What Leads Children to Adopt New Strategies? A Microgenetic/Cross-Sectional Study of Class Inclusion

Robert S. Siegler
Carnegie Mellon University

Matija Svetina
University of Ljubljana

Learning of class inclusion by 5-year-olds in response to empirical and logical explanations of an adult’s answers was examined. Contrary to the view that young children possess an empirical bias, 5-year-olds learned more, and continued learning for longer, when given logical explanations of correct answers than when given empirical explanations. Once children discovered how to solve the problems, they showed few regressions. Many children in the microgenetic experiment followed the path of change anticipated from previous cross-sectional studies, but children in the cross-sectional part of the study seemed to follow a different path. Reasons for the superior effectiveness of the logical explanations were discussed.

The primary goal of the present study was to compare young children’s learning from logical and empirical explanations of reasoning. A second main goal was to examine the relation between short-term microgenetic changes and long-term age-related changes. These goals were pursued in the context of the classic Piagetian class inclusion task, a task that can be solved either logically or empirically and that therefore allows direct comparison of the effects of encountering the two types of explanations.

This introductory section includes four parts. In the first part, we describe current understanding and unresolved issues regarding children’s adoption of new strategies in general. In the second, we examine current knowledge and unresolved issues regarding acquisition of class inclusion. In the third, we introduce the microgenetic/cross-sectional design as a way of addressing the main issues in the present study. In the fourth, we describe what we did in the study and our hypotheses regarding the results.

Acquiring New Strategies

A number of recent studies have examined how children discover new strategies (e.g., Alibali, 1999; Chen & Klahr, 1999; Goldin-Meadow & Alibali, 2002; Granott, 2002; Kuhn, Garcia-Mila, Zohar, & Anderson, 1995). These studies have shown that discovering a strategy is often only the first step in strategic change; prior strategies often persist, even when the newly discovered strategy has clear advantages and even when children can describe those advantages (Siegler & Jenkins, 1989). Thus, the adoption of new strategies is often quite slow.

It is also the case, however, that the rate of adoption of new strategies varies considerably across studies. Two factors that seem to be related to this rate have been identified: accuracy and efficiency. A recent literature review (Siegler, 2006) concluded that the largest factor in determining the rate of adoption of newly discovered strategies is the difference between their accuracy and that of previously available strategies. In the relatively few cases in which newly discovered strategies quickly replace previous approaches, the new strategies have entailed large improvements in accuracy, generally from consistently incorrect to consistently correct. Differences between the efficiency of new and previous strategies also appear to influence the rate of adoption of new strategies, although to a lesser degree than differences in accuracy do. When both new and old strategies generate correct performance, new strategies that confer large advantages in speed or substantial reductions in processing steps are adopted more rapidly than new strategies that confer less dramatic advantages (Siegler, 2006).

In the present study, we examined a third factor that may influence the rate and completeness of adoption of new strategies: whether there are logical as well as empirical reasons for adopting them. Some strategies are based on the fundamental logic of the problem. For example, on conservation of weight problems, one correct logical strategy is to note that...
nothing was added or subtracted when the clay was remodeled; hence the weight must be the same. Other strategies are empirical; for example, rather than solving the conservation of weight problem logically, a child could take the empirical approach of weighing the clay before and after its shape was transformed, and noting that the weight remained the same. Many other problems can also be solved either empirically or logically. For example, number conservation tasks can be solved either by counting the objects in each row or by noting that nothing was added or subtracted; transitivity of length problems can be solved either by measuring the length of the sticks being compared or by drawing the logical inference “If A > B and B > C, then A > C”; arithmetic problems of the form A + B = B can be solved either by adding the first two numbers and subtracting the third or by deducing that the answer must be “A,” because adding and subtracting the same number cannot have an effect; and so on.

The specific issue addressed in the present study was whether young children learn more rapidly and completely when they are encouraged to use logical or empirical strategies for solving a problem. In particular, we compared 5-year-olds’ learning in response to logical and empirical explanations of correct answers to class inclusion problems. Contrasting acquisition of two strategies for solving the same problem was a more direct way of establishing factors that influence acquisition of new strategies than previous techniques, which involved reviewing findings from different tasks and populations, and then attempting to isolate the factors associated with more and less rapid and complete acquisition (Siegler, 2006).

There were reasonable arguments for expecting either logical or empirical strategies to be acquired more effectively. First, consider arguments that suggest that young children should acquire empirical strategies more easily. Age peers of the children whose learning was examined in the present study, 5-year-olds, have been found to rely on empirical strategies in a number of situations in which older children and adults use logical reasoning strategies. When presented tautologies (e.g., “I either am holding a green chip in my hand or I’m not”) or contradictions (“I am holding and not holding a green chip in my hand”), and asked whether they needed the experimenter to open her hands to tell whether the statement was true or false, 5-year-olds usually chose the empirical strategy; in both cases, most children said they needed the experimenter to open her hands to tell (Osherson & Markman, 1975). Most 5-year-olds also confuse logically determinate and logically indeterminate problems and are subject to “positive capture,” in which they think that a case consistent with a conclusion implies that the conclusion must be true (Fay & Klahr, 1996; Klahr & Chen, 2003). On class inclusion tasks in which children are asked whether a set or the larger of two subsets has more objects (e.g., “I have five dogs and three cats; do I have more dogs or more animals”), most 5-year-olds compare the number of objects in the two subsets (e.g., five dogs versus three cats) rather than reasoning that there must be more in a set than in any of its subsets (Piaget, 1952). Young children have also been found to use empirical strategies on other tasks somewhat similar to class inclusion. When asked whether they could place more spoons than silverware on the table (Markman, 1978) or whether they could draw a picture that has more cats than animals (Miller, 1986), most 5-year-olds said “yes,” rather than reasoning that this was logically impossible.

Also favoring the prediction that children should learn more from empirical explanations, 5-year-olds seem to have a particular fondness for the type of empirical strategy relevant to class inclusion, counting. They use counting even on tasks where such a strategy is inappropriate. For example, Miller (1984) found that 5-year-olds created “fair” divisions of food between two birds by counting out equal numbers of pieces of food, regardless of the sizes of the pieces. Similarly, Levin (1989) found that 5-year-olds often counted the passage of two temporal durations and judged the higher number to indicate the longer time, regardless of whether the counting had proceeded at the same speed. This reliance on counting strategies, and on empirical rather than logical strategies in general, suggested that 5-year-olds would learn empirical strategies more readily than logical ones.

On the other hand, Siegler and Svetina (2002) suggested that logical strategies might be learned more quickly than empirical ones. The above-cited finlings of young children using empirical strategies in situations where older children and adults use logical ones may reflect the young children not understanding the logical strategies. If the logic was made clear to them, they might prefer the logical strategies (or have no preference between the two). In addition, several of the microgenetic studies in which newly discovered strategies have been adopted most rapidly have involved acquisition of logical reasoning strategies. These include strategies for solving matrix completion, A + B = B, and 20 Questions problems (Siegler & Stern, 1998; Siegler & Svetina, 2002; Thornton, 1999). Thus, the fact that
young children use empirical strategies to solve problems that adults and older children solve via logical strategies does not imply that children would choose empirical strategies if they understood that the logical strategies could be used to solve the problems.

