Cognitive Development, 1, 1-29 (1986)

Young Children's Numerical Competence

Rochel Gelman
Elizabeth Meck
University of Pennsylvania

Susan Merkin
Teachers College, Columbia University

ABSTRACT

In their analysis of numerical competence, Greeno, Riley, and Gelman (1984) distinguish between conceptual, procedural, and utilization competence. Principled knowledge about a domain, for example, counting, serves as the basis of conceptual competence. Conceptual competence does not provide recipes for procedures but does set constraints on the class of procedures that procedural competence can generate. The ability to assess a task correctly (utilization competence) influences performance because it, too, sets constraints on procedure generation. These distinctions allow one to classify the source of erroneous or variable performances on different number tasks. Our hypothesis that early counting behavior is guided by counting principles, despite the child's limited skill, is tested in four experiments with children ranging in age from 3 to 5 years old. The experiments focus on knowledge of the order-irrelevance and cardinal count principles; children either discriminate between erroneous and correct counting efforts of a puppet, assess the effect of counting the same array in different orders, or solve a counting task that has a constraint. The results of these studies allow us to reinterpret evidence others cite against the principle-first hypothesis, and conclude that much development takes place in the name of procedural and utilization competence.

The research reported here was supported by NSF grant BNS8004881 to Rochel Gelman, who prepared this version of the manuscript while she was a Fellow at the Center for Advanced Study in the Behavioral Sciences. Fellowship support came from an NIHCD Senior Fellowship, the Spencer Foundation, and the Alfred P. Sloan Foundation. We thank the many parents and children who participated in the studies and the teachers and school administrators who facilitated our work at their schools. R. G. offers special thanks to Randy Gallistel and Zella Luria who influenced the preparation of the manuscript in a fundamental way; Ann Brown and Jim Greeno who read critically earlier versions of this work; and Muriel Bell for her editorial wisdom.

Correspondence and reprint requests should be sent to Rochel Gelman at 3815 Walnut St., Philadelphia, PA 19104.

Manuscript received May 23, 1985; revision accepted July 11, 1985

There is considerable evidence that preschool children can count (e.g., Fuson & Hall, 1983; Gelman & Gallistel, 1978; Schaeffer, Eggelston, & Scott, 1974). Two kinds of models have been used to explain this early development of counting. One proposes that skill acquisition is guided by an implicit understanding of the principles that underlie the ability to count and represent the cardinal value of a set (e.g., Gelman & Gallistel, 1978; Greeno, Riley, & Gelman, 1984). The other proposes that children lack initial understanding and learn, in a piecemeal way, various component counting skills, such as reciting the count list, habitually repeating the last tag in the list, pointing to objects, etc. This happens presumably as a function of their being reinforced (either extrinsically or instrinsically) for repeating in a rote manner response patterns they encounter in their environment. As more and more of these accrue habit strength, children abstract the generalizations common to them and hence come to a principled understanding (e.g., Baroody, 1984, Briars & Siegler, 1984; Fuson & Hall, 1983).

In support of their principle-first position, Gelman (1982) and Greeno et al. (1984) pointed to three lines of evidence. The first concerns the difference in attributions of competence that result when success measures vary in their stringency requirements. For example, if one requires that 3-year-olds *always* use as many count words (N) as there are items to count (X), fewer than 40% who attempt to count a set size of seven succeed; in contrast, if one scores the same children's use of ($X=N\pm1$) count words, 70% "succeed" on set sizes of 7, 9, and 11. To account for the discrepancy, Gelman and Gallistel (1978) appeal to the notion of performance demands, including in this case the ones involved in the ability to terminate simultaneously the reciting of tags and the transfer of items from the to-be-counted to the already-counted class of objects. Gelman and Meck's (1983) demonstration that children's ability to count three-dimensional objects was better when they could touch and move them—and hence keep track of what they had counted—supports this interpretation. So does Wilkinson's (1984) formal analysis of further cases of variable behavior, behavior that, as he shows, reflects at least partial knowledge of the underlying principles.

Second, Gelman and Meck (1983) reasoned that children who implicitly understood the counting principles should be able to detect errors in their application, just as speakers of a language detect word strings that violate implicit knowledge of the rules of syntax. In three separate studies, preschool children detected errors in the application of the one-one principle; the stable-order principle (the list of tags used over trials must be stably ordered); and the cardinal principle (the last tag in a count has the special status of representing the cardinal value of the set).

A third line of evidence comes from the ability of children to deal with novel tasks, for example tasks that require the invention of solutions that honor the counting principles. Children could not do this if they did not have implicit knowledge of the principles, anymore than speakers of the language could generate novel utterances in the absence of implicit knowledge of linguistically relevant rules (Greeno et al., 1984). Gelman and Gallistel asked children to count a display of objects so as to tag a particular object, for example, the second in a row, with a particular tag, (e.g., "five"). In this case, the child cannot count left to right and take advantage of the usual correspondence between order of tags in a list and order of items. Children could move items so as to create a correspondence or they could skip the second item when counting left to right and then return to it to finish. In either case, they had to count in a novel way to avoid violating the counting principles. Of the 3-, 4- and 5-year-olds in the study, 43%, 73%, and 100% succeeded on at least two of their ten trials. Only 12% and 44% of the 3- and 4-year-olds did so on all their trials; in contrast 75% of the 5-year-olds did so on all trials.

A common feature of most of the evidence cited by Gelman (1982) and Greeno et al. (1984) is that children are not perfect; they err, and the younger they are, the more often they do so. It is exactly this feature that is highlighted by proponents of the competing account—that learning of rote procedures precedes principled understanding. The view is that there is no reason to grant young children principled understanding if under a variety of conditions they are not competent. Thus, for example, Briars and Siegler (1984) emphasize that, despite the young child's ability to detect counting errors, some kinds of errors are harder than others. Fuson and Hall (1983) point to the limited ability of 2-year-olds to coordinate their pointing to objects with their recitation of tags.

Fuson and Hall also draw attention to the fact that when children are first asked to count a set, and then answer a "How many?" question, there is no effect of set size. (See also Ginsburg & Russell, 1981.) There are children who always repeat the last tag, even on set sizes they cannot count accurately, and there are children who never do. Fuson and Hall conclude this means young children are using a cardinal "rule" but not a cardinal "principle." The argument is that the child has a collection of unintegrated procedures and hence the "cardinal rule" is not meaningful. It is a rule to "repeat what you last said" but not a rule to determine the cardinal consequence of the counting that precedes its use. Baroody offers a similar arguement in a discussion of the child's understanding of the order-irrelevance principle—alterations in the order in which items in an array are tagged do not change cardinal value. He concludes that success on an easier version of a task in conjunction with failure on a harder version justifies granting the child a procedural rule, but not implicit understanding of the order-irrelevance principle. The position is that children may know that the order of counting is irrelevant but not recognize the implication that whatever the order in which items are counted, the cardinal value remains the same.

We will argue that variable performance, either within one task or across related tasks, does not by itself provide clear evidence in support of the rote-learning position. Indeed, we will show that variability is predicted by the principle-first position, this because the principles do not themselves serve as recipes for procedures. They simply set constraints on what form those procedures can take. We will also argue that the counter evidence offered by Baroody and others has an alternative explanation, one that leaves intact the principle-first position. To do this it is necessary to have a systematic way to consider the difference between principle and procedure. Following Greeno et al. (1984), we distinguish between *conceptual, procedural,* and *utilization* competence.

In the Greeno et al. model of counting, conceptual-competence characterizes knowledge of the counting principles; utilization-competence deals with the ability to assess the performance requirements of a particular task in light of the constraints imposed by conceptual competence; procedural competence deals with the ability to generate behaviors that are congruent with the requirements of conceptual competence and relevant task variables. Conceptual and utilization competence refer conditions to a planner that determines which procedures can be generated in accord with these conditions. Successful performance then involves the generation of procedures that satisfy the constraints imposed by the counting principles and a given task setting.

