How Does Change Occur: A Microgenetic Study of Number Conservation

ROBERT S. SIEGLER

Carnegie Mellon University

Microgenetic methods can illuminate the path, rate, breadth, variability, and sources of change. The present study illustrates the types of information the method can yield in the context of number conservation. Five-year-olds whose pretest performance showed that they had not mastered number conservation were presented four training sessions. Some were just given feedback on their number conservation performance; others were given feedback and asked to explain their reasoning; yet others were given feedback and asked to explain the reasoning that led to the experimenter’s judgment. Being asked to explain the experimenter’s reasoning produced considerably more learning than either of the other two procedures. The learning involved two distinct realizations: that relative length did not predict which row had the greater number of objects and that the type of quantitatively relevant transformation did. Individual children generated multiple types of reasoning about conservation, both on the pretest and throughout the training procedure; understanding the importance of the type of transformation did not lead to immediate rejection of less advanced forms of reasoning, even for a single problem presented several times during the experiment. This variability was positively associated with learning; children who showed greater variability of reasoning on the pretest, both within and across trials, learned more. Educational and theoretical implications of children’s efforts to understand other people’s reasoning were discussed. © 1995 Academic Press, Inc.

How change occurs is perhaps the single, fundamental issue in the study of cognitive development. Progress in understanding the issue has been slow, however (Brown & DeLoache, 1978; Flavell, 1984; Miller, 1992; Sternberg, 1984). Part of the reason is the inherent conceptual complexity of the subject. Understanding changes in children’s thinking presents all of the demands of understanding their thinking at any one time, plus the added demands of understanding what is changing and how the change is being accomplished.

This research was supported by grants from the Spencer Foundation and the National Institutes of Health (HD-19011). We thank the children, teachers, and administrators of the Carnegie Mellon Day Care Center, the Carnegie Mellon Children’s School, Beth El Nursery School, Temple Emanuel Nursery School, and St. Paul Nursery School, who made the project possible, as well as Alice Siegler for help collecting data. We also thank Shari Ellis and Patrick Lemaire for providing particularly helpful comments on previous versions of the manuscript. Correspondence and reprint requests should be addressed to Robert S. Siegler, Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213.
A further reason for the slow progress has been the difficulty of formulating appropriate methods for studying the change process. Most experiments on cognitive development compare the performance of children of different ages. For example, many studies have contrasted the understanding of conservation of 5-, 6-, and 7-year-olds. Sometimes the studies are cross-sectional, comparing the performance of different children; other times they are longitudinal, comparing the performance of the same children at different ages. Both types of studies are often revealing regarding the general outline of changes; the longitudinal designs also can yield valuable information regarding stability of individual differences over time.

In both cases, however, the studies are less revealing regarding exactly how changes occur. The problem is that examining children at widely-spaced intervals, such as the 1- to 3-year intervals that are typical in both cross-sectional and longitudinal studies of development, yields too low a density of observations to allow precise conclusions about the ongoing process of change. The type of information that emerges is like that in before-and-after snapshots; what we want is more like the nearly continuous flow of information in movies. The low density of observations reduces the potential of these studies to advance conceptual understanding of changes in children’s thinking, limiting what we can learn about particular changes and not allowing rigorous tests of general hypotheses about how change occurs.

Microgenetic methods offer a means for obtaining the type of fine-grain information that seems necessary to advance understanding of cognitive change (Siegler & Crowley, 1991). Such methods are defined by three characteristics: (1) Observations span as large a portion as possible of the period during which rapid change in the particular competence occurs, (2) the density of observations within this period is high relative to the rate of change of the phenomenon, and (3) Observations are subjected to intensive trial-by-trial analysis of both qualitative and quantitative properties, with the goal of inferring the processes that give rise to the change.

The second characteristic is especially important. Densely sampling changes while they are occurring allows the kind of temporal resolution needed to inform our understanding of change. It provides the data needed to discriminate among alternative hypotheses about what actually goes on during periods of rapid change, rather than limiting understanding to what can be inferred from performance before and after the change.

The central contention of this article is that microgenetic studies can substantially increase our understanding of how change occurs, even regarding developments that have been studied extensively with other methods. To illustrate this point, a microgenetic method is applied to one of the best-studied problems in the cognitive developmental literature:
number conservation. A survey 15 years ago (Murray, 1978) located more than 140 studies of the development of conservation beyond Piaget's (1952) pioneering studies of the topic. Many more have been conducted since then (see Field, 1987, and McEvoy & O'Moore, 1991, for reviews). Despite the size of this literature, however, understanding of how the change occurs has remained at a rather superficial level. The present investigation illustrates the types of gains in understanding of change that microgenetic analyses can bring.

THE NUMBER CONSERVATION PROBLEM

In Piaget's (1952) classic number conservation problem, 4- to 7-year-olds are shown two parallel rows of objects. Initially, the rows have the same number of objects (6-10); corresponding objects in each row are located directly opposite each other. After the child agrees that there are the same number of objects in each row, the experimenter lengthens or shortens one of the rows and asks whether they still have the same number. Piaget saw this task as a particularly good measure of children's understanding of number, because success on it required ignoring perceptually-misleading cues and focusing on logical properties such as reversibility, compensation, and the type of transformation that had been performed.

Piaget (1952) described children as progressing through three stages of understanding. Children in Stage I were said to be unable to integrate the dimensions of length and density and therefore to focus solely on one of them, usually the relative length of the rows. They typically would choose the longer row as having more objects. Children in Stage II were described in a large variety of ways: as vacillating from trial to trial between reliance on length and reliance on density; as admitting that they did not know which row has more; and as advancing correct answers but then being unable to explain them. Children in Stage III were said to judge correctly, because they understood the logic of reversibility, compensation, identity, and the type of transformation.

Subsequent studies have extended Piaget's (1952) observations in a number of ways. They have shown that 3- to 5-year-olds often perform more successfully if the rows include small rather than large numbers of objects (Gelman, 1972; Winer, 1974), if the transformations involve addition or subtraction rather than neither adding nor subtracting (Halford, 1982; Siegler, 1981); if the wording of questions is facilitative (Markman, 1979; Donaldson & Wales, 1970), if they think the transformation was accidental rather than intentional (Light, Buckingham, & Robbins, 1979; McGarrigle & Donaldson, 1974; Neilson & Dockrell, 1982), or if the children are trained through presentation of rules, feedback, modeling, social interaction with a conserving peer, discrimination learning sets, or any of
a host of other instructional methods (Beilin, 1977; Field, 1987). Investigators also have identified several characteristics of children that are correlated with the success of training procedures: age, prior relevant knowledge, and initial incongruity between gestures and verbal explanations, among them (Brainerd, 1977; Church & Goldin-Meadow, 1986; Inhelder, Bovet, & Sinclair, 1974; Strauss & Langer, 1970).

Thus, we know a considerable amount about task variables that influence success on number conservation tasks, about training procedures that allow children to acquire more advanced understanding, and about characteristics of children that influence the effectiveness of training. Our knowledge of how the transitions occur, however, remains severely limited, as illustrated in the next section.

A TAXONOMY OF DIMENSIONS OF COGNITIVE CHANGE

Using microgenetic methods forces us to think seriously about the dimensions along which changes occur. Five dimensions that seem particularly important are the path, rate, breadth, variability, and sources of change. Subsets of these dimensions have been examined within approaches emphasizing stages of development (e.g., Case, 1985; Piaget, 1952), stability over time of individual differences (e.g., Collins, 1991; McCall, Applebaum, & Hogarty, 1973), and effective training procedures (e.g., Beilin, 1977; Field, 1987). The types of data on change yielded by microgenetic studies, however, require consideration of all of them. Below, each dimension is briefly described and then considered in the context of number conservation.

The Path of Change

One basic issue regarding many types of changes concerns whether children progress through qualitatively distinct understandings on their way to mature competence, and if so, what the qualitatively distinct understandings are. This issue has been considered most often in the context of research on stages and sequences. For example, Flavell (1971) noted that one of the four main assumptions of stage theories was that children progress through an invariant sequence of qualitatively distinct knowledge states on their way to advanced understanding.

Since Piaget's (1952) description of the sequence of understandings leading to mastery of number conservation, several variants of the basic sequence have been proposed. Most, though not all, accounts agree that at a relatively early point in development, children base number conservation judgments on the relative lengths of the rows. Most accounts also agree that children eventually base judgments on logical properties, such as identity, reversibility, compensation, or the type of quantitative transformation.
No agreement has been reached, however, concerning the transitional period during which children move from one understanding to another. A number of paths have been suggested. One hypothesis is that reliance on relative length of rows is followed directly by reliance on the type of numerically-relevant transformation (Lawson, Baron, & Siegel, 1974). A second hypothesized path posits the same beginning and end states, but suggests that between them, children solve problems empirically by counting the number of objects in the two rows (Gelman & Tucker, 1975; Siegler, 1981; Simon, Newell, & Klahr, 1991). Three other developmental paths that begin with reliance on relative length and end with reliance on the type of transformation are suggested by Piaget's (1952) and Inhelder, Sinclair, and Bovet's (1974) descriptions. In some passages, children are described as vacillating among different answers and logics during the transitional period; in others, they are described as often saying that they do not know; in yet others, they are described as judging correctly during the transition but not being able to explain their judgments.

The greater agreement about the beginning and end of the sequence than about the middle is attributable in part to the inherently greater difficulty of assessing rapidly changing knowledge states. However, it also seems due in part to limitations of the knowledge assessment methods that have been used. Standardized cognitive assessment methods, such as rule assessments and chronometric techniques, work well when knowledge is stable over prolonged periods. However, they work less well when knowledge is changing rapidly. They often are unable to reveal brief-lived transitional strategies or to reflect the highly fragile, changeable, and internally-inconsistent behavior characteristic of transitions. Within these methods, such instability is relegated to the error variance term.

Different problems characterize the less standardized methods used by Piaget and his co-workers. These techniques have been useful for yielding hypotheses about transitions, but the lack of clearly defined, consistently applied, criteria for categorizing children’s behavior has made it extremely difficult to derive quantitative statements from them about the frequency of different approaches. The experimenter's varying role from child to child has added further uncontrolled variability to the assessments. Overall, the limitations of these methods and the inherent difficulty of the problem have resulted in numerous hypotheses being generated about what goes on during transitional periods, without data adequately for choosing among them.

One way out of this impasse may be provided by methods that assess children's thinking on a trial-by-trial basis according to clear, consistently applied criteria. Such dense sampling of children's behavior can reveal whether they (1) progress directly from one relatively enduring under-
standing to another, (2) proceed from one enduring understanding to one or more brief lived transition strategies before settling on another enduring understanding, or (3) vacillate between earlier and later enduring understandings during the transitional period.