Moreover, there were reasons to think that young children might prefer logical reasoning strategies if they understood how to execute and apply them to particular problems. Logical strategies often have large advantages in speed and accuracy. In the context of class inclusion, solving the problem logically both requires less time and avoids the possibility of errors in counting and numerical comparison. Anticipating these benefits, or learning about them through experience with the strategies, might lead children to consistently use new logical strategies more rapidly than new empirical ones. In addition, to the extent that children are searching for basic logical understanding of domains, as theorists such as Gelman and Williams (1998) and Keil (1998) have proposed, logical strategies might be adopted faster and more completely than empirical strategies precisely because they are logically compelling.

Development of Class Inclusion Understanding

The development of class inclusion was first studied by Piaget (1952). Piaget defined understanding of class inclusion as the ability to compare sets of objects that are at different levels of a hierarchical organization. He viewed it as a crucial type of reasoning because it brings together understanding of classes and relations, requires understanding of hierarchies, and requires reliance on logical rather than intuitive reasoning under conditions that somewhat favor intuitive reasoning (Inhelder & Piaget, 1964). These features have also made the task of continuing interest.

To investigate the development of class inclusion, Piaget (1952) and Inhelder and Piaget (1964) presented children with drawings of two types of entities from a common set, for example five dogs and two cats, and asked whether there were more entities in the set or in the larger subset (more animals or dogs). Children younger than age 7 or 8 tended to say that there were more objects in the larger subset, which Inhelder and Piaget interpreted as supporting their theory that preoperational children are unable to view a single object as being simultaneously a member of both a class and a subclass (e.g., as being both an animal and a dog).

Inhelder and Piaget’s findings triggered a flurry of research on the development of understanding of class inclusion. These studies replicated the original finding that children younger than 7 or 8 years rarely solve the Piagetian class inclusion task consistently correctly. Reviews of the class inclusion literature (Halford, 1993; Winer, 1980) indicate that not until around 8 years of age do 50% of children in the United States correctly answer the standard Piagetian version of the task, and that not until age 9 or 10 do 75% of children do so.

Subsequent research also extended Inhelder and Piaget’s findings in several directions. Success on class inclusion tasks was found to correlate positively with success on other measures of logical reasoning, including conservation of number, liquid quantity, and solid quantity (Tomlinson-Keasey, Eiser, Kahie, Hard-Brown, & Keasey, 1979). Many task variables were found to influence class inclusion performance (Trabasso et al., 1978; Wilkinson, 1976). One involved the number of objects in the subsets: performance was best when the subsets were equally numerous and worst when one subset had far more objects than the other (Ahr & Youniss, 1970). Linguistic variables also proved highly influential: describing the set via collection nouns (e.g., forest, army, team), which call attention to part—whole relations, greatly facilitated performance relative to describing the set via class nouns (trees, soldiers, children), as in the original Piagetian studies (Markman & Seibert, 1976). Another influential variable was the typicality of subordinate class members; using typical category exemplars (e.g., dogs and horses) was associated with better performance than using atypical ones (flies and bees) (Carlson & Abrahamson, 1976). Children were also found to use class inclusion reasoning more often on tasks that did not include misleading properties; for example, after being told “A pig is a kind of dog,” 4-year-olds respond “yes” more often than chance when asked, “Is a pug a kind of animal?” (Smith, 1979).

Much less is known about how children learn about class inclusion than about their performance at different ages. However, some informative data have been obtained. These data can be described within Siegler’s (1996) five dimensions of change framework, which characterizes changes in terms of their path, rate, breadth, source, and variability. The path of change concerns the sequence of knowledge states or problem-solving approaches that children use while gaining competence. The rate of change involves the amount of time/experience before children’s first use of a new approach and the amount of time/experience that separates initial use of a new approach from consistent use of it. The breadth of change involves changes in performance on related
tasks. The source of change concerns the causes that set the change in motion. The variability of change concerns between- and within-child variability on the other dimensions of change.

Analyzing the acquisition of class inclusion understanding in terms of these five dimensions indicates that current understanding is largely limited to two areas. The greatest amount of previous research has examined the path of change. Several investigators (Chapman & McBride, 1992; Hodkin, 1987; McCabe, Siegel, Spence, & Wilkinson, 1982) have suggested a three-phase progression. In the first phase, between ages 4 and 6 years, children are said to solve class inclusion problems by comparing subclasses, which leads to below chance performance. In the second phase, around 6 to 7 or 8 years, children answer correctly at about a chance level (50%). This has been hypothesized to come about through guessing (e.g., Hodkin, 1987), although it also could come about through some children answering consistently correctly and other children answering consistently incorrectly (or through all children using a strategy other than guessing that generates a chance level of performance.) In the third phase, beginning at around age 8 or 9 years, children are said to compare the set with the larger subset, and therefore to generate consistently correct answers and explanations.

Some information is also available regarding the source of change. In particular, explicitly telling children about the relation between superordinate and subordinate classes has been shown to help children learn about class inclusion (Judd & Mervis, 1979). Growth of working memory and acquisition of inclusion schema also appear to be sources of improved understanding of class inclusion (Halford, 1993; Rabinowitz, Howe, & Lawrence, 1989).

Little is known about the other three dimensions of cognitive change. With regard to the breadth of change, success at class inclusion correlates positively with success on other logical reasoning tasks, but whether gaining understanding of class inclusion has broader effects is unknown. With regard to the rate of change, the fact that 5-year-olds solve some variants of the problem but that many 8-year-olds do not solve the classic Piagetian task might be taken to imply a slow rate. However, it is unclear whether the classic task and the variants are solved via the same cognitive processes, and very little is known about the rate of acquisition of any particular task in response to relevant experience. With regard to the variability of change, essentially nothing is known. To advance an understanding of these five dimensions of cognitive change, we used a microgenetic/cross-sectional design in the present experiment.

**Microgenetic/Cross-Sectional Designs**

Over the past 2 decades, an increasing number of investigators have adopted the microgenetic approach as a means for studying cognitive development (Kuhn, 1995; Miller & Coyle, 1999; Siegler, in press). The main reasons are the precise descriptions of changing competence that the approach can yield and the implications of these precise descriptions for understanding change processes.

The microgenetic approach is defined by three primary characteristics: (1) observations span the whole period of rapidly changing competence; (2) the density of observation within this period is high, relative to the rate of change; and (3) observations of changing performance are analyzed intensively to indicate the processes that give rise to them. The second characteristic is especially important. Dense sampling of performance while the performance is changing provides the temporal resolution needed to describe adequately the process of change. Often, microgenetic designs track strategic change on a trial-by-trial basis (e.g., Adolph, 1997; Alibali, 1999). The data yielded by such dense sampling of strategy use allow identification of the precise trial on which a child discovered a new strategy, which in turn allows analyses of the events that led up to the discovery and of how the strategy was generalized beyond its initial use.

An especially encouraging characteristic of microgenetic studies is that despite the diversity of content areas and populations to which they have been applied, findings from them have been surprisingly consistent (Kuhn, 1995; Miller & Coyle, 1999; Siegler, 2006). A common finding regarding the path of change is that just before discovery of a new approach, performance becomes more variable (Goldin-Meadow, Alibali, & Church, 1993; Graham & Perry, 1993; Siegler, 1995; Siegler & Svetina, 2002). The rate of change tends to be gradual, with less adequate, initial approaches continuing to be used well after more advanced approaches have also emerged (Bjorklund, Miller, Coyle, & Slawinski, 1997; Kuhn et al., 1995; Siegler & Jenkins, 1989). The breadth of change is usually relatively narrow (Kuhn et al., 1995; Schauble, 1990, 1996; Siegler, 2006). Variability tends to be high: children learn via different paths, at different rates, and with differing amounts of generalization. Finally, certain sources of cognitive growth, such as encouragement to explain observations, operate over a wide age range and in
diverse content domains (Bielaczyc, Piroli, & Brown, 1995; Chi, de Leeuw, Chiu, & LaVancher, 1994; Pine & Messer, 2000; Siegler, 2006). Thus, examining cognitive growth along the five dimensions has proven useful for identifying regularities in how change occurs.