In distinguishing between procedural and conceptual competence, we introduce the idea that the counting principles represent a set of abstract constraints on procedures. Only those procedures which satisfy these constraints are recognized as instances of counting. Principles do not by themselves determine the generation of procedures. Instead the counting principles are represented in the model by specifications which define the classes of acceptable behaviors and number-specific goal structures, much as do rewrite rules in linguistics. For example, the schemata MATCH and KEEP-EQUAL-INCREASE specify the properties a procedure must exhibit if it is to be consistent with the one-one principle. Since the MATCH schema takes two sets as arguments, the sets can be some particular items to count and the words in the conventional count list. The sets could also be words, abstract thoughts, etc. on the one hand, and the alphabet or hatch marks on the other hand. KEEP-EQUAL-INCREASE requires that the two sets remain equal throughout a count and indicates a general way of keeping these sets equal; the idea is that when starting with two empty sets, the sets remain equal if each is increased by a single member. These schemata together capture the fact that the child will have to generate procedures to partition to-be-counted and already-counted items, and coordinate the use of count words and the partitioning of items. But they do not dictate how this will happen. The characterization of the principles in terms of abstract action schemata makes it possible to detail the constraints or requisites involved in their use and formally describe the class of acceptable procedures. But it is the work of procedural competence to generate the particular instances of this class.

Although principles do not themselves generate counting procedures, they still play an indispensible role in the generation of such procedures: The constraints determined by the principles are referred to the planning or executive part of the system, the part that must produce competent or effective behaviors. The

generation of behavior requires a planning or procedural competence, the ability to set goals and select action schemata that meet the prerequisite, requisite, and postrequisite conditions given to the planner. The quality of the planner sets the limits on procedural competence since it determines which procedures can be generated, how to use a procedure, etc. The planner has to take a number-relevant goal, for example, determine the cardinal value of the set, and then assemble action schemata that check to see whether conditions are satisified in the situation, and search for ways to achieve conditions that are not already present.

The generation of correct counting also depends on utilization competence, the ability to assess a particular setting in terms of the task demands relevant to putting a plan of action into effect. The items in front of a child could be arranged in a circle, in which case the child must recognize a counting-relevant feature of the task, that is the beginning and/or end of the set must be clearly marked so as to avoid double-counting items and hence violating the one-one principle. Since a child might know that he/she is not supposed to double-count, yet lack the procedural resources to avoid doing so, a distinction should be drawn between these abilities and conceptual competence. To recognize the problem, the child has to appreciate the difference between displays that have a clear beginning and end and those that do not; but this is a problem in perception and is not unique to counting. Deficiencies in perception of groupings will influence counting whether or not the child possesses the requisite conceptual competence. Hence, the Greeno et al. argument that they are best represented as deficiencies in utilization and/or production competence. Similarly, general biases to misinterpret instructions, memory problems, etc., will influence performance because of the limits they place on the child's utilization skills at the moment.

But even if the child has the wherewithal to assess all the requirements of a task setting and correctly interpret instructions, there is no guarantee that success will follow. In the case at hand, the child also has to generate an acceptable counting procedure. As indicated above this might involve moving items, putting a piece of paper on an item, etc. None of these skills is counting-specific. They could be used to improve the perception and memory of other materials as well. Yet, their absence in the present particular case would be crucial. For accurate performance to occur, at the very least, children have to refer their assessment of the task requirements to the planner which has to coordinate these with the constraints given it by its conceptual competence.

The above analysis leads to a renewed defense of the "competence-performance" distinction in developmental accounts. Indeed, since principles do not spell out how conceptual competence will be put into practice, it follows that there must be a distinction between "competence" and "performance." The terms "principled" or "conceptual competence" substitute for what is routinely taken as "competence." However, since conceptual competence is characterized in terms of the kinds of requirements a procedure must meet, it is not a passive structure sitting in the head somewhere. It offers guidelines as to how to connect competence to performance.

Our discussion of performance is much more complex than ones that simply consider limits on memory, attentional capacities, or processing pace (e.g., Case, 1984; Chomsky, 1957). It requires considering the ability to plan actions that are consistent with the constraints set down by the principles. The principles do not guarantee the child will set the relevant goals, this must be part of the planning activity. Further, a consideration of performance requires considering the separate ability to assess task settings and relate them to these constraints. Making a distinction between procedural and utilization competence allows us to begin to classify in a systematic way the different kinds of variables which will influence the selection and production of actions and the consequent performance.

Return to the evidence which challenges the principle-first view, for example Fuson and Hall's demonstration that children fail to relate their answers to "how many questions" to their prior counting activities. How to reconcile this with evidence that there are times when young children do achieve a meaningful cardinal representation of a set (Greeno et al., 1984), for example, when solving simple arithmetic problems? A procedural competence failure to integrate the goal of obtaining the cardinal value of the set with the application of the one-one and stable-order principles could lead a child to execute counting-relevant behaviors which are not suitably integrated with respect to this goal. Perhaps this is what happened in the Ginsburg and Russell, and Fuson and Hall assessments of the child's understanding of cardinality. If so, whether the results reflect a deficit in procedural or conceptual competence is not clear.

By distinguishing between utilization and procedural competence, we can shed light on the range of variables that fall under the rubric of performance demands. A child could full well set the goal to count the set, generate error-free counting, and yet fail a task by neglecting some feature(s) of the task requirements. Consider the constrained counting task above, where the child is told to "count these but make this (some item other than the first in a row) *the one*." The child who neglects to assess the situation, and simply counts left to right will fail this task, not because she does not understand the counting principles, but because of some limit in her utilization competence. But assessing the situation correctly does not guarantee that success will follow. For now the child may have to use novel procedures, for example, skipping the first item and then returning to it. Smith and Greeno (1984) show that success or failure on the constrained counting task can be dealt with by altering the Greeno et al. models of utilization and procedural competence *without changing that for conceptual competence*. This makes clear why failure on a task need not reflect a deficit in principled understanding.

The analysis of the competence-performance distinction presented here provides a tool with which to determine what contributes to a child's variable or

wrong performance on counting tasks. We use this tool to consider the apparent contradictory results reported above. In what follows we present experiments designed to separate problems children might have with utilization and procedural competence from those they might have with conceptual competence. We begin with a pair of studies that examine the hypothesis that the children in Baroody's test of the order-irrelevance principle failed because of utilization competence limits, ones that led them to misinterpret instructions. This study also shows that, when young children do set a goal of obtaining the cardinal value of a set, they coordinate their knowledge about the component counting acts. In the next pair of studies we assess Smith and Greeno's conclusion that failure on the constrained counting task can be attributed to difficulties young children might have with utilization and procedural competence. We conclude that the younger the child, the more likely they are to have problems with both procedural and utilization competence. A consideration of the utilization variables that most affect their behavior highlights the young child's dependence on supporting social variables (Brown & Reeve, in press; Donaldson, 1978).

EXPERIMENT: ORDER-IRRELEVANCE AND CARDINALITY

Initial Study

Elsewhere (Gelman & Meck, in press) we provide evidence that the discrepancy between the results reported by Briars and Siegler and those reported by Gelman and Meck derives, in part, from the way children interpreted novel error-detection trials in the two studies. We proposed that details of instructions as well as our use of an interactive testing procedure helped children realize that the trials we called *pseudoerrors*, and Briars and Siegler called *unusual*, were ambiguous. On these trials puppets started counting at some place other than the beginning of a row; hence they were not conventional, although they were correct since all items were eventually counted. Thus children had to decide whether to respond on the basis of their knowledge of convention or set this aside and respond only on the basis of the principle. Children in Gelman and Meck (in press) did better on such trials when given a second chance or when probed. Some even told us "It's a silly but OK way" to count. Our account of the difference between their immediate and best-trial responses was that our instructions, which suggested they knew more than the puppet, and an interactive testing procedure helped children correctly interpret these trials. The conclusion was that the children's problem was in the domain of utilization, not conceptual competence. The next experiments provide further evidence for this hypothesis by showing how subtle variations in question format lead children to do either very well or very poorly on an order-irrelevance task.