Previous studies that have used such trial-by-trial assessments suggest that the complexity of the developmental sequence may not be limited to brief-lived transitional periods. Rather than varied ways of thinking only being present briefly, children of a given age have been found to think in multiple ways about the same concept and to use multiple strategies to solve a given kind of problem over a period of years. Such prolonged use of multiple strategies has been found in domains as diverse as arithmetic, word identification, spelling, time telling, serial recall, and physics problem solving, and for people of all ages from 4 years to adulthood (Siegler, 1994). For example, 5-year-olds sometimes solve a simple addition problem such as $3 + 7$ by counting from one, sometimes by counting from the larger addend (i.e., by counting "7, 8, 9, 10"), sometimes by guessing, and sometimes by retrieving the answer from memory. Even a single child solving a single problem on two successive days quite often will use a more advanced strategy one day and a less advanced one on the next (Siegler & Jenkins, 1989). For example, a child might retrieve the answer to $3 + 7$ today, but count from one to solve it tomorrow. This variability is not limited to a short-lived transitional period—it lasts for a period of at least 4 years (Siegler, 1987; Siegler & Shrager, 1984). One of the main issues in the present study was whether children also think about classic Piagetian tasks, such as number conservation, in a variety of ways, and if so, how the path of change can be characterized.

The Rate of Change

A second basic issue regarding change concerns its speed. Flavell’s (1971) analysis of stage theories referred to their stance on this issue as the abruptness assumption. Relative to other approaches, stage theories depict important changes as occurring rapidly. Learning theories often reflect the opposite assumption—that changes occur only gradually.

Results of previous microgenetic studies have consistently supported the characterizations of cognitive change as gradual. This has been true even when children can verbalize early on why the new approach is more desirable than alternatives (e.g., Kuhn, Amsel, & O’Loughlin, 1988; Kuhn & Phelps, 1982; Metz, 1985; Schauble, 1990; Siegler & Jenkins, 1989).

The difference between the assumption of stage theories and the results that have arisen within microgenetic studies, however, may stem from differences in the types of problems emphasized within stage approaches and within previous microgenetic studies, rather than from any inherent
HOW DOES CHANGE OCCUR

gradualness of cognitive change. The tasks that have been studied microgenetically, such as acquisition of new arithmetic strategies and systematic scientific experimentation skills, confer some but not enormous benefits on how well children can solve the problems. For example, in single-digit arithmetic, counting from the larger addend is somewhat more accurate than counting from one, but not hugely so; both strategies produce accurate performance. The differences tend to be especially small when children first adopt the new strategy, because of a lack of practice in executing it.

Number conservation provides a quite different situation. Relying on reversibility, identity, or the type of transformation leads to much more accurate performance than relying on the relative lengths of the rows. Further, unlike the previous problems to which microgenetic methods have been applied, conservation is generally considered a matter of logic; if nothing is added or subtracted, the number of objects must remain the same. If children ever quickly adopt a new strategy and use it consistently, they would seem likely to do it when the superiority of the new approach lies in its basic logic. These differences between the domains may lead to a much faster consistent adoption of the new way of thinking in number conservation than in previously studied areas. Alternatively, gradual acquisition may prove to be the rule, even on logical tasks such as number conservation.

The Breadth of Change

Once children grasp a superior logic, how widely do they generalize it? Flavell (1971) labeled the basic assumption of stage theories on this issue the "concurrence assumption." The idea was that many cognitive capabilities that share a common underlying logic are acquired close together in time.

Subsequent research has shown that the concurrence assumption is not very viable. A great many factors influence the timing of acquisition of different competencies, and broad unities in the timing of change have been difficult to identify (though see Case, 1992; and Halford, 1993, for some exceptions). Thus, consideration has turned to whether understanding of a given task, such as number conservation, tends to be acquired in unified or piecemeal fashion.

Number conservation includes three types of transformations: addition, subtraction, and the null transformation (neither adding nor subtracting anything). Previous studies indicate that some children solve addition and subtraction problems but not problems involving the null transformation (Gelman, 1972; Siegler, 1981). These findings have been interpreted as indicating that children find it easier to learn about addition and subtraction than about the null transformation. This interpretation is
open to question, however. Children may solve addition and subtraction problems earlier not because the problems are inherently easier but because such problems are presented more often in the world outside the laboratory. Many parents present 4- and 5-year-olds with simple addition and subtraction problems of the form "Suppose you had two oranges and I gave you one more; how many would you have then?" In contrast, it seems unlikely that they present null transformation problems such as "Suppose you had two oranges and I didn't give you any or take any away; how many would you have then?" The microgenetic design of the present study allowed the relevant data to be obtained. If addition and subtraction transformations are in fact easier to learn, then accuracy in answering them should grow more rapidly during the training sessions than accuracy in answering null transformation problems. On the other hand, if the inherent difficulty is equal, but children have more experience with addition and subtraction problems, they should do better on such problems initially, but the difference in performance should not increase over the course of training in the experiment.

A second issue regarding the breadth of change in number conservation involves whether initial use of transformational reasoning is influenced by the relation between the relative lengths of the rows and their relative quantities. Children may initially rely on the type of transformation most often when the longer row has more objects; this would occur if absence of conflicting spatial cues helped children to use their most advanced reasoning. Alternatively, they may initially rely on transformational reasoning most often when the longer row has fewer objects; this would occur if children most often used their most advanced reasoning when answering the particular problem correctly demanded it. A third possibility was that children would be impervious to potential conflicting or supportive spatial cues once they realized the importance of the type of quantitative transformation. Again, a microgenetic study including all three types of transformations in situations with conflicting and supportive spatial cues would provide the type of data needed to determine how problem characteristics influenced initial application of the new form of reasoning.

The Variability of Change

Issues regarding the variability of change patterns have been investigated primarily within long-term longitudinal studies of individual differences (e.g., McCall et al., 1973; Nesselroade, 1990; Plomin, 1986). Such studies have depicted cognitive changes at the relatively aggregated level of stability of overall test performance or factor structure, rather than in terms of specific cognitive processes. They also have examined performance at widely separated times. Thus, they have been useful for pro-
viding an overview of the variability of cognitive change, but have been less useful for obtaining in-depth information on how particular changes occur.

In contrast, most depictions of changes in understanding of specific concepts and problem solving skills have ignored the issue of variability of the change process. They have characterized changes in unqualified terms that imply that the pattern, if not the rate or breadth of change, is universal. The monolithic character of these depictions seems unlikely to have sprung from any deep conviction that all children progress through the same path of change. Instead, it seems attributable to traditional methods not yielding sufficiently rich data to differentiate among individual children’s change patterns, and to investigators therefore having little to say about the variability of change processes. Microgenetic methods promise to provide richer data regarding the path, rate, and breadth of change, which should allow more precise depiction of individual differences in change patterns.

The Sources of Change

Issues regarding the sources of change have been focused on most intensely within training studies. As noted in the context of number conservation, these studies have shown that a large variety of factors can cause cognitive changes to occur. The studies have been less informative, however, regarding how the changes occur. The emphasis has been on the percentage of children who learned, the stability of the learning over time, whether the learning transferred to untrained tasks, and whether it withstood countersuggestions, rather than on how the transition was accomplished.

Microgenetic studies resemble training studies in that they generally involve identifying one or more types of experiences that might contribute to cognitive change, providing intense exposure to such experiences, and observing children’s reactions to them. However, the trial-by-trial assessments of thinking that characterize microgenetic studies allow them to yield useful data not only regarding which procedures are effective but also regarding how a given procedure exercises its effects. For example, training studies might indicate that receiving feedback or learning rules or observing a conserving model can lead to conservation learning. Microgenetic analyses also yield this type of information, but yield additional types as well. In particular, they can indicate whether exposure to given experiences leads children to progress through qualitatively different states of knowledge; whether short-lived transition strategies mediate learning; whether different types of thinking coexist with each other and continue to be expressed for substantial periods of time; and whether learning tends to be specific to individual problems or is general across
classes of problems. Put another way, they allow analyses of how sources of change exercise their effects.

THE PRESENT STUDY

This study was an attempt to obtain the data needed to address these basic issues about the path, rate, breadth, variability, and sources of cognitive change. Preschoolers were presented an eight-session microgenetic procedure. In the first four sessions, they were pretested in order to identify individuals who could not yet solve large-set number conservation problems. In the following four sessions, these children were presented conservation problems under one of three conditions. Some children received feedback only; they were simply told that their answer was correct or incorrect. Other children were first asked to explain their reasoning and then given feedback regarding the correctness of their answers. Yet others received feedback about their answers and were then asked by the experimenter “How do you think I knew that?”

This last condition, in which the child needed to explain the experimenter’s reasoning, was of greatest interest. Studies of adults learning physics and computer programming have demonstrated that better learners tend to more actively explain to themselves textbook passages (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi & Van Lehn, 1991; Pirolli & Bielaczyc, 1989). Previously published articles, however, have not indicated whether the simple instruction to try to explain the reasoning of another, more knowledgeable person would have similar effects on individuals who did not engage in the activity spontaneously. Thus, a major goal of the study was to determine whether attempts to explain other people’s reasoning could be a source of this and other aspects of cognitive development.

This study also extended research on the effects of explaining other people’s reasoning to a population quite different from the ones studied previously. Earlier studies have focused on college students; the present study examined much younger individuals—5-year-olds. Ironically, such young children may be the population on which encouragement to explain other people’s reasoning would have the largest effects. Without encouragement to generate explanations, young children may be less likely than older individuals to search for the causes of events. When parents read books to children, they often intersperse such questions as “Why did the elephants fight?” or “Why was Babar’s mother so sad?” Such questions may be motivated by the intuition that the young children are capable of identifying the causes but will not necessarily do so without explicit encouragement. Consistent with this view, simply asking preschoolers to try to figure out why an event occurred has been shown to significantly
enhance their learning (Richards & Siegler, 1981). Thus, it seemed of particular interest to examine whether young children’s learning would benefit from requests to explain another person’s reasoning.

The condition in which children were asked to explain their own reasoning provided a control for the effects of generating explanations per se. If constructing explanations of conservation reasoning in and of itself enhances understanding of the concept, then these children should answer correctly more often than children in the group not asked to generate explanations. If generating an explanation is the only factor leading to benefits of explaining the experimenter’s reasoning, then the gains produced by the two conditions should be equivalent. Previous empirical data have been inconclusive regarding whether generating explanations per se enhances reasoning. In some studies, positive effects of explaining one’s reasoning have been found (Ahlum-Heath & DiVesta, 1986; Berry & Broadbent, 1984), in others they have not (Ellis, Siegler, & Klahr, 1993; Teasley, in press). Regardless of whether generating explanations improved reasoning in the present study, it was not expected to be as helpful as being asked to explain the experimenter’s reasoning, because the latter task would lead children to try to see the situation from the perspective of an older, more knowledgeable person who knew the correct answer.