One question that has been raised about microgenetic studies concerns the similarity of changes within them to changes with age in traditional cross-sectional and longitudinal designs. There is widespread agreement that the changes are similar at a general level but much less agreement as to the extent of the similarity (Granott, 1998; Kuhn, 1995; Miller & Coyle, 1999). For example, after reviewing the microgenetic literature, Miller and Coyle (1999, p. 212) concluded, “Although the microgenetic method reveals how behavior can change, it is less clear whether behavior typically does change in this way in the natural environment” (italics in original).

To address this issue, Siegler and Svetina (2002) proposed the microgenetic/cross-sectional design. The basic strategy was to combine cross-sectional and microgenetic components within the same experimental design, using the same population, tasks, instructions, and measures of performance. In the first use of this design, Siegler and Svetina found that microgenetic and age-related changes in acquisition of matrix completion were highly similar. The question investigated here was whether microgenetic and cross-sectional analyses of acquisition of class inclusion would also show extensive parallels.

The Present Study

In the microgenetic experiment in the present study, 5-year-olds were presented 30 class inclusion problems, 10 problems in each of three sessions. The children were randomly divided into four groups; the groups varied in the explanations provided to children regarding why the correct answer was correct. Consider the explanation that children in each group received following incorrect answers to a problem that involved six dogs and three cats. Children in the empirical explanations group were told that there were more animals than dogs because there were nine animals but only six dogs. To illustrate the counting procedure that yielded this outcome, the experimenter pointed to each animal in turn before saying there were nine, and to each dog in turn before saying there were six. Children in the logical explanations group were told that there were more animals than dogs because dogs are just a type of animal; therefore, if there were dogs and some other animals, there must be more animals than dogs. Children in the logical and empirical explanations group were provided both explanations. Children in the no explanations group were given feedback on the correctness of their answer (as were children in the other three groups) but were not provided any explanation.

On trials on which children answered incorrectly, they were asked to judge how smart the experimenter’s explanation of the correct answer was. This was intended to provide a measure of conceptual understanding of the logical and empirical explanations. Previous studies have shown that conceptual understanding, as indexed by ability to answer questions about the intelligence of strategies, often diverges from procedural knowledge, as indexed by use of correct strategies. Sometimes conceptual understanding precedes use of correct strategies (Siegler & Crowley, 1994), sometimes the reverse is true (Briars & Siegler, 1984), and sometimes the two types of knowledge are highly similar (Cauley, 1988).

This microgenetic design allowed us to characterize the source, rate, path, breadth, and variability of change. The source of change was examined by comparing changes in performance among children who received the four types of explanations. The rate of change was examined by analyzing the number of trials that elapsed before children first used the correct strategy and the number of trials before they did so consistently. The path of change was examined by analyzing performance in the period of rapid learning—the trials immediately before and immediately after children discovered the correct solution to the problem. The breadth of change was examined by analyzing the relation between changes in strategy use and changes in conceptual understanding, as measured by children’s evaluations of the smartness of logical and empirical explanations as well as their answers to problems. The variability of change was examined by analyzing individual differences in the other dimensions of change.

In the cross-sectional part of the study, 5-, 6-, 7-, 8-, 9-, and 10-year-olds were presented 10 class inclusion problems without any feedback. The problems were randomly sampled from the set of 30 presented to 5-year-olds who participated in the microgenetic part of the study. Because children in the cross-sectional and microgenetic parts of the study were chosen from the same population, and because the identical problems, procedures, and measures were presented in both parts, it was possible to evaluate in an unusually direct way the similarity between microgenetic and age-related change.

We tested four hypotheses. The first hypothesis was that accuracy in all four groups in the micro-
genetic segment of the study would improve over trials and sessions. If 5-year-olds’ predominant strategy at the outset was to compare the two subsets, as the literature suggested, the feedback that children in all groups received would disconfirm their initial approach and therefore move them from below chance to chance or above chance performance. This was expected to be due to the feedback, rather than to simple exposure to class inclusion problems or experience answering them. Thus, no comparable improvements in accuracy were expected over trials for 5-year-olds in the cross-sectional part of the study.

The second hypothesis was that both logical and empirical explanations would increase learning. Both types of explanation illustrated strategies that would yield accurate performance. In contrast, feedback would disconfirm incorrect approaches but would not indicate a correct one.

The third hypothesis was that logical explanations would elicit greater learning than would empirical explanations. Previous findings in which 5-year-olds used empirical rather than logical strategies may have been due to the children not realizing that the problems could be solved logically. That is, the preschoolers might reflexively assume that questions asked by an adult must be solved empirically, because that almost always is the case. How often are young children (other than those of academics) presented questions where the answer can be deduced from the question itself? This view, that preschoolers may not consider the possibility that the information needed to answer a question is contained in the question itself, is different from saying that preschoolers cannot use such strategies. Indeed, we believe that if it is clear to young children that a problem can be solved deductively, they will find such strategies preferable to empirical approaches. The sources of appeal of the logical strategies for young children would presumably be the same as for older individuals: perfect accuracy, rapid execution, and intellectual elegance. Although this final characteristic is difficult to define, we hoped to measure it by obtaining children’s and adults’ evaluations of the smartness of empirical and logical strategies.

A final hypothesis involved the relation between microgenetic and age-related change. Evidence from microgenetic studies suggests that after using an incorrect strategy for a substantial period of time, children often begin to guess or oscillate among alternative approaches before generating a stable, advanced strategy (Siegel & Svetina, 2002; Spencer, Vereijken, Diedrich, & Thelen, 2000; Thelen & Corbetta, 2002; van Geert, 2002). It seemed likely that receiving feedback on an unfamiliar task such as class inclusion would lead after a short time to guessing or inconsistency in the present task as well. Thus, we expected to see the same three-state path of change in the present microgenetic context as has been proposed in the literature on age-related change in class inclusion performance—first reliance on subset—subset comparisons, which produce systematically incorrect performance, then, guessing or oscillation among approaches, which produces approximately chance performance, then, set—subset comparisons, which produce consistently correct performance.

Method

Participants

The children were 200 Caucasian 5- to 10-year-olds of Slovenian ethnic background who were recruited from kindergartens and elementary classrooms of several schools in a predominantly middle-class area in Ljubljana, Slovenia: one hundred 5-year-olds, M = 66.7 months, range = 60–71 months; twenty 6-year-olds, M = 79.4 months, range = 75–83 months; twenty 7-year olds, M = 87.4 months, range = 84–91 months; twenty 8-year-olds, M = 100.5 months, range = 96–107 months; twenty 9-year-olds, M = 111.8 months, range = 108–119 months; and twenty 10-year-olds, M = 124.4 months, range = 120–130 months. Most children in each school returned the consent forms; the 200 children were randomly selected from among those children. Children seemed to enjoy the study; the only children who did not complete the study were 5 who either were ill for an extended period or moved. A 35-year-old male researcher served as the experimenter.

Task and Procedure

Class Inclusion Task

Children in the microgenetic part of the study received 30 trials, 10 trials in each of three sessions, with the sessions presented over a period of 2 weeks, with an average of 5.4 days between successive sessions for each child. Children in the cross-sectional part of the study received a randomly chosen 10 of these 30 problems in a single session.