Gelman and Gallistel's order-irrelevance principle captures the fact that one can apply the how-to-count principles (one-one, stable, and cardinal) in any order to a set of objects. The consequence of this is that regardless of the order, the same cardinal value obtains. Since children in the Gelman and Meck (in press) error-detection study judged the pseudoerror trials correct and even, at times, gave accounts of why, one might conclude they understood the order-irrelevance principle. Baroody (1984) cautions against this, preferring instead to conclude that the children could be indifferent to the order in which items in an array are tagged and yet not believe that two different count orders should yield the same cardinal value. In Baroody's study, children were shown a row of eight items and asked to first count left to right and then indicate the cardinal value of the set. Then, while the experimenter pointed to the right most item, they were asked "Could you make this number one?" Children typically said they could and were then told "We got N (N stands for the child's cardinal value) counting this (first) way, what do you think we would get counting the other way?" The array was then covered to discourage children from counting. The large majority of 5-year-olds responded with some value other than N. Only older children were correct.

Again we introduce the possibility that failure on a counting task reflects an erroneous assessment of the task demands, not a lack of principled understanding. In this case we focus on social features of the task and suggest that the younger children took the Baroody instructions as a challenge that implied their first answer was wrong, especially since the children had had but one counting trial and could have been unsure of their answer in the first place. Donaldson (1978) provides many demonstrations of how communication factors lead the very young to fail, apparently because they lack confidence and/or the ability to assess the real intent of an ambiguous message (see also Shatz, 1983).

To test our hypothesis we ran children in three independent groups. The first group (Baroody) participated in a replication of the Baroody experiment. The second group (Count 3X), was first given three opportunities to count the array—on one of which they were asked for its cardinal value. The rest of the procedure was the same as Baroody's. Perhaps these children would be more confident they knew the correct value of the set size. In the final group (Altered-Question), children started out as did those in the Baroody condition by counting and answering the cardinal question but once. However, then the experimenter pointed to the last item and said "Can you start counting with N?" (N was the child's cardinal value), "How many will be there?" or "What will you get?" Note the absence of a comment that asked children to focus on the answer they gave last time and consider what they would get this time. Such contrasts might serve as a challenge or suggest to children they should hedge (cf. Labov & Fanshel, 1976). It is known that preschool children respond to and produce linguistic devices of this sort (Shatz, 1983); thus we reasoned that, despite the seeming minor variation in instructions between the original Baroody condition and the Altered-Question condition, children would interpret the task correctly in the latter but not the former.

We reasoned further that should we obtain the expected differences, following Baroody's (1984) suggestion, it would be good to have additional evidence to validate our conclusion. Hence all three groups of children were subsequently tested on two kinds of error-detection tasks that focused on their knowledge of the cardinal principle. One of these was a "trick" task; they had to infer that the puppet who answered N+1 to the "How many?" question following a second count of the same set must have made a counting error even though it was not seen. Otherwise the cardinal value could not have changed. The other task demonstrated variations in correct and incorrect trials regarding the cardinal principle.

Including these last two tasks allowed us to do more than validate procedures. Should children succeed on them, we will have evidence that it is not always that they simply monitor learned procedures in error-detection experiments; at least in this situation they would have to have integrated their counting behavior with their assumptions about the cardinal value of the set. For example, should children tell us that the puppet had to make a mistake in the Trick trial, it is less likely that they have a cardinal rule which simply says repeat the last tag no matter what the true value of the set. Our suggestion that the Ginsburg and Russell (1981) and Fuson and Hall (1983) results reflect a failure in procedural competence, that is, a failure to integrate the goal of achieving the cardinal value of the set, whould gain support.

Subjects and Procedures. We first tried to replicate the Baroody results using a sample of 5-year-old children attending daycare centers in the Greater Philadelphia area. These centers serve a basically middle-class population which is heterogenous with respect to ethnic background, race, and number of parents (one or two) that children live at home with. Children made too few errors compared to Baroody's 5-year-olds. Since Baroody's sample was rural, instead we thought it might be more appropriate to use sample of urban 4-year-olds. A pilot study confirmed our guess and hence 36 4-year-old children, drawn from the described sample, served as subjects in the experiment. So did a group of 12 5-year-olds whom we ran for normative purposes.

The younger children were randomly assigned to the three groups of 12 Ss each. Their mean ages were 54.6 months (51–58 months), 54 months (51–59 months) and 54.5 months (48–59) months for the Baroody, Count 3X, and Altered-Question conditions. The mean age of the 5-year-olds was 65 months.

Phase 1. In all conditions we used a set of eight objects as had Baroody. The children in the Baroody condition were tested as their counterparts had been in the experiment in question (details were worked out from the published account and from consultation with the author). The children in the other two conditions were tested as above.

Error-Detection. Every trial of the phase ended with the puppet answering a "How many?" question. The instructions to children were as follows:

> This is my friend Mr. Horse (Lion), and he would like you to help him in playing the game. Mr. Horse is gong to count the things on the table but Mr. Horse is just learning to count and sometimes he makes mistakes. Sometimes he counts in ways that are OK but sometimes he counts in ways that are not OK and that are wrong. It is your job to tell him if it was OK to count the way he did or not OK. After he is all done counting, he is going to tell us how many there are. It's your job to to tell him whether he was right or if he was wrong and made a mistake. Now remember, you have to wait until he is all done counting and has told us how many there are before you tell him whether he was right or wrong.

The children were asked to make judgments of counting trials with linear, heterogeneous displays of 5, 7, and 10 trinkets. There were three trial types and two examples of each. On the Correct trials, the puppet counted the array from one end to the other and responded correctly to the "How many?" question. On the one-one trials, the puppet either skipped or double counted an item and then answered the cardinal question by repeating the last tag generated on the trial (e.g., "6" for a double-count trial with five items). On the Cardinal-Error trials, the puppet made no one-one errors; instead his cardinal answer was either two more or two less than it should have been (e.g., "12" for a set size of ten). All problems within a set size were run in a block. Their order within a block was random as was that in which the different set size blocks were presented. Children were asked to explain each judgment they made.

The Trick Trial. In this task the child was shown an array of seven items and again asked to judge the correctness of the puppet's response to the "How many?" question. The puppet first counted and responded to the latter question correctly. He then counted the array again, this time starting on the other side. This time, however, he surreptitiously made a one-one error counting, "1, 2, 3, 4, 5, 6, 7, 8" and then repeated the last tag (8) when asked "How many?" To end this phase, children were asked to justify their second cardinal response.

