control for the effect of pretesting.

References

Alibali, M. W., & Goldin-Meadow, S. (1993). Gesture-speech mismatch and mechanisms of learning: What the hands reveal about a child's state of mind. *Cognitive Psychology, 25*, 468–523.

Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational Researcher, 25*, 5–11.

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.

Bassok, M., & Holyoak, K. J. (1989a). Interdomain transfer between isomorphic topics in algebra and physics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 153–166.

Bassok, M., & Holyoak, K. J. (1989b). Transfer of domain-specific problem solving procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 522–533.

Biederman, I., & Shiffrar, M. M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 640–645.

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

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

Brown, A. L., & Kane, M. J. (1988). Preschool children can learn to transfer: Learning to learn and learning from example. *Cognitive Psychology, 20*, 493–523.

Brown, A. L., Kane, M. J., & Echols, C. H. (1986). Young children's mental models determine analogical transfer across problems with a common goal structure. *Cognitive Development, 1*, 103–121.

Carraher, T. N., Carraher, D. W., & Schliemann, A. D. (1985). Mathematics in the streets and in the schools. *British Journal of Developmental Psychology, 3*, 21–29.

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.

Cognition and Technology Group at Vanderbilt. (1997). *The Jasper project: Lessons in curriculum, instruction, assessment, and professional development*. Mahwah, NJ: Erlbaum.

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.

Conlee, J. W., & Parks, T. N. (1981). Age-dependent and position-dependent effects of monaural acoustic deprivation in nucleus magnocellularis of chicken. *Journal of Comparative Neurology, 202*, 373–384.

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

Dowker, A. (2003). Young children's estimates for addition: The zone of partial knowledge and understanding. In A. J. Baroody & A. Dowker (Eds.), *The development of* *arithmetic concepts and skills: Constructing adaptive expertise* (pp. 243 – 265). Mahwah, NJ: Erlbaum.

Duncker, K. (1945). On problem-solving. *Psychological Monographs*, *58*, 1 – 113.

Erdelyi, M. H., & Becker, J. (1974). Hypermnesia for pictures: Incremental memory for pictures but not words in multiple recall trials. *Cognitive Psychology*, *6*, 159 – 171.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, *12*, 306 – 355.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*, 1 – 38.

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

Hendrickson, G., & Schroeder, W. H. (1941). Transfer of training in learning to hit a submerged target. *Journal of Educational Psychology*, *35*, 205 – 213.

Holyoak, K. J. (1978). Comparative judgments with numerical reference points. *Cognitive Psychology*, *10*, 203 – 243.

Kaminski, J. A. (2006). *The effects of concreteness on learning, transfer, and representation of mathematical concepts*. Unpublished doctoral dissertation, The Ohio State University, Columbus.

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.

Lave, J. (1988). *Cognition in practice: Mind, mathematics, and culture in everyday life*. Cambridge, UK: Cambridge University Press.

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.

Lessac, M. S. (1965). *Effects of early isolation and restriction on the later behavior of beagle puppies*. Unpublished doctoral dissertation, University of Pennsylvania.

Lessac, M. S., & Solomon, R. L. (1969). Effects of early isolation on the later adaptive behavior of beagles: A methodological demonstration. *Developmental Psychology*, *1*, 14 – 25.

Luchins, A. (1942). Mechanization in problem solving. *Psychological Monographs*, *54* (6, Serial No. 248).

Luchins, A. S., & Luchins, E. (1950). New experimental attempts at preventing mechanization in problem solving. *Journal of General Psychology*, *42*, 279 – 297.

McNeil, N. M., & Alibali, M. W. (2004). You'll see what you mean: Students encode equations based on their knowledge of arithmetic. *Cognitive Science*, *28*, 451 – 466.

McNeil, N. M., & Alibali, M. W. (2005). Why won't you change your mind? Knowledge of operational patterns hinders learning and performance on equations. *Child Development*, *76*, 1 – 17.

Meiklejohn, A. (1908). Is mental training a myth? *Educational Review*, *37*, 126 – 141.

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

National Council of Teachers of Mathematics. (1980). *An agenda for action: Recommendations for school mathematics in the 1980s*. Reston, VA: Author.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (2001). *Principles and standards for school mathematics*. Reston, VA: Author.

Nieder, A., & Miller, E. K. (2003). Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron*, *37*, 149 – 157.

Opfer, J. E., & DeVries, J. M. (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. (2004). Revisiting preschoolers' *living things* concept: A microgenetic analysis of conceptual change in basic biology. *Cognitive Psychology*, *49*, 301 – 332.

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

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249 – 255.

Roediger, H. L., III, & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1155 – 1159.

Roediger, H. L., III, & Payne, D. G. (1982). Hypermnesia: The role of repeated testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*, 66 – 72.

Rubel, E. W. (1984). Ontogeny of auditory-system function. *Annual Review of Physiology*, *46*, 213 – 229.

Schoenfeld, A. H. (1985). *Mathematical problem solving*. New York: Academic Press.

Shrager, J., & Siegler, R. S. (1998). SCADS: A model of children's strategy choices and strategy discoveries. *Psychological Science*, *9*, 405 – 410.

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York: Oxford.

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

Siegler, R. S. (2006). Microgenetic analyses of learning. In W. Damon & R. M. Lerner (Series Eds.) & D. Kuhn & R. S. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (6th ed., pp. 464 – 510). New York: Wiley.

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

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., & Shipley, C. (1995). Variation, selection, and cognitive change. In T. Simon & G. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling*. (pp. 31 – 58). Hillsdale, NJ: Erlbaum.

Singley, K., & Anderson, J. R. (1989). *The transfer of cognitive skill*. Cambridge, MA: Harvard University 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.

Spiro, R. J., Feltovich, P. J., Jacobson, M. J., & Coulson, R. L. (1991). Cognitive flexibility, constructivism, and hypertext: Random access instruction for advanced knowledge acquisition in ill-structured domains. *Educational Technology*, *11*, 24 – 33.

Thompson, C. A., & Opfer, J. E. (in press). Costs and benefits of representational change: Effects of context on age and sex differences in symbolic magnitude estimation. *Journal of Experimental Child Psychology*.

Thorndike, E. L. (1922). The effect of changed data upon reasoning. *Journal of Experimental Psychology*, *5*, 33 – 38.

van den Boom, D. C. (1994). The influence of temperament and mothering on attachment and exploration: An experimental manipulation of sensitiveness among lower-class mothers with irritable infants. *Child Development*, *65*, 1457 – 1477.

van den Boom, D. C. (1995). Do first-year intervention effects endure? Follow-up during toddlerhood of a sample of Dutch irritable infants. *Child Development*, *66*, 1798 – 1816.