On each trial, children in both the microgenetic and cross-sectional segments were presented a piece of paper with a black-and-white drawing. Each drawing included 8–10 objects (e.g., animals), of which 2 or 3 belonged to one subclass (e.g., cats) and
6 or 7 to a different subclass (e.g., dogs) (figure 1). The drawings included diverse categories of objects from children’s everyday experience. The categories were trees (apples and pines), food (pieces of bread and sausage), toys (bears and balls), musical instruments (violins and trumpets), adults (men and women), children (boys and girls), clothing (jackets and pants; skirts and pants), footwear (shoes and boots), garden tools (rakes and spades), tools (hammers and saws), furniture (chairs and chests; tables and chairs), headwear (caps and hats), silverware (forks and spoons), fruit (apples and pears; bananas and strawberries), vehicles (trucks and buses), vegetable (beets and carrots), birds (crows and owls), kitchen tools (pots and pans), flowers (tulips and daisies), letters (A’s and G’s; B’s and E’s), numbers (1’s and 2’s), and animals (geese and chickens; horses and cows; rabbits and turtles; dogs and cats; cats and dogs). (Dogs and cats are listed twice, because on one problem there were more dogs and on another problem there were more cats.) The ratios of subclasses to classes were 2:6, 2:7, 3:6, or 3:7.

Problems were presented in random order, with the randomization performed separately for each child. On every item, children were asked to count the number of objects in one subset, then the number in the other subset, and then the number in the set. For example, on one problem, the experimenter said (pointing to the corresponding objects), “Look at these pictures; first count all the dogs, then all the cats, and then all the animals.” After each answer, the experimenter said, “OK.” Next, the experimenter asked the child either (e.g.,) “Are there more dogs or more animals in the picture?” or “Are there more animals or more dogs in the picture?” followed by the question “How did you know?” The set was mentioned first on half of the questions within each session; the larger subset was mentioned first on the other half.

Experimental Conditions

The eighty 5-year-olds who were in the microgenetic experiment were randomly assigned to the four experimental conditions. Children in all four groups were told, “Yes, that’s right” when they answered correctly and were told “No, that’s wrong” when they answered incorrectly. Where the groups differed was in the explanation that the experimenter provided on trials where a child answered incorrectly. In the empirical explanations condition, the experimenter said, “A couple of days ago, I asked a teacher about that problem. She said there were more animals than dogs, because there were nine animals (pointing to each of the animals) and only six dogs (pointing to each of the dogs).” In the logical explanations condition, the experimenter said, “A couple of days ago, I asked a teacher about that problem. She said there were more animals than dogs, because dogs are a type of animal, so if there are some dogs and some other animals, there must be more animals than there are dogs.” In the both explanations condition, the experimenter provided both empirical and logical explanations by saying that she had asked two teachers and one had provided one explanation and the other had provided the other (each type of explanation was stated first on half of the trials.) In the no explanations condition, the experimenter said, “A couple of days ago, I asked a teacher about that problem. She said that there were more animals than dogs because he/she just knew it.” In all four groups, children were asked after hearing the explanation, “Do you think this answer was very smart, not smart, or kind of smart?”

The one hundred and twenty 5- to 10-year-old participants in the cross-sectional segment of the study were presented 10 class inclusion problems, as in Session 1 of the microgenetic experiment. However, they were not provided explanations or feedback concerning the correctness of their answers.

Adults’ Evaluations of the Explanations

To test the assumption that adults would see logical explanations as superior to empirical ones, we asked 60 college students, aged 18–21 fall Caucasian, in their first semester in the School of Arts, Ljubljana, Slovenia, to rank how smart different types of explanations to class inclusion problems were. Specifically, we asked them to rank four
potential responses to class inclusion problems: correct answer and logical explanation; correct answer and empirical explanation; correct answer and no explanation; and incorrect answer and no explanation. For example, the college students were presented an overhead showing eight animals and the following story: "There were three horses and five cows in the yard. A group of children came by. They stopped and a teacher asked the children whether there were more cows or more animals in the yard. John said there were more animals in the yard because cows were just a type of animal. Ann said there were more animals in the yard because there were eight animals and only five cows. Brigit said there were more animals in the yard because she just knew it. Dennis said there were more cows in the yard because he just knew it." The college students were asked to rank the responses on each problem from 1 (smartest) to 4 (least smart). The overhead displaying a given problem and the four responses to it remained visible until the college students completed their ranking. Each student was presented four such problems, each including the same four types of answers and explanations. Cronbach’s alpha coefficients indicated that students were consistent in their judgments across the four stories—coefficient as for the logical and empirical explanations were .84 and .91, respectively.

Results

To test our assumption that adults viewed the logical explanations as more intelligent than the empirical ones, we first analyzed the college students’ ratings of the four answers and explanations. An analysis of variance (ANOVA) yielded differences among their evaluations of the four responses, $F(3, 177) = 310.92$, $p < .001$, $\eta^2 = .84$. The logical explanations (mean ranking of 1.07 on the 1–4 scale) were ranked as smarter than the empirical explanations (mean 2.08), which were ranked as smarter than correct answers without explanations (mean 2.88), which were ranked as smarter than incorrect answers without explanations (mean 3.94). These data supported our expectation that adults would find logical explanations smarter than empirical ones, and empirical explanations smarter than no explanation.

The remaining analyses examined children’s performance in the main experiment. On each trial, we coded children’s answers and their explanation for it. However, the explanations proved uninformative. The large majority of children said, “I don’t know,” or failed to respond when asked why they answered as they did. Efforts to motivate children to provide more informative explanations were unsuccessful. For this reason, we relied exclusively on the pattern of answers to infer strategy use.

The results are presented in two sections. In the first section, we describe findings from the microgenetic experiment; these results are organized in terms of the source, path, rate, breadth, and variability of change. In the second section, we describe changes with age in class inclusion performance and compare microgenetic and age-related changes. If not otherwise stated, post hoc comparisons are Newman–Keuls tests with $p < .05$.

Microgenetic Experiment

Source of Change

To examine the effects of the logical and empirical explanations on children’s class inclusion performance, we first conducted a 2 (logical explanations: present or absent) × 2 (empirical explanations: present or absent) × 3 (session) ANOVA on each child’s number of correct answers. The analysis revealed main effects of logical explanations, $F(1, 76) = 4.94$, $p < .05$, $\eta^2 = .06$, and session, $F(2, 152) = 35.20$, $p < .001$, $\eta^2 = .32$. Children who received logical explanations generated more correct answers than children who did not, and the number of correct answers increased over sessions.

A Logical Explanations × Session interaction was also present, $F(2, 152) = 5.85$, $p < .01$, $\eta^2 = .07$. To better understand the source of the interaction, separate Logical Explanations × Empirical Explanations analyses were undertaken for each session. In Session 1, performance did not differ as a function of the type of explanation that children received. In contrast, in both Sessions 2 and 3, children who received logical explanations answered the class inclusion problems more accurately than those who did not, $F(1, 76) = 6.23$, $p < .05$, $\eta^2 = .08$, and $F(1, 76) = 6.10$, $p < .05$, $\eta^2 = .07$. As shown in Figure 2, children who received logical explanations already answered somewhat more accurately than those who did not in Session 1, but the size of the difference was greater in Sessions 2 and 3. No main effect of empirical explanations and no interaction involving that variable were found.