Results and Discussion

Phase 1. Scores were based on whether a child gave the same or different answers to the two target "How many?" questions. Children were scored correct if they gave the same answer to both. Figure 1 shows the results for the three groups of 4-year-olds. Only 1 child (of 12) in the Baroody condition was correct; all others changed the value of the cardinal number they gave the second time. Children in the Count group benefited somewhat from their prior counting experience; still, half changed their answer. In contrast, there was little tendency to do so in the Altered-Question group: 10 of 12 gave the same answer they had initially, the same number as answered correctly in our 5-year-old Baroody condition. Children's justifications of their second answer corroborates these

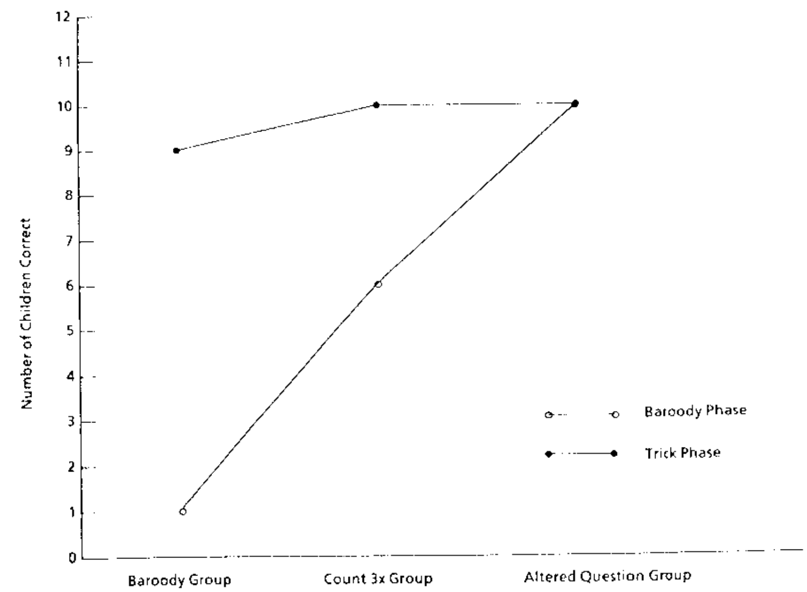

Figure 1. Number of children (n=12) who passed the Barоody Replication and trick phases of the initial experiment. The first phase compares children in Baroody's original condition with those in two new conditions.

differences. Those who did not change their answer could say the cardinal value had to be the same. Protocols of children in the Baroody group suggest they were confused and thought they were supposed to change their answer; their modal response to the second cardinal question was N±1. The following protocols illustrate the difference between the kinds of response styles.

(*G.J*) [53 months; Altered-Question condition.] (Could you start from this side and count that way?) Yes. [S. starts and E. stops S.] (How many are going to be there?) Eight. (How come eight?) Because that way there was eight and there's no way you can try and change numbers. (Why not?) Watch. [S counts] (How come you knew there would be eight?) Because I knew that. (How did you know?) Because I'm smart. (But you knew even before you counted them. How did you . . . How could you change the number? What would you have to do?) You would have to put more things on the table or take things away.

(*A.G.*) [52 months; Altered-Question condition.] Still eight. (How come still eight?) Because they are not going to move. (How did you know without even counting that there would be eight?) Because you didn't take none away.

(*A.R.*) [54 months; Baroody condition.] Seven. (How come seven?)[no answer]. (How come?) Because this fish (S points to what is now the first item]. (What about the fish?) He has to be in the game.

Trick Trial. Figure 1 also shows the number of children in each of these groups who said the puppet who changed his answer the second time he counted the same display was wrong. Nine of the 12 4-year-old children in the Baroody condition caught the error; 10 in each of the other two groups did likewise. Eleven of the 5-year-old Baroody control children did too.

One might object that children who did well with the trick must have seen the puppet's error. To determine whether this was so, the experimenter kept track of children's verbal and nonverbal reactions during the experiment. We coded what children said for clear statements that something odd had happened. Examples are provided for E.O. (51 months) and G.T. (58 months), respectively.

(*E.O.*) Nine this way and eight that way. . . .I thought both of them were going to be the same. . . .I think he made a mistake. I think he counted two times.

(*G.T.*) No. Supposed to be eight. (How come?. . . How do you know there's supposed to be eight?) Because he counted this way [Referring to the previous count]. (Did he make a mistake?). Yes. (What was his mistake?) I don't know. [This same child said "He forgot to count this one" during the Error-Detection phase, illustrating that his failure to describe the trick cannot be due to a general inability to detect and describe errors.]

Using a combination of notes taken during the experiment, what children said during this phase, and a contrast between what they could say about errors they actually saw during the Error-Detection phase with what they did not say in this phase, we determined that the trick was just that. All but two of the 4-year-old children who were right gave verbal or nonverbal evidence of treating this phase very differently from the Error-Detection one; in fact, 76% of them gave the kind of salient verbal evidence portrayed in the above protocols. Thus the trick was successfully disguised and children's detection of an error depended on the puppet's wrong answer to the "How many?" question. It was based on the discrepancy between the result obtained on the first count and the result obtained on the second, reverse count. The children would be bothered by this discrepancy *only if they believed the two counts should yield the same result.*

Error-Detection. Children's responses were scored twice; once on the basis of their immediate answer to the question of whether the puppet was right or wrong, and once on the basis of their best response on that trial. The reason there could be a difference between these two scores follows from the fact that the testing procedure was interactive and occasionally the experimenter repeated the question. To get credit on the best-score analysis, children had to justify their answer; otherwise they were given the same score they received to start. By requiring independent evidence that they knew why they changed their answer, we rule out children's getting credit for the wrong reasons.

Children did very well on this task. The average number of immediate correct responses (out of 12) were 10.3, 11.0, and 10.75 for the Baroody, Count, and Altered-Question groups. Their best response scores were only slightly higher,

10.4, 11.1, and 10.8. A one-way ANOVA was nonsignificant ($F(1,33) = .60$), indicating there were no overall differences in numerical ability between the groups—at least as assessed by the task.

Figure 2 summarizes the results as a function of trial type and set size. The advantage gained on best-scores analyses are slight and are indicated at the tops of the bar graphs.

There was no effect of set size. When children did err they tended to do so on the one-one problems. Indeed of the 19 of 36 children who made at least one error, 17 found this kind of problem the hardest. Figure 2 suggests no reliable differences between the other two classes of errors, and there were not.

The results of the Trick and Error-Detection phase validate our interpretation of the differences between the groups during Phase 1 of the experiment; failure on the Baroody task reflects problems young children have assessing the demands of a task and not their understanding of the order-irrelevance principle. If the latter were in question, they should have failed the Trick trial but they did not.

Trick-trial and Error-Detection data also support our view that young children can integrate the act of repeating the last tag in a sequence with the way they count to achieve a representation of cardinality. Otherwise they could not have passed these phases. Although one-one trials during the Error-Detection phase were hardest for the children, they nevertheless achieved an overall success level

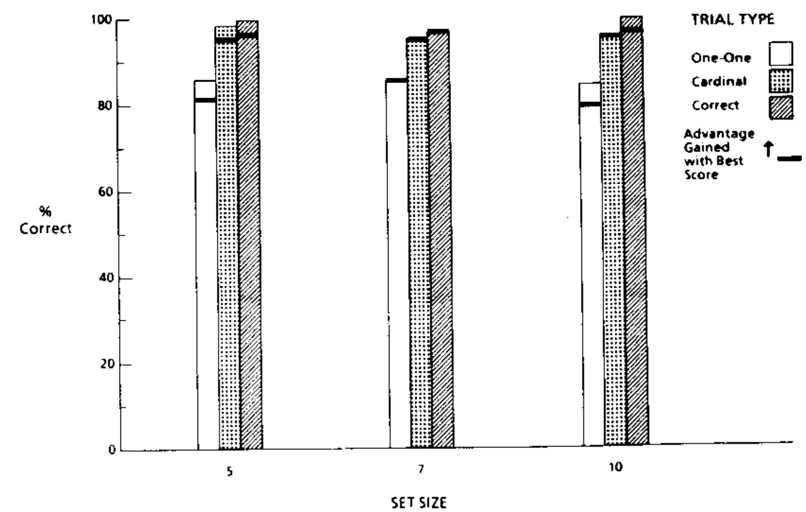

Figure 2. Percent correct judgments as a function of set size, trial type, and criterion (immediate vs. best) during error detection phase of experiment which was based on Baroody (1984).

of 82% (immediate responses). If they were simply monitoring whether the puppet repeated the last tag, they should have erred on all of these trials. Further, their justifications rule out the interpretation that they decided these trials were wrong simply because the puppet made a one-one error.

Children's accounts of why the puppet erred on one-one cardinal error trials included corrections of the puppet's answer to the "How many?" questions, statements that one needed to add or subtract an item in order to accept the answer, or complaints that the puppet had not used a long enough list. These were in addition to their descriptions of the one-one errors themselves, as is illustrated in the following protocols.