METHOD

The experiment was organized into two parts: pretest and training. The pretest included the first four sessions and the training the final four.

Pretest

The pretest served three main purposes. One was to identify children who did not yet know how to solve large-set number conservation problems; these were the children who would participate in the subsequent training sessions. The second purpose was to determine what conservation knowledge these children did have, so as to identify predictors of learning in the subsequent training sessions. The third purpose was to provide a base of comparison for children’s performance in the training sessions.

Participants

The 97 children who were pretested attended a university-based preschool, a university-based day care center, or a day care center in a middle class community. All ranged in age from 54 to 73 months, with a mean of 64 months. The experimenter was a 42-year-old female research assistant.

Task and Materials

The basic task and materials were similar to those used in the classic Piagetian format, though the task involved a wider range of set sizes and transformation. Children were initially shown two parallel rows of white buttons, each 1.1 cm in diameter, with the individual buttons in each row in one-to-one correspondence. The buttons were placed on a large sheet of green construction paper so that they would be clearly visible. The experimenter and child sat on opposite sides of a small table, and the two rows of buttons were perpendicular to a hypothetical line between them, so that the child could refer to the row
"close to me" or the row "close to you" when explaining an answer. The distances between buttons within a row (measured from the center of one button to the center of the next) ranged from 2 to 3.1 cm, depending on the problem. The lengths of the rows for the small set size problems (2 to 4 objects), ranged from 3.8 to 10.2 cm, and the density of objects within each row from .26 to .6 objects/cm. The length of rows for the large set size problems (7-9 objects) ranged from 13.5 to 35.6 cm, and the density of objects within a row ranged from .2 to .6 objects/cm.

On each conservation problem, one of the two rows was transformed spatially and/or quantitatively, with the experimenter describing the transformation as she performed it. For example, on the standard Piagetian problem, the experimenter said "Now watch what I do. I'm going to spread out this row and not add anything to it and not take away anything from it." Following the transformation she asked "Now do they both still have the same number of buttons, or does one row have more buttons than the other?" If the child indicated that one row had more but not which one, the experimenter asked "Which one has more?" After the child answered, he experimenter asked "How did you know that?"

For each trial in the pretest and subsequent training sessions, two types of data on children's explanations were coded. One involved whether the child advanced a single explanation or more than one. On the large majority of trials, children advanced a single explanation, but on 6% of pretest trials and 4% of training session trials, children advanced more than one. The second type of data was the main explanation that the child advanced. On trials on which a child advanced multiple explanations, the last explanation advanced was used to represent the child's reasoning on the trial. Independent raters agreed on their coding of 99% of a sample of 100 trials randomly chosen to assess reliability.

Problems

Past research (e.g., Siegler, 1981) has shown that conservation judgments sometimes are based on spatial properties of the post-transformation spatial arrangement of the two rows (e.g., choosing the longer row), sometimes on the type of quantitative transformation that was performed (e.g., choosing a row to which an object has been added as having more, even if it is shorter), and sometimes on a combination of approaches (e.g., basing judgments of small sets on the type of quantitative transformation but basing judgments of large sets on the relative length of the rows). For this reason, problems were designed to allow detection both of reasoning based on the spatial configuration of the final arrangement and of reasoning based on the type of quantitative transformation that was performed.

In each of the four pretest sessions, children were presented 12 conservation trials. The trials in two sessions involved small sets (2-4 buttons); those in the other two sessions involved large sets (7-9 buttons). Within each set size, problems varied in two critical ways: the type of quantitative transformation and the type of spatial transformation. The quantitative transformations were adding an item to a row, subtracting an item from a row, or neither adding nor subtracting any items (null transformation). The spatial transformations were lengthening a row, shortening a row, or moving items in a row back and forth.

The six types of problems that were presented are illustrated in Table 1. Classification of problems was based on their spatial properties following the transformation. The problem types were defined as follows:

Equal problems. Items on which one row had its items moved back and forth, ending in the initial position, so that after the transformation, both the length and the density of transformed and untransformed rows were equal.

Length problems. Items on which the final configuration involved a longer and a shorter row, with the two rows having an equal density of buttons but the longer row having one more button than the shorter row. This was produced through one of two combinations of quantitative and spatial transformations: adding an object to a row being lengthened, or
subtracting an object from a row being shortened. As on each type of problem that involved addition and subtraction, half the trials involved addition and the other half subtraction.

**Density problems.** Items in which the final configuration involved two rows of equal length, but in which one row had a greater density (and therefore a greater number) of buttons. This was produced either by adding an object to the interior of a row and rearranging the objects to maintain equal spacing, or by subtracting an object from the interior of the row and performing the analogous rearrangement.

**Conflict-length problems.** Items in which the final arrangement involved one row being longer, the other having a higher density of objects, and the longer row having more objects. This was created either by adding an object to a row being lengthened or by subtracting an object from a row being shortened. This, like the other two types of conflict problems, is named for the length cue pointing to one answer and the density cue to the other (thus the "conflict") and for the type of cue that points to the correct answer (thus the "length").

**Conflict-density problems.** Items in which the final arrangement involved one row being longer, the other having its objects more densely packed, and the row with the more densely arranged objects having more objects. This was created either by adding an object to a row being lengthened or by subtracting an object from a row being lengthened.

**Conflict-equal problems.** Items on which the final arrangement involved one row being longer, the other having its objects more densely distributed, and the rows having equal numbers of objects. These classic Piagetian number conservation problems were created by simply lengthening or shortening a row without adding or subtracting any objects.

These six types of problems were chosen because they allowed assessment of strategies based either on spatial properties or on the type of quantitative transformation. In particular, if children consistently chose the longer row, they would be very accurate on the three types of length-consistent problems—equal, length, and conflict-length—and very inaccurate on the three types of length-inconsistent problems—density, conflict-density, and conflict-equal. Reliance on the density of the rows would produce the exact inverse pattern. Understanding the effects of addition and subtraction, but not of the null transformation, would lead to correct answers on the four types of problems in the middle of Table 1; if children relied on length when nothing was added or subtracted, they also would answer correctly on equal problems. Finally, accurately counting the number of objects in the two rows would

### Table 1

<table>
<thead>
<tr>
<th>Problem type</th>
<th>Initial configuration</th>
<th>Operation</th>
<th>Final configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equal</td>
<td>⚪️ ⚪️ ⚪️ ⚪️ ⚪️</td>
<td>Back and forth</td>
<td>⚪️ ⚪️ ⚪️ ⚪️ ⚪️</td>
</tr>
<tr>
<td>Length</td>
<td>⚪️ ⚪️ ⚪️ ⚪️ ⚪️</td>
<td>Add or subtract</td>
<td>⚪️ ⚪️ ⚪️ ⚪️ ⚪️ ⚪️</td>
</tr>
<tr>
<td>Density</td>
<td>⚪️ ⚪️ ⚪️ ⚪️ ⚪️</td>
<td>Add or subtract</td>
<td>⚪️ ⚪️ ⚪️ ⚪️ ⚪️ ⚪️</td>
</tr>
<tr>
<td>Conflict-length</td>
<td>⚪️ ⚪️ ⚪️ ⚪️ ⚪️</td>
<td>Add or subtract</td>
<td>⚪️ ⚪️ ⚪️ ⚪️ ⚪️ ⚪️</td>
</tr>
<tr>
<td>Conflict-density</td>
<td>⚪️ ⚪️ ⚪️ ⚪️ ⚪️</td>
<td>Add or subtract</td>
<td>⚪️ ⚪️ ⚪️ ⚪️ ⚪️ ⚪️ ⚪️</td>
</tr>
<tr>
<td>Conflict-equal</td>
<td>⚪️ ⚪️ ⚪️ ⚪️ ⚪️</td>
<td>Lengthen or shorten</td>
<td>⚪️ ⚪️ ⚪️ ⚪️ ⚪️</td>
</tr>
</tbody>
</table>
yield correct answers on all kinds of problems, as would deducing the outcome from the type of quantitatively-relevant transformation that had been performed.

**Procedure**

In each pretest session, children were presented either 12 large-set problems or 12 small-set problems. One-half of children received the set sizes in the order large-small-large-small; the other half received them in the order small-large-small-large. Before each session, the child was brought individually to a small room within the preschool or day care center and asked to sit across a small table from the experimenter. The experimenter then gave the following instructions:

"Hello. Today we are going to play a game with these buttons. First I will start with two rows of buttons, with the same number of buttons in each row. Then I'll do something to one of them that may or may not change the number of buttons. Your job is to tell me whether that row then has the same or a different number of buttons.

Now there are three things I can do. I can add more buttons and make the line a different length; I can take away buttons and make the line a different length; or I can change the length of the line and not add or take away any buttons.

Following this, the experimenter presented the 12 trials of the pretest session. On these trials, unlike in the subsequent training sessions, no feedback was given. Instead, children were told noncontingently at frequent intervals that they were doing "a good job" because they were "trying hard." The pretest procedure did not produce learning—children were correct on 48% of problems in the first pretest session and 45% in the second.

**Training Sessions**

**Subjects**

For children to be included in the study, their pretest performance had to meet two criteria: no more than 20% of large-set, length-inconsistent trials on which both correct answers and transformational explanations were given, and 0% trials on the standard Piagetian conservation problem (the large set, conflict-equal problems) on which both correct answers and transformational explanation were given. Of the 97 children who were pretested, 45 met these criteria. These 24 boys and 21 girls ranged in age from 54 to 73 months ($M = 5.17$ years, $SD = .46$ years).

**Problems**

In each of the four training sessions, all children were presented 12 problems, 2 each from the same 6 problem-types as on the pretest. The particular items were identical to those on the pretest. In the course of the experiment, each child encountered each large-set item three times: once during the pretest, once during the first two sessions of the training period, and once during the last two sessions of the training period. This allowed examination of changes within-subject within-item. The order of presentation of items was randomly derived for each session, subject to the constraint that each of the six problem-types appear once in the first six items and once in the second six items of each session.

1 The reason that only large-set items were given in the training sessions was that pretest performance on small-set problems was already quite accurate (83% correct). This was in line with previous findings (e.g., Gelman, 1972; Siegler, 1981; Winer, 1974) that young children often succeed on small-set but not large-set problems.
Procedure

A randomly chosen 15 children were assigned to each of the three experimental groups: feedback only, feedback plus explain-own-reasoning, and feedback plus explain-experimenter's-reasoning. Children in all groups received the same problems in the same order and received feedback on their answers. They were told either "That's right, that row does have more." "That's right, they do have the same number of buttons." "No, actually this row has more buttons," or "No, actually they have the same number of buttons." The four training sessions were administered over a period of no more than 2 weeks.