Children were exposed to the logical and empirical explanations only when they answered incorrectly; therefore, the session variable only roughly corresponded to children’s experience with the experimental manipulations, because different children in a given group received different numbers of explanations. To examine more directly the effect of the history of explanations on children’s accuracy,
The effect of logical explanations and the lack of effect of empirical ones were not due to preexisting differences among the children who were presented the two experimental manipulations. Performance on the first trial of the study, the only trial before children in the four groups received different treatment, was closely similar; percent correct of children who did and did not receive logical explanations was identical, 40% correct, and percent correct among children who did and did not receive empirical explanations was almost identical, 42% and 38%, respectively.

The data also allowed us to examine whether feedback in the absence of explanations was sufficient to improve children’s accuracy. Surprisingly, feedback was insufficient for learning to occur. The evidence came from a comparison of the performance of 5-year-olds in the cross-sectional comparison, who did not receive feedback, with that of peers in the no explanations condition of the microgenetic experiment, who did receive feedback. The 5-year-olds in the cross-sectional comparison received 10 class inclusion problems, selected from the same set as the 10 problems received by peers in the no explanations condition in Session 1. The problems were not identical, but did not differ in any systematic way.

The number of correct answers on the 10 problems of 5-year-olds in the cross-sectional comparison did not differ from the number of correct answers of peers in the microgenetic experiment who received feedback without explanations, t(38) = 1.01, p > .05, d = 0.32. In addition, the accuracy of both groups declined slightly from the first five to the second five problems that they received, suggesting that neither exposure to the problems nor receiving feedback elicited learning. Moreover, in the no explanations condition in the microgenetic experiment, percent correct did not increase over the three sessions, F(2, 38) = 2.55, p > 0.05, η² = .12. Thus, feedback on whether responses were correct was insufficient for 5-year-olds to improve their class inclusion performance.

Rate of Change

The trial of discovery was operationally defined as the first trial within the first set of five consecutive correct answers that the child generated. Thus, if a child answered incorrectly on Trials 1 and 4 and then correctly on Trials 5–9, the child would be said to have made the discovery on Trial 5. Given that one of the two alternative answers on each problem was correct, the chance probability of generating five
consecutive correct answers was .5³ or approximately 3%. This criterion for discovery was chosen to minimize the probability of both false positives and false negatives. The results did not vary substantially with the particular criterion; for example, 70% of children made the discovery at some time during the three sessions by the five consecutive correct answers criterion, and 74% would have made it if a three consecutive correct answers criterion had been used.

Most children who ever met the criterion met it quite early in the experiment. In particular, of the children who met the criterion in any of the three sessions, 82% met it in Session 1.

To examine the rate of discovery more precisely, we graphed the cumulative percentage of discoveries that had occurred by each of the 30 trials within the microgenetic experiment (Figure 3). In addition to illustrating that more children in the two groups that received logical explanations made the discovery, Figure 3 also illustrates that the difference between those who did and did not receive logical explanations arose through the logical explanations condition stimulating discoveries over a longer period. The difference is particularly clear in comparing the rates of discovery of children who only received the logical explanations with that of children who only received the empirical ones. Through the first 5 trials, more children who only received empirical explanations than who only received logical ones made the discovery (35% vs. 20%). In contrast, many more children who received logical explanations made the discovery after the fifth trial (45% vs. 15%). Viewed from a different perspective, 69% of children who received logical explanations and who had not yet made the discovery by Trial 5 made it afterward, whereas only 30% of children who received empirical explanations did. \( \chi^2(df = 1, N = 23) = 3.49, p < .10, V = .39 \).

Once children met the discovery criterion, they used the correct approach with striking consistency. This can be seen by examining performance in the next session, which was removed by a mean of 5.4 days from the session in which the discovery occurred. Children in previous multisession microgenetic studies have often shown a great deal of regression to previous strategies from one session to the next (e.g., Siegler & Jenkins, 1989; Siegler & Stern, 1998), but this was not the case in the present study. Children who discovered the correct strategy in Session \( N \) answered correctly on 97% of the first five questions in Session \( N+1 \). The findings did not depend on the original discovery criteria requiring a relatively long string of consecutive correct answers. When a looser discovery criterion of three consecutive correct answers was used, children who met the criterion in one session answered 91% of problems correctly on the first three trials of the following session.

To examine the effect of the instructional variables on the consistency of use of the newly discovered correct strategy, we performed a 2 (logical explanations: present or absent) \( \times \) 2 (empirical explanations: present or absent) ANOVA on the number of incorrect answers after the child met the discovery criterion. Consistency of use proved to be unaffected by these variables. Once children discovered the correct solution, they used it consistently, regardless of the type of explanation that they were given; in none of the four conditions did children average more than 1.3 errors in the more than 20 trials (on average) following their discovery.

**Path of Change**

To examine where the path leading to the discovery began, we compared with chance children’s percent correct on all trials before the discovery. The results showed that the percentage correct on trials before the discovery (34%) was below chance, \( t(40) = 3.85, p < .001, d = 0.60 \). However, the data appeared to include two subperiods: an earlier period in which accuracy was far below chance and a later...
period in which it was at a chance level. It also appeared that the five consecutive correct trials that operationally defined the discovery marked the transition to a third period, one of consistently correct performance. Thus, acquisition of understanding of class inclusion appeared to include three periods: a period of consistently incorrect responding, a period of chance responding, and a period of consistently correct responding.

Several sources of evidence were consistent with the conclusion that most children began the study with a systematically incorrect approach to the problems. On the first item, the only item presented before children received feedback, accuracy was below chance. Excluding the 15 children who came to the experiment already knowing how to solve the problems (as defined by their generating five correct responses on the first five items), only 26% of children answered the first problem correctly, which was well below chance, $t(64) = 4.34$, $p < .001$, $d = 0.55$. Another source of evidence was that among the 30% of children who never met the criterion of five consecutive correct answers, percent correct was far below chance, 6% correct, $t(23) = 23.03$, $p < .001$, $d = 4.65$. Lending further support to the conclusion of an initial state that produced below chance responding, the accuracy of age peers in the cross-sectional sample, who were not presented feedback on any of the 10 problems they encountered, was also well below chance, 30%, $t(19) = 3.57$, $p < .01$, $d = 0.56$. More than half of the 5-year-olds in the cross-sectional sample (55%) were correct on 10% or fewer of the problems.

Evidence for a period of chance levels of accuracy and a period of consistently correct performance came from a comparison of the percentage of correct answers on the five trials before the discovery and the five trials immediately after it. This comparison was limited to the 28 children who made the discovery, performed at least three trials before it, and performed at least three trials after it; the reason for the last two criteria was to allow reasonably stable estimation of performance both before and after the discovery. Performance on the trials immediately before discovery was at a chance level, 46% correct, $t(27) = 0.85$, $p > .05$, $d = 0.16$. Performance on the trials immediately after the discovery was very accurate and far above chance, 90% correct, $t(27) = 7.45$, $p < .001$, $d = 1.99$. Again, the particular criteria for inclusion in the analyses were not critical; the same pattern emerged with a criterion of at least two trials before the discovery, 43% correct, $t(32) = 1.44$, $p > .05$, $d = 0.27$; and after it, 90% correct, $t(32) = 8.38$, $p < .001$, $d = 2.13$.