(*P.J.*) [48 months, responding to a puppet having skipped an item on a 7-item set and then saying there were six items.] He missed the cow. [S nods No to indicate an error.] (He missed the cow? But he counted six and he said six). . . No. He had to go 1, 2, 3, 4, 5, 6, 7. And that's seven [pointing to the display] because all these animals are out (here on the table).

(*S.Z.*) [57 months in response to the puppet's double-counts on a 10-item set.) No. Ten. He went this and this and this and this and then all the way back to it again and said 11. You are not allowed to change the number when it is one number and that is right. (Why can't you change it?) Because it is supposed to be the same number. Only if you take some away you would have to count it.

Of the 36 4-year-old children in the experiment, 30 talked about the one-one errors they saw on at least one of their trials. Additionally, 28 made explicit reference at least once to the fact that the resulting count sequence and/or the consequent cardinal number was wrong on such trials; 21 of these corrected the puppet. Finally, as can be seen in P.J.'s protocol, the experimenter sometimes challenged children and suggested that the trial was right as long the puppet simply repeated the last tag in the particular list used on the error trial. In all, 17 children were challenged; 14 of them resisted the challenge as did P.J. Given these details about the protocols, we conclude that the children used expectations they had regarding the cardinal value of the display to reach conclusions that simply repeating the last item in a list is not necessarily the cardinal value of the set in question. In other words, they did honor the cardinal principle and since they did, they have the requisite conceptual competence to apply the order-irrelevance principle.

Individual Analyses. To determine whether the above group results reflect individual patterns of responses, we scored children as having passed a phase or not. This was straightforward for the Baroody and Trick phases; a child was either right or wrong on the critical question. For the Error-Detection phase, we required that a child obtain 11/12 correct trials; (a less stringent requirement of 10/12 yielded the same outcomes). Cochran-Q tests were sensitive to the different levels of success for those children who were in the Baroody control condition during Phase 1 ($Q_r(2)=12.667$, $p<.01$), reflecting the fact that these children had trouble only with the Baroody task. The differences between the

conditions for the other two groups was not significant. The Q values were 5.2 and 3.6 for the Count and Altered-Question groups.

Summary. The above results provide support for the view that success on a counting task reflects a combination of conceptual, procedural, and utilization competence. Young children may fail a task like Baroody's because they err in their assessment of its requirements and not because they lack the requisite conceptual competence. We suggest that the children in the Ginsburg and Russell (1981) and Fuson and Hall (1983) experiments did not integrate the relevant goal of obtaining the cardinal value of the set with their prior counting behavior. This implicates a faulty execution of a plan or a problem in procedural competence—not one in conceptual competence. It may be that the children in the present study did so well because they were asked to monitor errors; if so, there was an interaction between the task demands and the way the children responded. This possibility highlights how utilization competence can interact with procedural competence.

The 3-Year-Old Study

Our conclusion that children in the above study had the requisite conceptual competence vis-a-vis the cardinal and order-irrelevance principles may well hold for children 4 years old and up. But the criticisms of our attribution of competence regarding order-irrelevance and its relationship to cardinality could hold for still younger children. Hence our decision to do a follow-up study with younger children.

Children in the previous experiment were in one of two experimental groups: Count 3X and Altered-Question. The first served to make sure the child knew what set size was involved; the second to remove a challenge during the critical phase of questioning. Although children benefited some from the memory experience, they did not need it or they would not have done so well in the Altered-Question condition. We thought it was possible that still younger children would be even less confident about their cardinal judgment and hence more likely to give it up simply because they were being questioned again, no matter what the instructions. Pilot work confirmed this possibility. Thus, in this study we ran only two conditions: the standard Baroody condition and a combined Count 3X, Altered-Question one.

Ss and Procedures. Twenty-one Ss were tested. One refused to count and the remaining were randomly assigned to the two groups. The mean age of the Baroody children was 42.9 months (38–46 months); the other group 42.7 months (36–47 months). Children were drawn from the same kinds of centers as above. Owing to personnel needs, there was a two- to three-month delay between the running of Phase 1 and the other two (Trick and Error-Detection). A set size of 5 was used during the Baroody and trick phases; set sizes of 5 and 7 during the

Error-Detection task. Otherwise, the experiment was exactly as described above for the older children.

Results and Discussion. Eight (of 10) children in the experimental condition were correct during the Baroody phase: All said there were five items before the critical question and then predicted there would still be five. Five of them explained this would be the case because they got five the last time they counted. The two children who erred predicted there would only be four objects. Four of the Barody subjects passed this phase, the difference between the groups is significant (Fischer test, $p<.05$).

Despite the difference in Phase 1, there were none on subsequent tasks. Seven in each group (of 10) were correct on the Trick trial. Experimental children were correct (immediate responses) on 82.5% and 100% of their one-one and cardinal-error trials; Baroody control Ss were correct on 78% and 90%, respectively. On one-one trials, more than half the children in both groups (7 in the Experimental group, 6 in the Control group) corrected the puppet's cardinal answer and/or set of count words.

We conclude that Baroody's assessment of children's understanding of the order-irrelevance principle is not valid. The method seems unusually sensitive to factors which influence utilization competence. We think this is because it leaves too much room for misinterpreting questions. We will return to this point.

EXPERIMENT: ORDER-IRRELEVANCE: THE CONSTRAINED COUNTING TASK

In these studies we focus on the constrained counting task (sometimes called *the doesn't matter task*) used by Gelman and Gallistel. Their subjects had to count a display of five items while tagging the second (or fourth) item ''one'' on the first trial, ''two'' on the second trial, ''three'' on the third trial, etc. A child who succeeds on all of these trials will perforce keep altering the order in which tags are assigned to objects and hence provide evidence of applying the principle. As indicated in the introduction, it was only in the group of 5-year-olds that a significant majority of the children in the Gelman and Gallistel study received perfect scores. One could take this to mean that an understanding of the order-irrelevance principle develops late in the preschool years. Alternatively, given the novelty of the task as well as its strategic demands, failures of utilization and procedural competence might account for the younger children's difficulties.

If the culprits in the doesn't matter task are utilization and procedural competence, variations designed to reduce the novel and strategic features of the task should affect levels of success. A comparison of the strategic and performance demands of a 3- versus a 5-item version of the task illustrates one way to accomplish this goal. In each of the displays in Figure 3, a circle represents one of the heterogenous objects in a linear display. The path of arrows indicates the order in which items might be tagged. An X inside a circle marks the object which is to be assigned a particular tag and which tag that should be is given in the column to the right of a row. For example, in Figure 3.1, on trial 1, an individual is supposed to assign the tag ''one'' to the middle item.