The procedures presented to children in the three groups differed only in a quite subtle way: the type of explanation that was requested on each trial. Children in the feedback-only group were not asked to provide any explanation; they saw the same problems as on the pretest, were asked the same question concerning which, if either, row now had a greater number of objects; advanced the same type of judgment; and were provided feedback concerning which, if either, row actually had more. Children in the feedback-plus-explain-own-reasoning group received the identical procedure except that they were asked after their judgment (but before the feedback) to explain their own reasoning. In particular, they were asked "How did you know that?" Children in the feedback-plus-explain-experimenter's-reasoning group saw the same transformations and received the same feedback; however, after the experimenter indicated the right answer, she asked them to explain her reasoning. In particular, she asked "How do you think I knew that?"

RESULTS

The data were analyzed at two levels: the group level and the individual level. The group-level analyses provided an overview of the effects of the experimental conditions on changes in performance. The individual-level analyses provided detailed depictions of how changes occurred and allowed analysis of children's varying reactions to a single experimental treatment.

In most analyses, the six types of problems were divided into two sets: the three types of problems on which reliance on relative length produced the correct answer (length-consistent problems), and the three types of problems on which reliance on relative length produced incorrect answers (length-inconsistent problems). The reason for this division was that both initial performance and patterns of changes over sessions differed greatly depending on whether the length cue led to a correct answer on the particular problem. Neuman-Keuls post hoc comparisons were used to identify the sources of main effect and interactions that appeared within the analyses of variance. All results reported are significant beyond the .05 level.

Group Level Analyses

Percent correct judgments. To determine whether children benefited from requests to explain the experimenter's reasoning, and whether the benefits varied with the type of problem, a 3 (Group: feedback only, feedback-plus-explain-own-reasoning, or feedback-plus-explain-experi-
menter’s reasoning) × 2 (Problem-type: length-consistent or length-inconsistent) × 5 (Session: pretest, training session 1, 2, 3, or 4) ANOVA of percent correct judgments was conducted. As in all of the subsequent analyses, group was a between-subjects factor and session and problem-type were within-subjects factors. The analysis yielded significant main effects for all three variables: group, F(2,42) = 7.17, problem type, F(1,42) = 106.25, and session, F(4,168) = 12.09. As predicted, children who were asked to explain the experimenter’s reasoning answered correctly more often than did children who explained their own reasoning or who were not asked for any explanation: 62% vs 48% and 49% correct. Children also answered correctly more often on length-consistent problems than on length-inconsistent ones (72% vs 34%), and more often during the last three training sessions than during the pretest or the first training session (47% and 50% on the pretest and first training session vs 55%, 58% and 61% in the last three training sessions).

Significant interactions between group and session, F(8,168) = 2.78, problem-type and session, F(4,168) = 32.49, and group, problem-type, and session F(8,168) = 2.60, were also present. Inspection of the means indicated that the effects of experimental condition and sessions were entirely different on the length-inconsistent problems, where the length cue pointed to the wrong answer, than on the length-consistent problems, where it pointed to the right answer. Therefore, separate ANOVAs were conducted for length-consistent and length-inconsistent problems to explicate the sources of the interactions.

First consider the analysis of performance on the length-inconsistent problems. Significant effects were present for group, F(2,42) = 7.21, session, F(4,168) = 35.03, and the interaction between group and session, F(8,168) = 3.13. Children who were asked to explain the experimenter’s reasoning were correct significantly more often than those asked to explain their own reasoning or just given feedback (Fig. 1). Percent correct was higher in all four of the training sessions than it had been on the pretest, and was also higher in the last three training sessions than in the first one. Finally, and most important, the group by session interaction on these length-inconsistent problems reflected the fact that the three groups did not differ on the pretest, but that children who were asked to explain the experimenter’s reasoning advanced a significantly higher percentage of correct answers than children in the other two groups in each of the four training sessions. Performance improved from 17% correct on the pretest to 72% correct in the last training session for the group asked to explain the experimenter’s reasoning, from 13 to 31% correct in the group asked to explain their own reasoning, and from 9 to 41% correct in the group that was not asked to provide an explanation. Thus, on the length-inconsistent problems, percent correct steadily improved over sessions,
with the improvement more marked in the group asked to explain the experimenter's reasoning than in the other two groups.

The pattern was quite different on the length-consistent problems (Fig. 2). There was no significant main effect for experimental group. There was a significant main effect for session, $F(4,168) = 7.02$, but it reflected decreasing accuracy over sessions. Percent correct decreased from the pretest level of 80 to 66, 68, 75, and 65% in the four training sessions; the decrease was significant from the pretest to the first, second, and fourth training sessions. The session by group interaction was also significant, $F(8,168) = 2.20$, reflecting a greater overall decline in percent correct from the pretest to the final training session in the group asked to explain the experimenter's reasoning (84 to 67% correct) and in the group asked to explain their own reasoning (82 to 61%) than in the group not asked to advance any explanation (75 to 67%).

**Percent judgments consistent with length.** The fact that percent correct on the length-consistent problems declined over the course of training hardly fit the prototype of a training effect. One interpretation was that number conservation involves two key realizations, and that many children only came to one of them. Specifically, the two key realizations for understanding number conservation are (1) relative length does not predict relative number, and (2) the type of numerically relevant transformation (or counting) does predict it. A number of children may have learned
the first but not the second of these lessons. This would entail fewer judgments consistent with relative length, but not necessarily a greater percentage of correct judgments. Children who came to the first but not the second realization would have become more accurate over sessions on the length-inconsistent problems, where relying on length initially produced 0% correct answers, but less accurate over sessions on the length-consistent problems, where relying on length produced 100% correct answers.

Examining the percentage of children's errors that were consistent with the length cue allowed a rigorous test of this explanation. On length-consistent problems, all errors were by definition inconsistent with length. On length-inconsistent problems, however, an error could either be consistent with length or inconsistent with it. If a row was spread out and nothing added or subtracted, for example, a child could err either by saying that the longer row had more objects or that the shorter one did. If learning about the unreliability of length was separate from learning how to generate correct answers, the effect should be evident in the specific errors that were made, as well as in the percentage of correct answers. In particular, of the errors that children made, a higher percentage should have been to the choice not predicted by the length cue.

Results of a 3 (Groups) × 5 (Sessions) ANOVA on the percentage of erroneous judgments that were consistent with length supported this pre-
diction. It yielded a single significant main effect for sessions, $F(4,168) = 24.62$. Percentage of erroneous judgments that were consistent with length decreased from 77% on the pretest to 48, 42, 36, and 31% in the four training sessions. The difference between the percentage of length-consistent errors on the pretest and on each of the four training sessions was significant, as was the difference between the percentages in the first and last training sessions.

The large decrement in the percentage of errors consistent with length, the fact that most of this decrement occurred earlier in training than the increase in correct answers, and the fact that the decrement was of equivalent magnitude for all three groups (as evidenced by the lack of a group by session interaction) rather than being disproportionately present in the group asked to explain the experimenter's reasoning were all consistent with the interpretation that conservation learning involved two separate realizations—what not to rely on (the relative lengths of the rows) and what to rely on (the type of transformation or counting). The data also indicated that some children learned that relative length did not predict relative quantity without learning what did predict it. These children continued making errors, but by the final training session, far fewer of their errors were consistent with length.

**Percent correct for each transformation.** The design of this study allowed analysis of whether the pattern of acquisition of the traditional Piagetian conservation problem, involving the null transformation, differed from that of other length-inconsistent problems that involved addition or subtraction transformations. A $3 \times 5 \times 2$ ANOVA on percent correct answers was conducted to examine the influence of the type of transformation.$^2$ Because the effects for group and session were the same as in the analysis of performance on length-inconsistent problems reported earlier, only effects involving the type of transformation are reported here.

The analysis indicated significant effects for the type of transformation, $F(1,42) = 43.78$, and for the interaction between transformation and session, $F(4,168) = 4.20$. Children generally answered correctly more often on problems involving addition and subtraction transformations than on ones involving the null transformation (45% vs 21%). The interaction appeared due to the differences between the two types of transformations increasing during the training period beyond those that were present initially. On the pretest, percent correct judgments was already somewhat higher for the addition and subtraction problems than for the null transformation ones (16% vs 6% correct). During the first training session, the

---

$^2$ Preliminary analyses revealed similar performance on addition and subtraction transformations; therefore performance on the two types of transformations was collapsed.
difference grew considerably (42% vs 8% correct), before leveling off in the second and third training sessions and finally decreasing in the fourth sessions (52, 60, and 54% correct on the addition/subtraction problems versus 28, 28, and 37% on the null transformation problems). The finding that pretest performance on addition/subtraction problems was more accurate than on the traditional Piagetian problem replicated previous findings (Gelman, 1972; Siegler, 1981). The finding that learning of addition/subtraction problems was also faster extended the earlier findings.

Explanations

Children in two of the three groups—the explain-own-reasoning and the explain-experimenter’s-reasoning groups—advanced explanations as well as judgments on each trial during the pretest and training sessions. The explanations fell into five categories (Table 2). Separate 2 (Group) × 5 (Session) × 2 (Problem type) ANOVAs were conducted for the three most common types of explanations, which for simplicity of reference are referred to as “transformational,” “length,” and “don’t know” explanations. These analyses were conducted on the single main explanation on each trial, as explained under Method.

Transformational explanations. Number of transformational explanations was significantly influenced by group, \( F(1,28) = 12.02 \), session, \( F(4,112) = 8.30 \), and problem type, \( F(1,28) = 10.57 \). Children advanced a greater number of transformational explanations to account for the experimenter’s reasoning than to account for their own (29% vs 5%) and a greater number on the length-inconsistent than on the length-consistent problems (21% vs 13%). They also advanced a greater number of transformational explanations in later sessions than in earlier ones (4, 13, 19, 26).

<table>
<thead>
<tr>
<th>Explanation</th>
<th>Example</th>
<th>% Use on pretest</th>
<th>% Use in training sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transformational</td>
<td>You added one to that row, so it has more.</td>
<td>4</td>
<td>19</td>
</tr>
<tr>
<td>Length</td>
<td>That row is longer.</td>
<td>52</td>
<td>24</td>
</tr>
<tr>
<td>Counting</td>
<td>I counted, and that row has 7 and that one has 6.</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Back and forth</td>
<td>It doesn’t matter, because you put them back where they started.</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Don’t know</td>
<td>I don’t know why they’re the same.</td>
<td>36</td>
<td>49</td>
</tr>
</tbody>
</table>

TABLE 2
Most Frequent Explanations
21, and 25% on the pretest and the four training sessions respectively), with the number on the pretest differing significantly from the numbers in the second, third, and fourth training sessions, and the number in the first training session differing significantly from that in the fourth.