**Breadth of Change**

To better understand the breadth of change, we examined children's evaluations of the smartness of the logical and empirical explanations before and after the discovery. Evaluations of the smartness of strategies have been used to measure conceptual understanding in a number of contexts, including counting, simple addition, and complex subtraction. Findings based on this measure of conceptual understanding often diverge from findings regarding children's use of appropriate procedures. Sometimes, children use correct procedures before they show conceptual understanding, as indexed by judgments of smartness. For example, Briars and Siegler (1984) and Frye, Braisby, Love, Maroudas, and Nicholls (1989) found that many preschoolers correctly use the standard counting procedure at ages at which their evaluations of the smartness of the strategies reveal a lack of conceptual understanding. In other cases, the opposite pattern has emerged. For example, in a study of children's addition strategies, Siegler and Crowley (1994) found that 5-year-olds judged a puppet's use of the min strategy to be smarter than alternatives despite not using the strategy themselves. On yet other tasks, judgments of smartness are closely related to use of correct procedures; for example, Cauley (1988) found such a relation between second and third graders' knowledge of the long subtraction algorithm. Thus, examining children's judgments of the smartness of logical and empirical explanations provided a means of assessing conceptual understanding of the task, and thus to broaden the knowledge assessments based on children's use of correct or incorrect procedures.

As noted previously, college students, who presumably understood class inclusion, rated logical explanations as being smarter than empirical ones. The question was whether 5-year-olds, who did not initially solve class inclusion problems, would also evaluate such explanations as smarter. To find out, we compared the evaluations of logical and empirical explanations of children who received both types of explanations. Before the discovery, children in this group judged logical explanations to be smarter than empirical explanations, means of 2.81 versus 2.19 on a 3-point scale, with 3 being very smart, $t(12) = 2.85$, $p < .05$, $d = 0.91$. The difference in evaluations before the discovery, together with the greater learning from hearing logical rather than empirical explanations among children who heard one but not both, suggested that even before children consistently solved class inclusion problems, they, like college students, found logical explanations more compelling than
empirical ones. After the discovery, evaluations of the two explanations did not differ, \( t(6) = 0.22 \). This lack of a difference seems attributable to the low power that resulted from only 7 children in the group ever erring after discovering how to solve the problems, and even most of those children erring only once and therefore providing few evaluations.

Variability of Change

Variation in rates. To analyze variability in the rate of change, we examined individual children’s learning in the three sessions. Four distinct patterns appeared (Table 1). Twenty-eight percent of children fit the fast-learner pattern, which was operationally defined as 90% or more correct answers in all three sessions. These children learned extremely quickly—the criterion of at least 90% correct in all three sessions meant that a child in this group could only advance one incorrect answer in the first session. Another 29% of 5-year-olds fit the medium-rate learner pattern, which was defined as fewer than 90% correct answers in Session 1 and 90% or more correct answers in Sessions 2 and 3. An additional 6% of children fit the slow-learner pattern, which was defined as fewer than 90% correct answers in Sessions 1 and 2, and at least 90% correct in Session 3. Finally, 28% of children fit the nonlearner pattern, which was defined as 30% or fewer correct in all three sessions.

The learning of 90% of children fit one of these four patterns. This percentage is quite striking, because all four patterns depart markedly from a chance distribution, which would have an expected value of 50% correct in each session. The four learning patterns also indicated that strategy acquisition was highly stable—regressions from 90% or more correct in an earlier session to below 90% correct in a later session were extremely rare—only 4% of the participants (3 children). Moreover, the bimodality of the patterns—in 51% of sessions, children answered 90% or more items correctly, and in 31% of sessions, they answered 30% or fewer items correctly—suggested that transitions were quite brief. Most transitions took place in Session 1, when 36% of children generated more than 30% but less than 90% correct answers; in Session 2, only 10% of children did, and in Session 3, only 9% did.

The remaining 10% of participants (8 of the 80 children) did not match any of the four learning patterns. These children usually ceased being systematically incorrect, but they never mastered the task. For example, 1 child was correct on 10% of the trials in Session 1, and 50% in Sessions 2 and 3.

As shown in Table 1, the number of nonlearners, slow learners, medium learners, and fast learners was associated with whether children encountered logical explanations, \( \chi^2(3, \ N = 72) = 8.27, \ p < .05, \ V = .34 \). The finding could be interpreted in two ways. One was that whether children did or did not receive logical explanations influenced whether learning ever occurred, \( \chi^2(1, \ N = 72) = 6.55, \ p < .05, \ V = .30 \), but not whether children were fast, medium, or slow learners, \( \chi^2(2, \ N = 50) = 1.79, \ p > .05, \ V = .19 \). The other interpretation was that whether children received logical explanations was associated not only with whether the children learned but with their learning slowly rather than not at all. The percentage of fast learners was identical for children who did and did not receive logical explanations (Table 1). Of the remaining children, the large majority who received logical explanations (67%) learned eventually, whereas far fewer of those who did not receive logical explanations (32%) ever learned.

Variation in paths. Analyses of variations in children’s paths of change involved setting criteria for each hypothesized level of performance—consistently incorrect, chance, and consistently correct—and examining the fit of the criteria to individual children’s performance. The period of below-chance performance was defined as the trials before children responded correctly on two consecutive trials; chance performance was defined as the period in which children had responded correctly on two consecutive trials but not on five consecutive trials; and consistently correct performance was defined as the period following children having met the discovery criterion of five consecutive correct answers.

All children’s responses conformed to one of five patterns. One group of children (\( N = 15 \)) was consistently correct from the outset of the experiment. They met the discovery criterion on the first 5 trials of the first session, and they answered 95% of subsequent problems correctly. \( t(14) = 18.81, \ p < .001, \ d = 5.00 \). A second group of children (\( N = 24 \)
answered consistently incorrectly throughout the experiment. Their percent correct throughout the experiment was well below chance throughout the 30 trials, 6% correct, $t(23) = 23.03$, $p < .001$, $d = 4.65$; none of them met the discovery criterion.

A third group ($N = 11$) progressed through all three hypothesized periods. In the first period, their percent correct was below chance, 12%, $t(10) = 7.03$, $p < .001$, $d = 2.11$; in the second period, their percent correct was at chance, 52%, $t(10) = 0.28$, $p > .05$; and in the third period, they were consistently correct, 89%, $t(10) = 6.73$, $p < .001$, $d = 2.05$. Children in a fourth group ($N = 24$) progressed directly from below chance to consistently correct performance. Their first two consecutive correct answers were immediately followed by three more consecutive correct answers. Before the discovery, they answered correctly on 16% of trials, a figure well below chance, $t(23) = 8.89$, $p < .001$, $d = 1.70$. After the discovery, they answered correctly on 94% of trials, which was well above chance, $t(23) = 18.01$, $p < .001$, $d = 3.67$. It should be noted that the relative sizes of the third and fourth groups depended on the exact criteria for chance responding; if a criterion for a period of chance responding had been at least one correct response but not five consecutive correct responses, the numbers of children in the two groups would have been reversed. Finally, children in the fifth group ($N = 6$) answered the first two questions correctly, then regressed to chance level performance for an average of five trials, 47% correct, $t(5) = 0.29$, $p > .05$, and then, after the discovery, answered consistently correctly, 97% correct, $t(5) = 39.29$, $p < .001$. These children may have begun with the correct approach, regressed to chance level responding, and then answered consistently correctly.