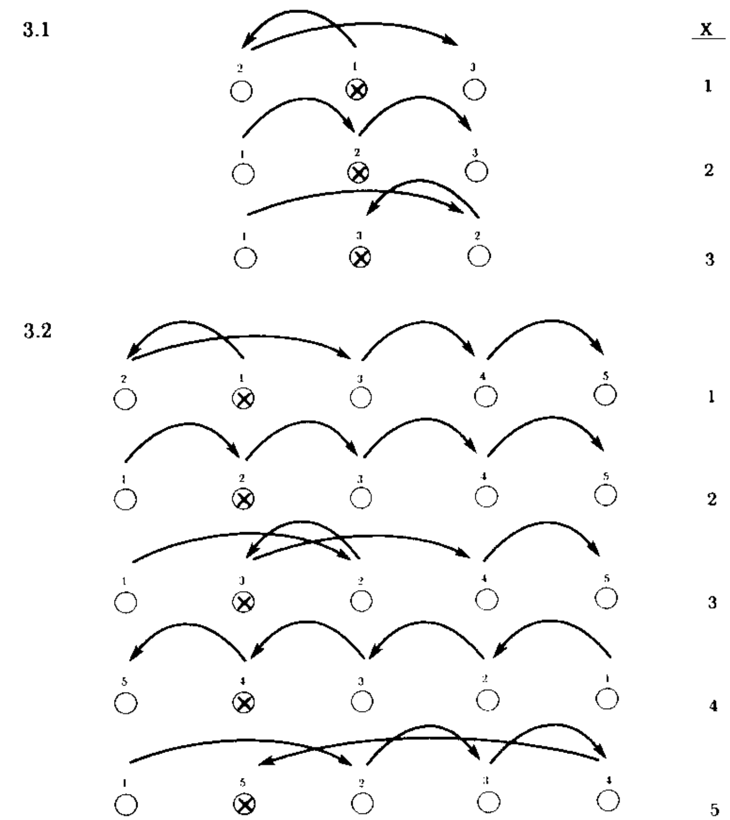

Figure 3. Schematic illustrations of possible correct strategic solutions on the constrained counting task with 3- and 5-item display. The X inside a circle indicates which object is to be tagged with the value shown in the right-hand column. The path of arrows indicates the order in which items are tagged.

One way to solve the constrained counting problem is to tag items out of order and jump. This is shown in line 3 of Figure 3.1. One can also assess whether the spatial order of the target item would correspond to the order in which the designated tag appears in the count list. If children have this utilization competence, then they can count smoothly from one end to the other of the display, as in line 2 of Figure 3.1. Note that, even though one could use the first solution on even these trials, it would not yield as efficient an execution. In the instances of set size 3, the jump-around solution could be used when X=1 and 3; the matching of orders in the count list and the display could be used for X=2. In the case of set size 5, the jump-around solution could be used for X=1, 3, and 5; the matching-orders solution when X=2 and 4.

Despite the commonality of solution types for the two set sizes, their implementations vary in complexity for the different set sizes. As can be seen in Figure 3.2, children have to cover more distance when moving back and forth to apply strategy 1 on set size 5. Further, in order to apply the matching strategy, they have to switch the side from which they start their count when X is the second item from the left and it is supposed to be tagged "four." In the case of set size 3, they need simply recognize the correspondence but in the case of set size 5, they have to go one step beyond this.

There are other solutions children could invent. They could move items and produce a correspondence between the position of X and its tag. They also could start to count a bit in from an end so as to establish a correspondence between the order of tags and tagged items; this would be the case should a child respond to a request to make the middle item of a 5-item display "the two" by starting the count with the second item and tagging it "one," etc., and finally returning to count the first item in the display. Use of such correspondence-producing strategies has the advantage of allowing the child to count efficiently, that is, without jumping around and running the risk of making errors. Thus, although such solutions make more demands on procedural competence, they do have an advantage on larger set sizes. For this reason as well as those given above, we conclude that a 3-item task is strategically simpler than a 5-item one. Still, even the 3-item task requires children approach the task in a novel way and select an acceptable counting procedure in response to the novelty.

A Preliminary Study

Ss and Methods. The above analysis led to the first study with 3- and 4-year-old children to ascertain whether 3-item displays would be easier than 5-item displays. For each age group there were two different conditions with 20 children each. One condition (Easy) started with a 3-item constrained counting task; the other (Hard) a 5-item task. The average ages of the 3-year-olds in the Easy and Hard conditions were 42 months (36–47 months) and 42 months (38–47 months); and the 4-year-olds, 55 months (49–59 months) and 54 months (49–59 months), respectively. Children attended one of five centers serving the same population as described above.

After a child had first counted displays of N items, he/she was introduced to a puppet, asked if she wanted to show him tricks, and told "Here's the trick. Start counting with this (E points to X) and make it number one." This continued with E increasing the required value of X until it was N+1. In the last case, the child had to realize that there were only N items. Children who resist this request provide strong evidence that they know the cardinal value of the display remains constant across trials. The experiment ended with a test designed to assess whether children knew that any count word could be assigned to any object whereas the same was not true for the labels of objects. For details and evidence that they could see Gelman & Meck (in press).

In the Easy condition children were tested on a 3-item, then a 4-item, etc. In the Hard condition they were tested only on set size 5. Half of the latter group were asked to simply count a 3-item array before testing. Since this had no apparent effect, no further reference will be made to this variable.

All sessions were video-taped and then transcribed, trial by trial, in the same fashion as those shown in Figure 3. The only time this was difficult was when children did not point. Such trials had to be scored as errors. A random spot check by an independent transcriber of several tapes, produced a 96% agreements level.

Results and Discussion. Responses to a target question were scored twice—once on the basis of the child's immediate reply, and once on the basis of the best reply. One effect of repeating the question was to clarify the instruction for those children who thought they were simply supposed to tag the item with the given value. The other was to lead the child to repeat the trial. In neither case were the children given clues about what to do. Hence, if simply repeating a request leads a child to use a successful strategy, it seems reasonable to give them credit and use their best effort. The percent repeated trials was 5.8 and 2.5 for the 3- and 4-years-olds in the Easy conditions. Figure 4 displays the percent children who were perfect on each problem for either their immediate or their best responses to a set of questions. To be scored as perfect a child had to be correct on all three standard requests for X=3, all four for X=4, and all five for X=5. The response to the X=N+1 trick trial was scored separately.

It can be seen in Figure 4 that children in the Easy conditions did better than those in the Hard ones. This is true whether we compare performances on the first problem of the Easy group or success levels onset size 5. Eleven of the 20 3-year-olds were able to solve all trials with the 3-item displays on their first attempt; this contrasts with only 1 of the same-aged children tested only with a 5-item display ($\chi^2(1)=16.7$, $p<.001$). Of the older children, 16 and 10 in each of the groups could negotiate correctly each of the 3-item and 5-item requests ($\chi^2(1)=4.81$, $p<.05$). Clearly, the children found the 3-item arrays easier.

When tested on the 5-item arrays, 6 of the 11 3-year-olds who were correct on the 3-item array and 14 of the 16 4-year-olds also scored perfectly. Thus success on the easier, but nevertheless novel task transferred to the harder version of the task. The effects of task difficulty and locus of transfer are clearer if we score the children a little more leniently; i.e., on the basis of their best performance for a target item. As can be seen in Figure 4, by this measure, almost all 4-year-olds in

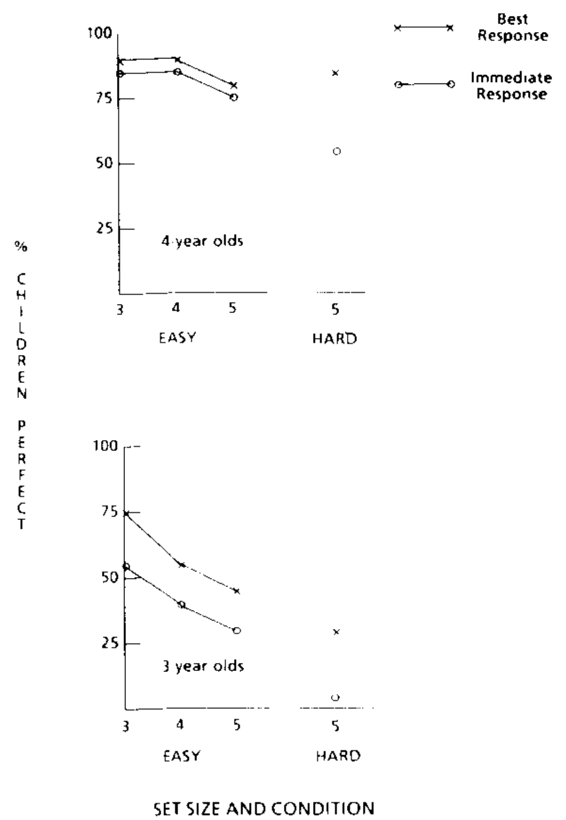

Figure 4. Percent children correct in each condition of the initial constrained counting task, as a function of age and response measure.

both conditions and on all set sizes solve for every value of X. Of the 3-year-olds, 15, 12, and 9 solved every item at each successive set size of 3, 4, and 5. Only 3 of the same age group who were simply tested with a 5-item display did as well by this criterion, even though almost *half* the items were repeated.