Significant interactions were also present between group and session, $F(4,112) = 5.58$, and among group, session, and problem type, $F(4,112) = 4.08$. As illustrated in Fig. 3, the basis of both of these interactions appeared to be that percentage of transformational explanations was at a constant, low level in the group in which children were asked to explain their own reasoning, but was influenced by both the session and the type of problem in the group in which children were asked to explain the experimenter’s reasoning.

To test this interpretation, 2 (Problem type) X 5 (Session) ANOVAs on the frequency of transformational explanations were performed separately for children in the two groups. No significant effects were present in the analysis of performance of children who were asked to explain their own reasoning. In contrast, the parallel analysis of the frequency of transformational explanations among children asked to explain the experimenter’s reasoning indicated significant main effects for both session, $F(4,56) = 11.41$, and problem type, $F(1,14) = 8.42$, as well as for the interaction between session and problem type, $F(4,56) = 3.57$. The interaction was due to these children advancing similar numbers of transformational explanations on length-consistent and length-inconsistent
problems on the pretest, but advancing more transformational explanations on length-inconsistent problems in each of the four training sessions.

The differences in frequency of transformational explanations for the two groups in which children advanced explanations could not be accounted for by the children who explained the experimenter's reasoning more frequently explaining correct rather than incorrect answers. When the analysis was limited to trials on which children had advanced correct answers, children explaining their own reasoning still only advanced transformational explanations on 8% of trials, far fewer than the 29% of trials for those explaining the experimenter's reasoning. Again, the groups' percentages of transformational explanations did not differ significantly on the pretest, but did differ significantly in all four training sessions. Nor was the difference in frequency of transformational explanations limited to one type of problem. Children asked to explain the experimenter's reasoning advanced far more transformational explanations both on length-consistent problems (25% vs 4%) and on length-inconsistent ones (42% vs 6%).

Length explanations. The ANOVA on percentage of explanations that cited the relative lengths of the rows yielded a single significant main effect for session, $F(4,112) = 23.25$. Such explanations decreased from 52% of trials on the pretest to 29, 26, 22, and 19% in the four training sessions. The difference between the pretest and each of the training sessions was significant, as was that between the first and last training sessions.

Significant interactions were also present between group and session, $F(4,112) = 4.93$; group and problem type, $F(1,28) = 4.81$; and group, session, and problem type, $F(4,112) = 8.17$. The three-way interaction appeared to be due to the percentage of explanations citing length declining especially greatly on length-inconsistent problems among children asked to explain the experimenter's judgments (from 70% length-based explanations on the pretest to 12, 12, 12, and 7% in the four training sessions). This was not surprising when the children's task was considered. On length-inconsistent problems, the experimenter's judgment never coincided with the length cue. For children to explain the judgment in terms of length, they either needed to misrepresent the experimenter's judgment or to reason that the experimenter had chosen that alternative because it was the shorter row. Citations of length also declined on length-consistent problems and among children who explained their own reasoning, but the decrement occurred progressively over the training sessions and never reached as low a percentage of total explanations.

Don't know explanations. The ANOVA on percentage of trials on which children said they did not know how to explain their own or the experimenter's judgments showed a single significant main effect for ses-
HOW DOES CHANGE OCCUR

The percentage of such explanations increased from 36% on the pretest to 51, 49, 48, and 49% in the four training sessions. The percentages in all four training sessions were higher than that on the pretest.

A significant interaction was also present between problem-type and session, $F(4,112) = 3.79$. It reflected such explanations becoming more frequent over sessions on length-inconsistent problems but not on length-consistent ones. The general increase in such explanations from the pretest to the training sessions seemed attributable to the feedback leading children to realize that they did not know how to solve the problems. That the effect was greater on the length-inconsistent problems may have reflected the correct answers on these problems often not being explainable in terms of the factor that most children initially emphasized, the relative lengths of the rows.

Stability of transformational explanations on identical problems. Once children explained their own or the experimenter's reasoning in terms of the type of transformation that had been performed, did they consistently continue to do so, at least on that same problem? The fact that each problem was presented once on the pretest, once in the first two training sessions, and once in the last two training sessions made it possible to find out.

Previous microgenetic studies have frequently included qualitative descriptions of the instability of use of new approaches. However, they have lacked any generally applicable quantitative index of stability/instability. The present design, in which the same set of 24 problems was presented to each child once in each third of the experiment, made it worthwhile to formulate such a quantitative index. The analyses of stability within child within problem involved a comparison of two conditional probabilities:

1. $P($Transformational explanation on later trial on a problem $| \text{Transformational explanation on earlier trial on that problem})$

2. $P($Transformational explanation on later trial on a problem $| \text{Nontransformational explanation on earlier trial on that problem})$

Such comparisons were used both to examine stability of use of transformational explanations from the pretest to the training sessions, and to compare stability from earlier training sessions to later ones.

Some, but not great, stability of transformational explanations was evident from the pretest to the training sessions. On 43% of problems on which the child advanced a transformational explanation on the pretest, the child also advanced a transformational explanation when the problem was again presented during the training session (with the two presenta-
tions of the problem during the training sessions viewed as separate instances in which the child might or might not again advance a transformational explanation). This was considerably greater than the 18% of problems on which children advanced transformational explanations in the training sessions after not having done so on the pretest. On the other hand, it indicated far from perfectly consistent reliance on transformational reasoning once it was employed, even on the identical problem.

Considerably more stability was evident across training sessions. When a child advanced a transformational explanation on a problem in one of the first two training sessions, that child did so again on 76% of trials when the problem reappeared in one of the last two training sessions. In contrast, when a child did not advance a transformational explanation on a problem presented in the first two training session, that child advanced such an explanation on only 13% of trials when the problem was again presented in the third or fourth training session. Although still not indicative of perfect stability, these data indicated that within the context of the training procedures, early use of a transformational explanation on a problem was quite predictive of later use of such reasoning on that problem. Even in this most favorable situation for demonstrating stability of reasoning, however, in which the child, the problem and the training procedure were all constant, children reverted from using transformational explanations to not using them in 24% of cases.

Number of types of explanations. Children differed not only in the frequency with which they advanced each explanation but also in how many explanations they advanced. Over the five sessions, they advanced a mean of 3.14 of the 5 main explanations.3 Children asked to explain the experimenter’s reasoning advanced more distinct types of explanations than those asked to explain their own reasoning (means of 3.67 and 2.60, t(28) = 2.78). Almost all (13 of 15) children who were asked to explain the experimenter’s reasoning advanced at least three distinct explanations, versus less than half (6 of 15) of those asked to explain their own reasoning.

The fact that each of the 24 problems appeared once on the pretest, once in the first two training sessions, and once in the last two training sessions allowed examination of changes in number of types of explanations over sessions. The analysis revealed no difference in the number of types of explanations used; children used a mean of 2.43 of the 5 main explanations on the pretest, 2.47 in the first two sessions, and 2.30 in the last two sessions (the means are lower than the overall mean be-

3 When all explanations were considered, the mean was 3.45 different explanations per child. In the remainder of the analyses, consideration is limited to the five main explanations, which together accounted for more than 90% of total explanations.
cause children used some explanations in some sessions and not others). The findings indicate that children possessed multiple ways of thinking about number conservation before they received any training (pretest performance), and continued thinking about it in multiple ways throughout the training sessions.

Relation between accuracy and use of transformational explanations.
The analyses presented above suggested that children learned conservation through coming to understand that the type of transformation is what determines whether the number of objects increases, decreases, or remains unchanged. Analyses of changes over sessions in correlations between percent correct and percent use of transformational reasoning strongly supported this interpretation. Among the 30 children who were asked to provide explanations, the correlations between percent correct and percent use of transformational explanations increased from \( r = .12 \) on the pretest to \( r = .28, .62, .64, \) and \( .78 \) for the four training sessions. The low levels on the pretest and the first training session reflected the restricted range of the number of transformational explanations that children advanced. Children were selected for participation on the basis of advancing few or no transformational explanations on the pretest, and the percentage of transformational explanations increased gradually over sessions. Thus, the correlations in the later sessions were the best indicators of the strength of the relation.

The strong correlations in the later sessions were not attributable to children who were asked to explain the experimenter’s reasoning generally learning better, and therefore both answering correctly more often and advancing more transformational explanations. The pattern was present within each training group. Among children asked to explain the experimenter’s reasoning, the correlations for the pretest and four training sessions were \( r = .01, .03, .48, .55, \) and \( .85 \), respectively. Among those who explained their own reasoning, the corresponding correlations were \( r = .13, .39, .35, .45, \) and \( .53 \). Thus, the relation between accuracy and use of transformational explanations steadily increased over the training sessions, reaching high levels by the end of the experiment.\(^4\)

Individual Patterns of Change

The microgenetic design used in this study allowed detailed analysis of changes in individual children’s performance, as well as of differences in children’s reactions to a single experimental condition. Since by far the

\(^4\) The somewhat higher correlations in the group asked to explain the experimenter’s reasoning were likely due to the more continuous distribution of percentages of use of transformational explanations in that group.
largest changes occurred in the group asked to explain the experimenter's reasoning, the analyses of individual patterns of change focused on children in this group.

Preliminary examination of the change patterns suggested that for 12 of the 15 children, the changes over sessions in both accuracy and explanations fit one of two patterns. These patterns were defined in terms of which explanation (among the five main ones) showed the greatest increase from the pretest to the fourth training session, and which the greatest decrease. Below, these two relatively common patterns are described at both the subgroup and individual levels; the focus then shifts to the idiosyncratic patterns shown by the remaining three children in the experimental condition.

Transformational explanations increased length explanations decrease. The most common pattern among children who were asked to explain the experimenter's reasoning involved increasing use of explanations emphasizing the type of transformation and decreasing use of explanations emphasizing the lengths of the rows. Slightly more than half (53%) of the children in this group showed this pattern of change. Their performance reflected both of the realizations described above: that relative length does not predict relative number, and that the type of transformation does.

As illustrated in Fig. 4, six of the eight children showed increases in transformational explanations in the first training session. The increase continued for most of them into the second training session, with some showing further increases in the third and fourth sessions. Of the other two children who showed substantial increases in transformational explanations, one began the increase in the second training session and one in the third. The increase continued for both of these children through the final session.