**Changes With Age**

In the cross-sectional part of the study, groups of 20 children of each age between 5 and 10 years were presented a single session of 10 trials without feedback or explanations. As shown in Figure 4a, mean percent correct increased gradually with age—from 30% correct among 5-year-olds to 89% correct among 9- and 10-year-olds, $F(5, 114) = 7.00$, $p < .001$, $\eta^2 = .24$. The percent correct among the youngest children (the 5-year-olds) was well below chance, $t(19) = 3.57$, $p < .01$, whereas that of the oldest children (the 10-year-olds) was well above chance, $t(19) = 6.24$, $p < .001$, $d = 1.39$. Thus, as in previous studies, overall percent correct improved from below chance to about chance to consistently correct in this age range (Figure 4a).

The data on individual performance told quite a different story (Figure 4b). Simply put, at no age did many children appear to guess or to use strategies that were sometimes correct and sometimes incorrect. At all six ages, at least 70% of children answered either no more than 10% or at least 90% of problems correctly. Summed across the six age groups, this was true of 88% of children. Thus, at the level that could be revealed by cross-sectional sampling, transitions between stable below chance performance and stable correct performance appeared to be brief, just as they were in the microgenetic data.

**Discussion**

The data from this study provided strong support for several of the hypotheses that motivated the study. In particular, not only did the 5-year-olds prove capable of learning from logical explanations, they learned more from them than from empirical expla-

![Figure 4](image-url)
nations of the same answer or from feedback alone. On the other hand, children’s learning failed to support other hypotheses, including ones that seemed very likely, for example that learning would be greater if children received feedback than if they did not. In this concluding section, we apply the five aspects of change framework to summarizing the findings and their implications for understanding children’s learning.

Sources of Change

Previous studies have shown that 5-year-olds use empirical strategies to solve many problems that older children and adults solve with logical strategies. The problems on which young children rely on empirical, rather than logical, solutions include tasks that are highly similar to class inclusion, for example being asked whether they can draw more cats than animals or place more spoons than silverware on the table. Conversely, 5-year-olds rely on counting strategies in many situations in which older children and adults would not use them. These findings have been interpreted as meaning that young children have an empirical bias, which would seem to imply that they should learn more effectively from empirical justifications of answers to problems than from logical explanations of the answers.

The observed pattern was exactly the opposite. Exposure to logical explanations led to large improvements in 5-year-olds’ performance on class inclusion problems, much larger improvements than did exposure to empirical explanations for answers to the same problems. Moreover, the benefits of logical explanations were not limited to a few children who were “ready” to learn from them. Instead, the logical explanations led to more of the children learning and to the learning continuing over more trials and sessions.

These findings suggest that 5-year-olds’ reliance on empirical solutions in situations in which older individuals rely on logical approaches is not due to the young children having an empirical bias or to their not being able to understand nonempirical explanations. Instead, the findings suggest that when young children understand that logical solutions are possible, and understand how to apply them, they readily adopt them. Consistent with this interpretation, the 5-year-olds who were exposed to both logical and empirical explanations for the correct answer rated the logical explanations as smarter even before they themselves began to solve the problems correctly.

Why were the logical explanations rated as smarter than the empirical ones, and why did they lead to greater learning? It seems unlikely that inductive generalization from experience using the logical and counting strategies was the reason. The superior ratings of the smartness of the logical explanations came before children made the discovery, often within the first five trials of the experiment, and therefore before children had used either the logical or the empirical approach at all often. Instead, it seems either that the logical explanations enabled children to foresee the advantages in the speed and accuracy of solving the problem logically or that the logical strategy sounded smarter to the children for less pragmatic reasons, such as intellectual elegance. Determining how children judge the relative smartness of newly encountered strategies, and how such judgments influence the rate of discovery of new strategies, should enhance understanding of the strategy acquisition process.

One caveat about these conclusions regarding the effectiveness of the logical and empirical explanations should be noted: the conclusions are based on the logical and empirical explanations that were used in the study. If the particular empirical explanation that the experimenter advanced was less effective relative to the overall class of empirical explanations than the particular logical explanation was relative to the class of logical explanations, the findings might not generalize to other empirical and logical explanations. Contrasting the effectiveness of other empirical and logical explanations can help address this issue.

Rate of Change

Symbolic information processing models often depict children as changing strategies in response to a single incorrect answer (Newell, 1990; van Lehn, 1988). Connectionist models often have the opposite bias, typically depicting strategy change as an extremely slow process that requires thousands of trials (e.g., McClelland, 1995; Rogers & McClelland, 2005). In contrast to both depictions, in the present study, as in the large majority of previous ones, the rate of discovery proceeded on a “human scale”—slower than in the symbolic models, but much faster than in the connectionist models.

The rate of discovery in the present experiment was higher than in most microgenetic studies. This was probably because the explanations provided in three of the four conditions described specific strategies that children could use to solve the problems. In most previous microgenetic studies, children
needed to generate new, superior strategies for themselves.

In addition to being relatively fast, the rate of discovery in the present study was quite variable, both within and across experimental conditions. As shown in Figure 3, many children's discoveries came quickly, but other children's discoveries came considerably later, particularly in the logical explanations condition. This finding suggested two effects of the logical explanations: they were capable of stimulating discoveries in the large majority of 5-year-olds—more than 80%—but they often did not do so quickly. The pattern in the empirical explanations condition was different; there, children either made the discovery in the first five trials or were unlikely to ever make it. The contrasting patterns suggest that the empirical explanations may have been easier to understand but that the logical explanations ultimately evoked greater cognitive change.

Once children met the criterion for discovery of the correct strategy, they answered consistently correctly; this was true in all four conditions. The rapidity with which the new approach was adopted was unusual, probably the fastest that has been documented in a microgenetic study. On the other hand, the rapidity was consistent with previous microgenetic findings that consistent use of a newly discovered strategy generally occurs more quickly when the new strategy is much more accurate than alternatives (Siegler, 2006; Siegler & Svetina, 2002). This was the case in the present task; changing from subset-subset comparisons to subset-set comparisons produces a change from 0% to 100% correct solutions. The extremely fast uptake of the newly discovered strategy was also predicted by the view that children are less likely to revert to less advanced strategies when new, more advanced strategies embody a more advanced logic, as well as having pragmatic advantages (Siegler & Svetina, 2002). Yet a third likely reason was the stringency of the original discovery criterion; needing to generate five consecutive correct responses to be said to have made the discovery, or even three as in the alternative analysis, already demanded a fairly consistent use of the new approach.

Path of Change

In the microgenetic experiment, some children's path of change matched the hypothesized sequence. These children progressed from an initial period of below chance performance to a period of roughly chance performance to a period of stable correct responding. Other children skipped the intermediate period and progressed directly from a systematically incorrect to a systematically correct approach. The relative frequency of the two sequences depended on the exact criterion used to denote the period of chance responding, but by both lenient and stringent criteria, substantial numbers of children showed both sequences. The first type of path—from a stable less advanced approach to an unstable approach involving guessing or oscillation among several strategies to a stable more advanced approach—has often emerged in microgenetic studies. Among the cognitive developmental tasks on which such a path has been documented are basic addition, number conservation, matrix completion, mathematical equality, reaching, kicking, and ramp descent problems (Adolph, 1997; Albali, 1999; Graham & Perry, 1993; Siegler & Jenkins, 1989; Siegler & Svetina, 2002; Spencer et al., 2000; Thelen & Ulrich, 1991; van der Maas & Molenaar, 1996). The second type of path—from a stable less advanced approach to a stable advanced approach—has also emerged in some studies (Albali, 1999; Albali & Goldin-Meadow, 1993; Opfer & Siegler, 2004; Perry, Church, & Goldin-Meadow, 1988). In most of these cases, as in the present study, analyses of individual performance revealed that some children progressed through all three phases, whereas others progressed directly from the least to the most advanced.