The previous results place the main effects of set size and experience with easier problems in the younger group. The 4-year-olds show the same pattern of results only when scored entirely on the basis of their very first response to an item. These results support our hypothesis that difficulty on the doesn't matter task is very much a function of introducing demands on the child's utilization and procedural competence. If we simplify the task, very young children are strategic. Further, they seem to transfer their recognition of how to deal with the task to harder problems. Finally, when, on occassion, the experimenter clarifies

instructions or lets the child try a trial again, children often improve. The latter is especially noteworthy in terms of the levels of success children reach. Four-year-olds are almost always able to solve the problems presented by a trial; and 15 of 20 3-year-olds pass all of their 3-item problems. It is hard to deny these children an ability to apply the order-irrelevance principle and therefore the requisite abilities to honor the one-one, stable-order, and cardinal principles. This conclusion is buttressed by data from the final trick trials.

As indicated, children were asked to make $X=N+1$. Fifty and 75% of the 3- and 4-year-olds in the Easy condition either refused to do so, said they could not, or asked for another item on at least one occasion. None of the children in the younger Hard condition did so—in part because they could not finish their set of initial problems; 70% of the 4-year-olds in the more difficult condition did pass this item.

We should be tentative about the source of transfer in the Easy condition. While 3-year-olds may have done reasonably well on larger set sizes because they started with an easier task, it may be that the critical variable was the opportunity to do the task repeatedly. Hence, in the following study, a condition where children did the 5-item task three times is included. So is one where groups of children count set sizes of 3 and then 4, each three times, before they are tested on just the 5-item "doesn't matter" task. This allows us to assess the effect of children simply counting repeatedly the relevant set sizes before starting testing on the target display. It also makes it more reasonable to compare the data from these studies with that reported by Gelman and Gallistel, since they had their subjects first count a 5-item display six times.

The Follow-Up Study

Subjects and Methods. Thirty children in two age groups were drawn from the same sources as above; 10 each were randomly assigned to three separate conditions. The mean ages of the three groups of 3-year-olds were 43.6 months (39–46 months), 44.6 (38–47 months), and 43 months (39–47 months) in the Experimental, Repeat-Control, and Count-Control conditions, respectively; the mean ages for the 4-year-olds were 54.8 months (48–59 months), 53.1 (51–59 months), and 53.6 (52–57 months), respectively.

Children in the Easy groups were tested exactly as those in the above study. Children in the repeat-control conditions were tested three times with a set size of $N=5$. Children in the count-control condition were tested but once on the order-irrelevance task; before that they were asked to count a linear set size of 3 three times, and then 1 of 4 four times. At the end of each count trial, the array was covered and the child was asked to answer "How many?" items were present. Between count trials the heterogeneous collection of trinkets was rearranged. Here, as above, the displays were always left in place during the successive trials of the constrained counting task.

Data from the order-irrelevance trials were scored in the same ways as in the initial experiment. The differences between immediate-and best-response scores reflect the fact that, for the 3-year-olds, a total of 5 trials from three Easy Ss, 8 trials from 4 Repeat Ss, and 12 trials from 8 Count Ss were repeated. The comparable figures for the 4-year-olds were 6 trials from 3 Ss, 7 trials from 4 Ss, and 8 trials from 5 Ss in the Experimental, Repeat, and Count conditions.

Table 1 summarizes the results. Those from the Easy groups replicate those from the initial experiment (see Figure 4). As before, the older children did better than the younger ones. Still, 3- and 4-year-olds did well on 3-item displays and then transferred their skills to larger set sizes. The fall-off as a function of set size observed in the first experiment is not as notable here—especially when performance is assessed on the basis of best trial data.

Table 1 allows us to compare children in different groups when they were first tested with 5-item displays. If we consider best-trial data, 3-year-olds in the Easy group transfer their skills to larger set sizes. They do quite well on this hardest task of theirs and outperform both control groups—although they do so reliably only when compared to the Count group ($p<.05$, Fischer exact test). Hence, the transfer effect seems due to the opportunity to do the task repeatedly, a conjecture which is supported by the fact that the second round of the young Repeat group's performance compares favorably with that ' the Easy group's. The benefit dissipates by the third repetition—at least in part because some children in the Repeat group (and only this group) did not want to keep doing the exact same task. A similar but nonsignificant trend can be seen in the older group.

Table 1. Percent Children in Second Constrained Counting Task Who Were Perfect on Each Set Size by Either Measure

	Response Measure and Trial Block[a]					
	Immediate			Best		
Conditions	First	Second	Third	First	Second	Third
Easy						
3-year-olds	70	50	30	80	60	50
4 year-olds	90	70	80	100	80	80
Repeat						
3-year-olds	30	40	20	30	50	20
4-year-olds	60	70	60	60	80	70
Count						
3-year-olds	0	—	—	10	—	—
4-year-olds	40	—	—	50	—	—

[a]Trial Blocks in Easy Condition were with increasing set sizes 3, 4, and 5; Repeat Condition with 5, 5, and 5; Count Condition with 5 and once.

Now compare performances for the problems where children first encounter a 5-item display. Children who first had extensive counting, did worse than those who did not. The difference between the Easy and Count groups is reliable for the best-trial, 3-year-old results, and the immediate 4-year-old results (Fischer exact probabilities were at least $<.05$). Still, given that it took different indices (best and immediate) to reveal reliable effects of the prior counting in each age group, one should ask whether the statistical results reflect a real effect. To find out, we considered the kinds of strategies children in the different groups used.

Strategy Analysis. Across all correct trials, children used either one or a combination of four kinds of strategies. The first was restricted to the instruction to make $X=1$; children started their count at X, continued from there to one end of the row and then back to the other side until they got to the item before X, where they finished. This meant that children started counting somewhere in the middle of a row and then counted all items in as smooth a path as possible. Strictly speaking, this is not strategic. Although children who do this are not concerned about starting at the beginning of a row and hence honor the principle, they can simply start at the designated object and count until they have counted everything. Hence, no further mention will be made of these solutions, and only more complex ones that occur on the $X=1$ trials are scored in subsequent analyses.

The second class of strategies also allowed children to count smoothly but depended on their recognizing a correspondence between an item's position in the display and the position of the target count word in the standard count list. In one case, this involved recognizing that the target item was already in the correct position if the child counted from her preferred side, as for example when the $X=2$ request was made for the item in the second position from the left. In the second case, this required the realization that a correspondence would exist if she switched the side from which she began her count. In the third case, the child did not start at either end, but rather at the correct number of items before the target; e.g., when a child who, when asked to make the fourth object in a 5-item row "three," started the count on the second item and came back to tag the first item "five" to finish. We will refer to this second class of strategies as Correspondence-Capitalize strategies. To use these kinds of solutions, children must first recognize that there is an order correspondence and then take advantage of it and execute a smooth sequence of count actions. In doing so, they capitalize on a feature of the display they may or may have not noticed before.

The third class of strategies all involved children jumping over one item or another (see examples in Figure 3) and then returning either immediately or later to tag the bypassed items. We refer to these as Skip-Around strategies. Finally children used Correspondence-Create strategies. They rearranged the display so as to have the position of the targeted item correspond to the ordinal position of the number tag it would receive. Note that in contrast to the Correspondence-Capitalize strategy, children here produce the correspondence instead. They coordinate their assessment of the task with their conceptual ability to detect the correspondence and then produce a solution that takes advantage of their knowledge. In this way both variables that affect utilization and procedural competence are involved in the generation of a solution. In the case of the Correspondence-Capitalize it is mainly utilization competence that is involved—this of course on the assumption that conceptual competence is such that it can represent ordinal features, a point to which we will return.