The decrease in length explanations was even larger than the increase in transformational ones. The decrements were invariably present, and usually substantial, even in the first training session. For seven of the eight children, a decrease of more than 30% in absolute terms was present between the pretest and the first training session. The mean absolute decrement in length explanations between pretest and first training session (45%) was even larger than the mean absolute increment in transformational explanations during the same period (27%). The pattern was also evident at the individual level; seven of the eight children showed larger decrements in reliance on length explanations than increments in transformational ones between the pretest and first training session. The decrements in length explanations did not stop there. By the end of the training sessions, only two of the eight children advanced length explanations more than occasionally.
One other noteworthy feature of these eight children’s explanations was an increment from the pretest to the first training session in percentage of trials on which they said “I don’t know” (from 12 to 36%). Frequency of this explanation increased for all eight children. The first training session represented the highpoint of this explanation for most of these children; by the last training session, most of them did not use it at all. Thus, the experimenter’s reasoning was something of a mystery to these children initially, but a mystery that they progressively figured out.

Even among these children, who learned the most of any of the participants in the experiment, changes toward adoption of transformational explanations occurred gradually rather than suddenly. This is evident in the graphs of individual children’s explanations (Fig. 4). None of the children progressed in a single session from rarely advancing transformational explanations to consistently doing so. Rather, the increases in use of transformational explanations generally continued over several sessions. Half of the children advanced both explanations based on length and explanations based on the type of transformation in at least three of the five sessions. Even when children advanced transformational explanations quite often (for example, on half of trials in a session), they did not generally advance them consistently in the next session. The data converge with the earlier-described finding that use of a transformational explanation by a given child on a given problem on an earlier trial does not imply use of a transformational explanation by the same child on that problem on a later trial.

These children’s increasing use of transformational explanations to account for the experimenter’s reasoning was accompanied by impressive increases in the accuracy of their own judgments (the numbers in parentheses in Fig. 4 represent percent correct for that child in pretest and Training Sessions 1, 2, 3, and 4). Between the pretest and the final training session, their percent correct increased from 49 to 86%. The pattern was highly consistent across individuals. Absolute percentage correct increased at least 30% for seven of the eight children (the eighth child showed a similar increase from the pretest to the third training session, but for unknown reasons did less well in her last one). Thus, children in this subgroup applied the reasoning they found useful for explaining the experimenter’s judgments to generating their own.

Don’t know explanations increase; length explanations decrease. Children in the second subgroup also decreased their reliance on length as an explanation of the experimenter’s judgments. Rather than increasing their reliance on transformational explanations, however, this group of children most often stated that they did not know why the experimenter judged as she did (Fig. 5). Thus, these children learned that length did not
predict relative quantity, but did not learn that the type of transformation did predict it.

Interestingly, three of these four children had advanced at least one transformational explanation of their own judgments on the pretest, but none of the four ever advanced such explanations of the experimenter's judgments in the subsequent training sessions. The frequencies of transformational explanations of their own pretest judgments were low, but no lower (and sometimes higher) than those of other children who did sub-
substantially increase their use of transformational explanations during the training sessions.

Consistent with their inability to explain the experimenter's conservation reasoning, the accuracy of these children's own judgments did not increase over trials (Fig. 5). They answered 43% of problems correctly on the pretest and 40% correctly in the last training session. The difference between this pattern and that of children who increasingly advanced transformational explanations of the experimenter's behavior provided converging evidence for the interpretation that learning to explain the experimenter's reasoning in terms of transformations was critical to the
children learning to make their own judgments correctly. The performance of children in this group also gave particularly clear evidence of the separation between learning that length did not predict relative number and learning that the type of transformation did.

Idiosyncratic patterns. The remaining three children generated performance that differed both from that of children in the first two subgroups and from that of the other two children who showed idiosyncratic patterns (Fig. 6). One child answered correctly approximately 50% of problems,
both length-consistent and length-inconsistent, in each of the four training sessions. He increasingly advanced transformational explanations of the experimenter's reasoning on problems where his own judgments were correct (6 of 7 in the final session), but always said he did not know the reason for the experimenter's judgments when his own judgment was wrong. A second child was somewhat more accurate during the four training sessions, but was unable to explain either her own or the experimenter's judgments at any point in the experiment. The third child was very accurate, answering correctly 92, 92, 100, and 75% of training ses-
sion trials. He was the only child to frequently explain the experimenter’s judgments by saying that she had counted the objects in the two rows. His next most frequent explanation was to cite the type of transformation; he also cited length fairly often. These idiosyncratic patterns further illustrate the variability of children’s learning in response to a single experimental procedure.

Predictors of learning. Why did some children benefit so much from being asked to explain the experimenter’s judgments and others little if at all? To address this problem, the predictive relation between a number of aspects of each child’s pretest performance and that child’s subsequent performance in the training sessions was examined. In particular, stepwise linear regression analyses were conducted to identify measures of pretest performance that predicted total percent correct over the four training sessions. The measures of pretest performance that were used as predictors were number correct, number of transformational explanations, number of different explanations used, whether a child ever used two or more explanations on a single trial, and the child’s age (in months).

Three factors significantly predicted each child’s percent correct over the four training sessions. The first factor to enter the equation was number of different explanations used on the pretest, which accounted for 34% of the variance in percent correct in the subsequent training sessions. The greater the number of distinct types of explanations children used on the pretest, the higher their percent correct in the training sessions. The second factor was whether the child ever advanced two or more explanations on a single pretest trial. Children who sometimes advanced multiple explanations on a given pretest trial learned more effectively during the subsequent training period. The two factors together accounted for 55% of the variance in percent correct in the training sessions. The third predictor was age; older children learned better. Together, the three factors accounted for 65% of variance in percent correct over the four training sessions.

It should be noted that the positive relation between number of different explanations used on the pretest and learning during the training period could not be reduced to some children already advancing transformational as well as other explanations on the pretest and therefore being better able to learn. If this explanation were valid, the number of transformational explanations on the pretest should have predicted subsequent learning, but it did not. Nor could the relation be reduced to children who generated more correct answers during training having already been more accurate on the pretest. Again, had this been the case, percent correct on the pretest should have predicted subsequent learning, but it did not.

Examination of which children generated multiple explanations within a single pretest trial revealed that the measure was related to the individ-
ual difference classifications described earlier. Among children in the increased-transformational-explanations/decreased-length-explanations subgroup, six of eight (75%) advanced two types of explanations on at least one pretest trial. Only three of the other seven children (43%) did so. The difference became more striking when performance during the training sessions was included. All eight children who during training substantially increased their use of transformational explanations advanced multiple explanations on a single trial at some point during the study, versus three of seven of the other children ($p < .03$, Fisher Exact Test). Over the course of the experiment, these eight children also advanced multiple explanations on considerably more trials than did the other seven children (means of 7.50 vs 1.29 trials, $t(13) = 2.42$). In sum, generation of varied explanations, both between and within trials, was positively related to learning.

DISCUSSION

The microgenetic method used in this study yielded the type of in-depth information needed to increase understanding of how change occurs. Consideration of the findings and their implications can be organized around the five dimensions of change described earlier: the path, rate, breadth, variability, and sources of change.

Path of Change

Change in this experiment did not fit the usual model of a developmental sequence, in which children progress from Level 1 to Level 2 to Level 3 reasoning. Instead, individual children and the group as a whole were both best described as using a variety of explanations at all points in the experiment. Over the sessions, children on average used more than three types of explanations. The variety of explanations was also evident within each session, with the average number advanced by each child remaining almost constant over the pretest and four training sessions.

Individual children’s paths of change over sessions, even within a single experimental condition, also showed great variability. For example, a number of children who were asked to explain the experimenter’s reasoning progressed from most often explaining it in terms of length to vacillating among several explanations to explaining it most often in terms of the type of transformation. Other children, however, moved from most often advancing length-based explanations to not knowing how to explain what they observed. Yet other children moved from most often citing length to a variety of idiosyncratic patterns. The overall pattern described in previous studies—movement from reliance on length to reliance on the
type of transformation—was evident here as well, but the results revealedar more variability in the path of cognitive change, both in terms of
performance within any one session and in terms of change over sessions,
than recognized in previous descriptions.

These findings suggest that the traditional construct of a developmental
sequence is oversimplified, at least as it applies to number conservation.
Rather than progressing through a series of discrete knowledge states, in
which they consistently think of a phenomenon in a single way at a given
time, children know and use several types of reasoning before as well as
during transitional periods. Such use of diverse strategies for prolonged
periods has been observed previously in other domains, such as arithmetic,
time telling, and serial recall (Siegler, 1994). However, its appearance
on number conservation is especially telling. Number conservation is a
classic Piagetian task; descriptions of children’s thinking on it and other
Piagetian tasks provided much of the empirical basis for the standard
concept of a developmental sequence. It also is a logical task, in which
one type of reasoning is inherently superior to another. Yet on it as on
other tasks, children use a variety of types of reasoning, of varying de-
greses of sophistication, and use them both before and during training.

Rate of Change

Changes in understanding of number conservation came quite gradu-
ally. Even the subgroup of children who eventually most completely relied
on the type of transformation required several training sessions be-
fore they consistently did so. The moderate rate of change was evident
despite the logical superiority of thinking of conservation in terms of
transformations, and despite the fact that focusing on the type of trans-
formation, rather than length, produced much greater accuracy.

Even within individual items, change could not be characterized as a
sudden shift from not relying on transformational reasoning to relying on
it. Children who advanced a transformational explanation on a pretest
item did so on less than half of subsequent trials when that item was
presented in the training sessions. Even within the four training sessions,
children who cited the type of transformation on an earlier trial advanced
a different type of explanation on one-quarter of instances when the same
item was presented later. It is important to realize that the ’’same prob-
lem’’ here refers to literal identity: same number of objects, same length
of rows, same transformation, same instructions, same question asked by
the same experimenter in the same room. Even here, children who un-
derstood the problem well enough to state a transformational explanation
fairly frequently advanced a less advanced explanation a few days later.

This relatively slow progress was not attributable to the measure being
HOW DOES CHANGE OCCUR

unreliable or to children just not wanting to repeat the same explanation. A parallel analysis of conditional probability of advancing a correct answer on a problem in a later training session, given that the child advanced such an answer in an earlier training session, yielded almost identical results. There, on 22% of occasions in which the child had answered a problem correctly in the earlier training session, the child answered incorrectly in the later session (versus 24% for explanations). This variability, both in children's judgments and in their explanations, was part of the reason for the relatively slow rate of change.

These findings do not mean that the rate of adoption of advanced ways of thinking is the same in all domains; the rate of increase of transformational explanations in number conservation was considerably greater than the rates of adoption of new strategies previously observed in arithmetic, scientific reasoning, and other domains (Kuhn & Phelps, 1982; Kuhn, Schauble, & Garcia-Mila, 1992; Metz, 1985; Schauble, 1990; Siegler & Jenkins, 1989). Even in number conservation, however, with many factors favoring rapid adoption of transformational reasoning, older ways of thinking continued to coexist with newer ways for several sessions. The findings suggested that the gradual adoption of new approaches, evident in previous microgenetic studies, will to varying degrees characterize change patterns on a wide variety of problems, including ones where the new approach is logically superior and brings with it large increases in accuracy.