Contrary to conclusions from previous studies, the cross-sectional segment of the present study provided little evidence for an intermediate period of instability between the period of systematically incorrect responding and the period of systematically correct responding. At all ages, a large majority of children answered either systematically incorrectly or systematically correctly. The percentage of children who followed the two approaches changed gradually with age, yielding the smoothly increasing percentage of correct answers that has been typical in previous studies of class inclusion. However, examination of individual performance suggested that this apparent gradual change was an artifact of averaging over children. These data would appear to argue against the position (Chapman & McBride, 1992; Hodkin, 1987; McCabe et al., 1982) that a period of guessing or oscillation between correct and incorrect approaches appears between periods of systematically incorrect and consistently correct responding.

It remains possible that children in the cross-sectional segment of the study did progress through a brief period of guessing or oscillation between incorrect and correct approaches. Such a transitional period might have been too brief to detect in a cross-
sectional sample in which successive age groups differed on average by a year. Alternatively, receiving feedback that indicated that one’s existing approach yielded incorrect answers might have provoked a period of guessing or oscillation in some children that does not exist without feedback that disconfirms one’s usual (incorrect) approach. One way to test whether there is a period of oscillation in the absence of feedback would be to present children with class inclusion problems on several days close in time. The ideal age group for such a comparison would be 6- and 7-year-olds, the age groups that generate the most equal balance of systematically correct and systematically incorrect answers. If many children answer systematically correctly on some days and systematically incorrectly on other days, it would provide evidence of a period of instability of class inclusion strategies that was not provoked by feedback.

Breadth of Change

The present study included 5-year-olds’ and college students’ evaluations of the smartness of alternative strategies for answering class inclusion problems. This provided a measure of conceptual understanding of the logical and empirical explanations and thus broadened the knowledge assessment beyond the measures of strategy use.

As anticipated, the 5-year-olds, like adults, evaluated the logical strategy as being smarter than the empirical strategy, despite the fact that both approaches would always yield correct outcomes if executed correctly. The very high absolute evaluations of the smartness of the logical explanations before the children answered class inclusion problems consistently correctly, together with the fact that these evaluations were significantly higher than those of the empirical explanations, indicate that the children possessed conceptual understanding of class inclusion that was not apparent in their initial strategy use. The evaluations also suggested that the children understood that there was something appealing about the logical strategy before they understood the strategy well enough to use it themselves. This sense that there was something appealing about the strategy may have motivated children to continue thinking about the logical explanations long enough to learn from them. The sustained number of trials over which children learned from repetitions of the logical explanations, and the fact that this period was considerably longer than the period over which children benefited from repetitions of the empirical explanations, is consistent with this interpretation.

More generally, the process of discovery may often begin with children having a vague sense that a strategy they encounter or generate is a good one, without much additional understanding of why it is good or even how to execute it. Such intuitions may guide children toward attending to, and thinking about, such strategies long enough to understand and begin to use them. This idea is related to previous proposals regarding skeletal principles (Gelman & Williams, 1998; Keil, 1998), utilization deficiencies (Miller & Seier, 1994), and goal sketches (Siegel & Crowley, 1994). In all cases, the core idea is that children at times have a rudimentary understanding of the goals that new strategies could serve and their potential advantages over previous strategies, which increase the children’s attention to the goal of generating new strategies and lead children to persist in using them even if the strategies do not immediately pay off in faster and more accurate performance.

It would have been revealing to know how the young children reached the conclusion that the logical explanations were smarter than the empirical ones. Unfortunately, their explanations of their ratings were uninformative. Prior findings regarding 5-year-olds’ evaluations of the smartness of arithmetic strategies (Siegel & Crowley, 1994) revealed similar inarticulateness regarding the reasoning underlying their evaluations of strategies. It was unclear whether a different phrasing of the question would have elicited more informative explanations, whether 5-year-olds are unable to access the reasoning underlying their strategy evaluations, or whether they lack the vocabulary to describe their reasoning.

Variability of Change

Microgenetic studies allow the examination of individual differences in learning patterns. As shown in Table 1, in the present study about 30% of children learned extremely quickly, answering at least 90% correctly in all three sessions. These fast learners were distributed roughly evenly among the four experimental conditions. About half of these children probably already knew how to solve class inclusion problems coming into the experiment, like the 15% of 5-year-olds who answered correctly on at least 90% of trials in the cross-sectional sample despite receiving neither feedback nor explanations. Others may not have known the solutions beforehand but learned extremely quickly.
Another roughly 30% of children answered consistently correctly during both the second and third sessions, but not the first. These were primarily children who received the logical explanations. A third, smaller group of children, roughly 10%, answered consistently correctly by the third session; again, these tended to be children who received logical explanations.

The fourth group of children, about 30%, never learned to answer consistently correctly. Most of these children answered no more than 30% of problems correctly in any of the three sessions. Despite 30 trials of feedback, each of which indicated that the number of entities in the larger subset did not exceed the number in the entire set, these children continued to claim that it did. Such resistance to feedback, in which children not only fail to learn but also continue to base responses on a consistent incorrect rule or strategy, is surprisingly common among 5-year-olds. Consider some results from previous studies. Many 4- and 5-year-olds continued to rely solely on weight to solve balance scale problems despite encountering feedback on which the weight rule led to consistently wrong answers (Siegler, 1976; Siegler & Chen, 1998). Many 5-year-olds continued to rely on the height of the liquid columns to solve conservation of liquid quantity problems and on the relative stopping points of two trains running in the same direction to solve time problems, despite receiving numerous feedback problems on which these approaches led to incorrect answers (Siegler, 1983, 1996). Perhaps the most extreme example comes from Siegler and Svetina (2002), in which most 5-year-olds continued to use the same incorrect approach after feedback on 88 matrix completion problems. Older children and adults are generally much quicker to drop incorrect approaches and to try new ones (Siegler, 1996).

Why do 5-year-olds cling to rules and strategies that have been repeatedly disconfirmed? One reason that has emerged in a number of previous studies is failure to encode relevant dimensions (Kaiser, McCloskey, & Proffitt, 1986; Siegler, 1976, 1983). In the present situation, children who consistently answered incorrectly despite disconfirming feedback may not have encoded the problems in terms of abstract set/subset relations. Instead, they may simply have encoded each problem as being about the entities named in it: dogs, cats, and animals; boys, girls, and children; tulips, daisies, and flowers; and so on. Failing to encode the abstract structure that was common to all the problems may have prevented these children from learning from the feedback. The encoding interpretation is consistent with Markman and Seibert's (1976) hypothesis that presenting collection nouns improves class inclusion performance because it calls attention to the inclusion relations that young children otherwise fail to notice. Helping young children to encode the common structural relations among items, rather than the details of the individual items, may improve the children's learning from logic-based explanations and from feedback not only on class inclusion but on a wide range of problems.

Analyses of individual children’s paths of change converged with the analyses of individual children’s rates of change in showing a subgroup of children who knew how to solve the problems from the outset and a subgroup of children who never learned how to solve them. The analyses of individual paths of change added the information that some children progressed through all three hypothesized knowledge states, that others progressed directly from the least to the most advanced approach, and that a few others may have regressed from correct to chance level responding before progressing to consistently correct performance. This variability was not evident in group-level analyses, which suggested that the three-level progression was more prevalent than the individual analyses revealed them to be. The general message is that examining variability in individual children’s learning, as well as group-level patterns, is essential for understanding the process of change.

References