The present method of repeatedly presenting the same set of problems, and of using conditional probabilities to measure stability of use of the new approach within individual problems, seems likely to be useful for microgenetic studies in general. It provides a quantitative index that can be used to compare the rate of change across experiments. It also provides a means for distinguishing between two interpretations for the frequently-observed slow adoption of new approaches. In previous microgenetic experiments, conclusions regarding the gradual nature of change have been based on simple counts of the frequency of use of new and old strategies following discovery of the new approach. This made it impossible to determine whether low initial rates of use were due to inconsistent application of the new strategy to problems where it had previously been used or to the strategy initially being limited to problems with particular characteristics (but being consistently applied on them). In the present terminology, it did not allow distinction between the rate and the breadth of change. Repeatedly presenting the same items allows separate assessment of the two. Findings in the present experiment indicated that at least part of the gradual rate of the increase in transformational reasoning was due to the new approach not being applied consistently, even to individual problems.
Breadth of Change

This study's microgenetic design allowed investigation of several questions concerning how initially uncommon forms of reasoning are extended to new problems within a domain. One question was whether children would use transformational reasoning equally often on length-consistent and length-inconsistent problems. On the pretest, frequency of such reasoning was equal on the two types of problems. However, during the training sessions, children generated transformational explanations more often on length-inconsistent problems (e.g., more often when the longer row did not have more objects).

This difference was evident even among the subgroup of children who learned the most—those who were asked to explain the experimenter's reasoning and who substantially increased their reliance on transformations during the training sessions. By the final training session, these children cited the type of transformation in 90% of their explanations on length-inconsistent problems, but on only 46% of explanation on length-consistent items. Thus, even the best learners in the experiment applied the logically more-advanced reasoning much more often on problems where it was necessary to account for the particular judgment of the experimenter than on problems where less advanced explanations were also consistent with the experimenter's judgment on that item.

Similar unevenness of application was evident when examining application of transformational explanations to the three types of transformations: addition, subtraction, and the null transformation. Transformational explanations were slightly more frequent on the pretest on addition and subtraction problems than on problems that involved the null transformation. The difference steadily increased over the course of the training sessions. The increase could not be explained by greater exposure to addition and subtraction transformations, because within the experiment, the amount of exposure was equal. Instead, it appeared that the effects of addition and subtraction transformations are easier for children to learn.

Number conservation is not a domain in which there is a very broad variety of possible problems. From an adult perspective, knowing the effect of any one type of transformation implies the effects of the other two. The fact that knowledge spreads unevenly across such a narrow range of problems suggests that young children's applications of new forms of reasoning may not only be gradual, they may also generally be uneven across relatively small variations in the type of problem.

Variability of Change

As noted above, great variability in change patterns was evident at all levels of analysis. Perhaps the most striking type of variability was that in
the explanations advanced by a single child on a single problem. At first impression, such variability seems puzzling. Why should a child who understands advanced forms of reasoning sufficiently well to articulate them then go on to advance less advanced explanations on later presentations of the identical problem?

In thinking about this question, it may be useful to distinguish between the functions of two types of variability: variability between trials and variability within trials. First consider the functions of variability between trials. It often is difficult to foresee the effects of new approaches without actually using them; trying alternatives allows exploration of the task environment. Thus, on tasks such as class inclusion, map drawing, and use of causal verbs, children have been found to continue generating new approaches even after they have devised entirely adequate ones (Bowerman, 1982; Karmiloff-Smith, 1992; Markman, 1978). In some, though not all, of these cases, the alternatives proved even more useful than previously used approaches and eventually became the most frequent approach. Thus, applying varied approaches on different occasions contributes to learning by providing empirical evidence concerning which approaches work best.

Variability within a single trial is also associated with cognitive change, but its relation may be correlational rather than causal. Unusually great within-trial variability often immediately precedes discovery of new strategies (Caron & Caron-Pargue, 1976; Siegler & Jenkins, 1989). Hesitations, vague speech, and speech with unclear referents and internal contradictions are manifestations of this type of variability. Variability within a single trial has also been found predictive of learning; children who communicate different information in speech and in gesture on individual pretest trials have been found particularly likely to benefit from subsequent training (Alibali & Goldin-Meadow, 1993; Church & Goldin-Meadow, 1986; Goldin-Meadow, Alibali, & Church, 1993). A potential source of the relation of this within-trial variability to learning, the source emphasized by Goldin-Meadow and her colleagues, is that such within-trial variability reflects cognitive conflict, which makes children more amenable to new ways of thinking.

The present findings added to this growing body of evidence concerning the relation of both between-trial and within-trial variability to subsequent learning. The two best predictors of which children would benefit most from being asked to explain the experimenter’s reasoning were the number of different explanations the child used on the pretest and whether the child ever advanced two or more explanations on a single trial on the pretest. Together, these two predictors accounted for more than half of the variance in subsequent learning. The children who learned best continued to exhibit greater variability, both within and between trials, during
the training sessions. Thus, greater variability in children's explanations was associated with greater learning, both predictively and concurrently.

Sources of Change

Being asked to explain the correct reasoning of an adult was more effective in promoting learning than was just receiving feedback or receiving feedback and being asked to explain one's own reasoning. This finding extended in at least four ways previously-published results regarding self-explanations. It showed that (1) explaining other people's reasoning is causally related to better learning, not just correlated with it; (2) generating such explanations is helpful in acquiring a classic cognitive developmental concept, number conservation, as well as formal college-level computer programming and physics concepts; (3) the effectiveness of the procedure comes from explaining another person's more advanced reasoning, rather than from the act of explanation per se; and (4) the benefits of explaining other people's reasoning extend to 5-year-olds, as well as to the much older individuals who had been studied previously.5

The fact that young children can benefit from requests to explain another person's reasoning may prove to be especially important. In the everyday environment, children's effort to understand other people's reasoning may contribute to many aspects of cognitive development. Such efforts are not automatic, however, certainly not among 5-year-olds; if they were, the request to explain the experimenter's reasoning would have been redundant with the children's spontaneous activity, and would not have aided learning. To the extent that young children are less likely than older ones to generate such explanations spontaneously, but can derive similar benefits when they do so, it may prove especially worthwhile to encourage them to try to explain the behavior that they encounter.

The type of cognitive activity encouraged by the invitation to explain the experimenter's reasoning seems prototypic of what Tomasello, Kruger, and Ratner (1993) labeled "cultural learning." In their words,

In cultural learning, learners do not just direct their attention to the location of another individual's activity; rather, they actually attempt to see a situation the way the other sees it—from inside the other's perspective as it were . . . It is learning in which the learner is attempting to learn not from another but through another. (p. 496)

As Tomasello et al. note, a wide range of theorists have proposed that such cultural learning is central to cognitive development. Regardless of

5 Since this was written, another article (Chi, deLeeuw, Chiu, & Levancher, 1994), also demonstrating the causal impact of instructions to explain reasoning (in their case, the reasoning underlying statements in a textbook) has been accepted for publication.
whether the particular proposal emphasizes taking the other person’s role (Mead, 1934), taking the other person’s perspective (Piaget, 1932), simulating the other person’s thinking (Harris, 1991), reading the other person’s mind (Whiten, 1991), or appropriating the other person’s reasoning (Rogoff, 1990), the ideas converge on the hypothesis that much of children’s learning comes from trying to explain other people’s reasoning.

The present results bolster the empirical basis for this hypothesis. Most previous evidence was based on logical arguments for why such activity would be effective, on the observation that better learners engage in such activity more often than less good learners, and on the observation that older children engage in it more often than younger ones. The present results, however, demonstrated the causal role of generating such explanations; children who were randomly chosen for encouragement to explain another person’s reasoning learned more.

What factors influence the effectiveness of such encouragement to explain another person’s reasoning. One factor that probably is critical is the correctness of the judgment being explained. Explaining the reasoning underlying a correct judgment seems likely to promote better learning than explaining the reasoning underlying an incorrect one. A second class of factors that seem likely to influence the effectiveness of such explanatory activities is the relation between children’s existing reasoning and the reasoning they are trying to explain. Learning may be most effective if the reasoning that children are trying to explain either overlaps the best of their own reasoning or is somewhat, but not greatly, more advanced than that level. This interpretation could be tested by assessing in detail younger children’s preexisting knowledge of transformations and other quantitative skills, presenting them the current experimental manipulation, and determining which factors predicted their learning. A third factor that may influence benefits of attempts to explain another person’s reasoning may be the type of person whose reasoning is being explained. This variable could work in several different ways. Children may think more deeply in trying to explain the reasoning of an authoritative adult than that of individuals whose knowledge they question, on the logic that understanding the adult’s thinking is more likely to be worth the effort. Alternatively, children may think it more likely that they can figure out the reasoning of a peer or younger child than of an adult, and therefore try harder to explain the peer’s or younger child’s reasoning. Yet a third possibility is that the effect of the type of person whose reasoning is being explained may be more on the type of explanation that is advanced than on whether an explanation is advanced. In situations in which multiple explanations are consistent with the observed behavior, children may generate simpler explanations of younger children’s reasoning and more complex ones of the reasoning of older individuals.
In addition to being promising targets for future investigations, these questions place in sharp focus an issue that is central to understanding the source of change in the present study. Were children who were asked to explain the experimenter’s reasoning just trying to explain the correct answer, or were they trying to explain how the experimenter generated that answer? Relevant evidence was provided by differences in the frequency with which children advanced two types of explanations of the experimenter’s judgments: transformational explanations and counting explanations. Although both types of explanations were sufficient to account for all of the experimenter’s answers, children cited the type of transformation 10 times as often. If children were only trying to explain the correct answer, there would have been no obvious reason for this imbalance. Indeed, since counting has often been hypothesized to be a transitional strategy preceding reliance on transformations, children might have been expected to cite it quite often to explain the experimenter’s answers, on the logic that counting was closer to their existing reasoning.

A straightforward reason for children citing the type of transformation far more often than counting was that the topography of the experimenter’s behavior did not make counting a likely explanation of the way in which she had generated her answers. The audible and visible behaviors that typically accompany counting were not present. The lack of counting explanations thus suggested that in line with the question they were asked, children were trying to account for how the experimenter generated her answer, rather than simply trying to explain how the correct answer could have been generated.

Although this evidence suggests that children benefited from trying to explain the experimenter’s reasoning, another possibility that cannot be ruled out is that they benefited from being asked to explain the correct reasoning. Children in the explain-experimenter’s-reasoning condition were always trying to explain correct reasoning; those in the explain-own-reasoning group were not. This difference, rather than the target of the explanation, may have produced the differences in learning. An experiment in which some children are asked to explain the experimenter’s reasoning and others are asked to explain the correct answer (by saying after their answer, “The correct answer is X. See if you can tell me why X is the correct answer” ) is planned to discriminate between these interpretations.

In addition to asking about the conditions under which children’s learning will be enhanced by efforts to understand another person’s reasoning, we also can ask how such activity produces its effects. One possibility is that it works through producing an especially useful type of cognitive conflict. Since all three conditions in the present experiment involved
telling children that many of their answers were incorrect, all would be expected to produce some degree of conflict. However, challenging children to generate reasons that might have led the experimenter to her conclusion may have sharpened the conflict, forcing children to think about the differences between the reasoning underlying her conclusions and the reasoning underlying their own. It may also have led them to consider why they might want to use reasoning like that of the experimenter to generate their own future judgments. The cognitive conflict may have been especially high in the present situation because the other person was an authoritative adult who at least claimed to know the right answers. Such conditions would militate against children assuming their own reasoning was ideal when it conflicted with the other person’s.

Seen from a more general perspective, the way in which encouragement to explain another person’s reasoning produces its benefits may have much in common with the way in which other manipulations designed to promote meaningful learning exercise their effects. Such meaningful learning involves analysis of the semantics of the problem, the goals being pursued, and the way in which various strategies can be applied to the problems so as to meet the goals. Among the constructs used to describe such meaningful learning are deep processing (Craik & Lockhart, 1972; Craik & Tulving, 1975), elaboration (Bower, 1972; Paris & LIndauer, 1977), structure mapping (Gentner, 1989), mental modeling (Halford, 1993), and metacognitive analysis (Kuhn, Amsel, & O’Laughlin, 1988; Schauble, 1990). At the core of all of these ideas is the recognition that people can learn more than usual if they are induced to think about the task more deeply than usual. The present procedure seems to offer a widely applicable means for implementing this insight. In almost any domain, children can be asked to explain how other people came to their conclusions. On a practical educational level, by inducing children to explain other people’s thinking, we may often be able to help them to improve their own.

The Overlapping Wave Model

Two decades ago, Flavell (1971) advanced an argument regarding the depiction of change within stage theories of development that in several ways parallels the present position. He argued that the very brief transitions implied by stage models (Fig. 7A) were unrealistic and suggested that cognitive development should be viewed as including more gradual transitions, such as those depicted in Fig. 7B. This conclusion was definitely a move in the right direction, in that it recognized that children’s thinking is constantly changing. The present data, however, suggest that a more radical departure from the Fig. 7A model is necessary, one that
Fig. 7. Three schematic depictions of change. (A) Traditional stage model, (B) Flavell's (1971) proposed alternative, (C) Overlapping wave model.

recognizes that children possess multiple ways of thinking at any one time, with the frequency of the ways of thinking changing over time. Thinking of cognitive change in terms of overlapping waves, as illustrated in Fig. 7C, provides a potentially-useful alternative depiction.
Within this overlapping wave conception, some ways of thinking are prevalent early-on and then decrease in frequency; others rise from infrequent to frequent use and then fall to infrequent use again; others grow from infrequent to frequent use and remain frequent; still others are used only occasionally even at their peak. The advantage of this conception is that it allows depiction of the diversity of children’s thinking at any one time, the rate of change in the frequency of each way of thinking over time, and the introduction of new ways of thinking into children’s cognitive repertoires. For example, it fits well the patterns of change over sessions in individual children’s performance in the present experiment (Figs. 4–6). It also is sufficiently flexible to be able to describe the types of patterns illustrated in Figs. 7A and 7B, should they characterize a particular acquisition. Finally, it highlights a number of general issues regarding development, in particular issues regarding the origins of the different waves, the processes by which their relative frequency change, and the determinants in the natural environment of the rates and timing of the changes. The present data, along with previous findings, allow us to address each of these issues in the context of number conservation.

Origins of approaches. The basic goal of most research on number conservation, and many other classic cognitive developmental tasks, has been to determine when and how children acquire new, more advanced, forms of reasoning. In the case of number conservation, the new form of reasoning is that lengthening or shortening a row of objects does not change its quantity.

Traditionally, our thinking about this acquisition (and many others) has focused on relatively brief transition periods ranging from when the child “doesn’t have” the competency in question to when the child “has” it. This can be seen in the best-specified current model of acquisition of number conservation, Q-SOAR (Simon & Klahr, 1995; Simon, Newell, & Klahr, 1991). Q-SOAR is a running computer simulation of acquisition of number conservation. It begins with the assumption that before becoming conservers, young children believe that if a row is visibly changed, the number of objects in it will also change. In the absence of contradictory evidence, this leads them to base responses on length. However, when presented small sets of objects, they use measurement operations, notably subitizing and counting, to quantify the set before and after the operation and to note the similarity. Obtaining this contradictory evidence leads them to re-classify the effects of the transformations they have seen, such as spreading and compressing, as ones that do not affect quantity. The learning occurs through the mechanism of chunking. After two learning trials, the type of transformation becomes the basis for responding to all presentations of that transformation.

Q-SOAR has much to recommend it. It is based on a number of plau-
sible assumptions: that on each trial, a variety of dimensions including length, density, and number are encoded; that measurement operations, such as counting and subitizing, are critical to acquisition of number conservation; that the greater ease of applying these measurement operations to small than to large sets accounts for the earlier success of children on small number problems; that the relative ease of measuring number also accounts for number conservation being present earlier than liquid and solid quantity conservation; and that what is learned is the effect of particular transformations on particular dimensions (in this case, the effects of spreading and compressing on number). Further, chunking provides a well-specified mechanism that could contribute to learning about the effects of transformations.

Data from the present microgenetic study, as well as previous findings, however, suggest that the reality is considerably more complicated than Q-SOAR suggests. The pretest data from the present study showed that one-half of nonconservers at times cited the type of transformation to justify their response. Thus, while not often relying on such reasoning before the training sessions, they knew enough about the effects of quantitatively relevant and irrelevant transformations on large sets to sometimes explicitly cite the knowledge in the conservation context. The origins of the knowledge may go back years further. When presented small sets of objects, even infants and toddlers show some understanding of the effects of both quantity-preserving transformations, such as spreading and compressing, and quantity-altering transformations, such as adding and subtracting (Gelman, 1972; Wynn, 1992). For example, when 5-month-olds are repeatedly shown three objects, they do not dishabituate when the length of the row is changed, but do so when an object is subtracted. Thus, rather than conservation acquisition involving sudden generation of a novel way of thinking near the end of the preschool period, it appears that knowledge of quantity-preserving and quantity-altering transformations originates years earlier, and is sometimes manifested in the specific conservation context before children become "conservers".

Changing distribution of approaches. How does this early knowledge of transformations come to be used consistently? Siegler and Shipley's (1995) ASCM model of strategy choice specifies mechanisms through which this may occur. Within ASCM, strategy choices are based on three types of information: information about each strategy's effectiveness on all problems in the domain (e.g., on all number conservation problems), on problems with particular features (e.g., on length-inconsistent problems), and on particular problems (e.g., on the effects of spreading seven objects and neither adding nor subtracting any). Each strategy has a strength, which reflects the accuracy and speed that it has produced when
applied to problems in the same general domain, problems with the same features as the given problem, and the given problem (if it has been encountered before). Probability of relying on any given strategy is proportional to that strategy's strength relative to the strength of all of the strategies.

ASCM has been implemented as a model of children's learning of arithmetic, not number conservation, but the main phenomena it produces fit the number conservation data equally well. Multiple strategies coexist for a prolonged period; performance shifts gradually toward increasing reliance on the more effective approaches; even within a given problem, use of a more effective approach on an earlier occasion is often followed by use of a less effective approach on a later occasion; and the more effective approaches are used earliest and most often on the problems on which they produce the greatest gains in accuracy and/or speed (e.g., transformational reasoning would be used first and most often on length-inconsistent problems).

ASCM also suggests an explanation for the effectiveness of asking children to explain the experimenter's reasoning. Trying to retrieve a strategy that would generate her responses would increase the strength of the transformational approach, which was invariably consistent with both the experimenter's answer and her lack of visible counting. It would simultaneously weaken other approaches, such as relying on length, which often predicted answers other than the ones she advanced. The effect of explaining the experimenter's reasoning would be expected to be considerably greater on the length-inconsistent than on the length-consistent problems, as indeed it was.

Rate and timing of change. The general overlapping wave approach, as well as the more specific ASCM model, focuses attention on the relative advantages and disadvantages of each competing strategy, and how these advantages and disadvantages change with age, as critical determinants of the rate and timing of cognitive changes. It suggests reasons both for why it takes so long before children rely on the type of transformation and for why they typically do so around age 5 or 6.

One likely reason that it takes so long for children to rely on transformations on number conservation problems is that the main alternative approach is reasonably effective—longer rows usually do have more objects. Another reason is limited evidence for the superiority of relying on transformations with large sets. Most children do not count large sets accurately until age 4 1/2 or 5 (Fuson, 1988). Until then, they cannot verify that transformations that do not influence small quantities also do not influence large ones. Further, even when children do count large sets accurately, it takes them more time to determine relative quantity in that way than by relying on relative length. To the extent that relying on length
is sufficiently accurate for their purposes, they may not count often enough to learn the advantages of relying on transformations on large sets.

Given these impediments, why do children ever come to rely on the type of transformation on large set number conservation problems? As Simon and Klahr (1995) noted, increasing skill at related quantitative competencies seems likely to be important. As counting and magnitude comparison processes become quicker, less effortful, and more accurate, they are used more often to quantify sets. As this occurs, they provide increasing evidence of the imperfect correlation between length and quantity, and of the perfect correlation between the type of transformation and whether the quantity changes. Another source of support for reliance on the type of transformation may be 4- and 5-year-olds' increasing knowledge of arithmetic. As children learn that adding invariably increases the original number, and that subtracting invariably decreases it, they may deduce that doing neither should leave the amount unchanged. Problems involving addition and subtraction of "1" may play an especially important role here. Thus, once children can verify the effects of transformations on large sets, they find that relying on the type of transformation has the advantage of being quick and easy, as well as accurate, and of fitting together logically with their increasing knowledge of arithmetic. In sum, early-originating knowledge about transformations involving small sets, general strategy choice mechanisms, and developing skill at related quantitative competencies all seem likely to be involved in 5- and 6-year-olds coming to rely consistently on the type of transformation to solve number conservation problems.

REFERENCES


HOW DOES CHANGE OCCUR


HOW DOES CHANGE OCCUR


(Accepted August 3, 1994)