To assess the tendency of children to use different strategies, we first considered all correct trials ever produced by any child, even if it was a best-trial with a target question. We then deleted from the obtained total those X=1 trials on which children simply started counting at the designated item. For the remaining cases, the percent on which the three different classes of strategies were used, as well as the number of individual children who used the strategies in question, served as summary statistics. The overall tendencies to use the different strategies in the different conditions are shown in Table 2. The question of which combination of strategies individual children used is answered in Figure 5.

Consider the individual tendencies of the 4-year-olds (Figure 5). If they were in the Easy condition, they all used the Skip-Around and Correspondence-Capitalize strategies; only 1 (of 10) of these same children also used the Correspondence-Create strategy. In contrast, 6 of the 10 Count-Control children used the latter as well as the Capitalize strategy and only 3 of these children combined Skip and Capitalize solutions. Children's solution preferences were more evenly divided in the Repeat-Control group of 4-year-olds.

The picture for the 3-year-olds is somewhat different, owing to the fact that their overall tendency to use Create-Correspondence strategies was very limited

Table 2. Percent Correct[a] Strategy Types Used in Each Condition of Second Constrained Counting Experiment

	Strategy		
Conditions	Skip	Recognize Correspondence	Create Correspondence
3-year-olds			
Easy	57	40	2
Repeat	48	44	8
Count	(23)[b]	(54)	(23)
4-year-olds			
Easy	57	40	2
Repeat	33	22	44
Count	20	32	49

[a]Excluding $X = 1$ trials where no strategy used.
[b]Percents given in brackets since N's are small; 3 for S; 7 for R; 3 for C.

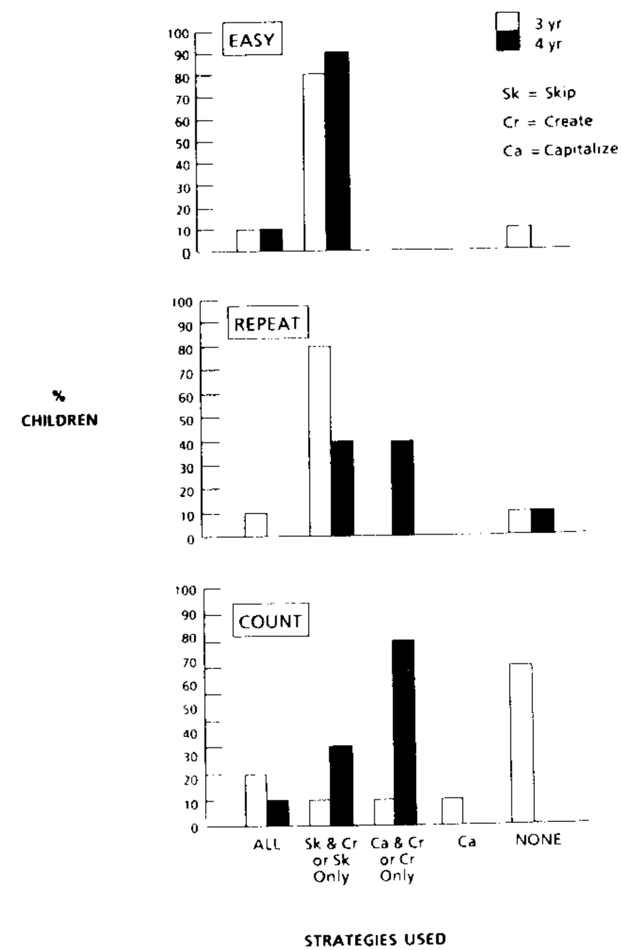

Figure 5. Percent children in each condition of the follow-Up constrained counting task who used each pattern of strategies on correct trials. Read the entries on the abcissa from left to right as follows: ALL Ss used each of the three Skip, Create, and Capitalize strategies; the next group of Ss used just the Skip strategy or both Skip and Create; the Ca or Ca and Cr Ss used only the Create strategy or both Capitalize and Create; Ca Ss used only the Capitalize strategy; NONE Ss failed to use any successful strategies.

(see Table 2). Like the 4-year-olds, many (7 of 10) used a combination of Skip and Correspondence-Capitalize solutions in the Easy condition; very few used Correspondence-Create solutions. In contrast to the 4-year-olds, the younger children did not use Create strategies in the Count-Control condition. And, instead of splitting their responses between Create and Skip, the Repeat-Controls relied on the latter solutions when not using Correspondence-Capitalize.

Why this pattern? We suggest that the younger children did well when they did because they learned to use the Skip-Around strategy (either on small sets or over repeated efforts with the larger sets) and/or they did not assume that they had to count in an orderly fashion from one end of the row to the other. Children in the Count-Control group came to assume they were supposed to be orderly, and since they were less able than older children to produce correspondences between the order of items and count words, they did very poorly. A similar account holds for the pattern of results in the older group, given that they were more able to produce correspondences. The older Count-Control children could do well even if they assumed their counts were supposed to be orderly, because they could create the order.

Support for the conclusion that Count-Control Ss developed a set for orderly left-right (or vice versa) counting comes from looking at what 3-year-olds in this condition did on their error trials. Seven of the ten children repeatedly produced trials in which they tagged items in a smooth path from one end of the arrary to the other and, to do so, either altered the count list by either skipping and/or rearranging words e.g., 1, 2, 4, 5, 6; 1, 3, 2, 4, 5) or assigned two count words in sequence to one item (e.g., 1, 2, 3, 4–5, 6 when asked to make the fourth item be "five"). Only three children of the same age in the other two groups did this at all; and only one child in each of the three groups of older children did.

X+1 Results. The above findings make it hard to deny the children an implicit understanding of the order-irrelevance principle. That task difficulty can influence the display of such knowledge is once again demonstrated by the results of the Trick trials. Six of the 3-year-olds in the Easy and Repeat conditions each refused to do the trial or said the number of items was wrong on at least one occasion; in contrast only one did so in the Count condition. As in the first study, condition did not affect the older children. Eight in each group rejected or commented appropriately on the request.

Summary. The results of this experiment allow us to conclude that there are conditions under which children as young as 3-years-old will honor the order-irrelevance principle. When small arrays are used, the strategic demands are fewer and children do well. Three-year-olds then use the same strategies that work on small set sizes (Skip and Capitalize) on larger set sizes. If, however, they develop a set to count smoothly from one end of the array to the other, they are at a disadvantage because they seem unable to create the correspondences they can capitalize on. The older children can create correspondences and do so when it is to their advantage, as in the Count condition and when they encounter larger sets. In other words, their production competence is ahead of that of the younger children. Hence, we see how development can involve improvements in both utilization and procedural competence.

CONCLUSIONS

We started this paper with the thesis that early skill at counting is guided by the availability of implicit counting principles. In response to the argument that young children's counting performances, both across and between tasks, is too variable to be consistent with an understanding of the principles, we introduced an analysis of the system that translates principles into practice (conceptual competence into behavior). This analysis distinguishes between conceptual, procedural, and utilization competence, thereby allowing us to pinpoint factors that determine the variability in performance. In a series of experiments, we have shown that variability in counting performances can be traced to problems in assessing the task (utilization competence) and planning solutions (procedural competence) that meet the constraints of conceptual competence.

The studies highlight the role of social factors which influence a child's assessment of a task. We suggested that tasks that provide novel conditions—as in the case of pseudoerror trials and constrained counting requests—are especially likely to yield misinterpretations; hence, subtle variations in instruction can have profound effects on performance levels. The suggestion here is that the younger the children, the more likely they are to require completely unambiguous instructions, presumably because they have less experience stepping outside a particular interpretation of an utterance and selecting another (cf. Donaldson, 1978; Shatz, 1983).

The strategy analysis of our last study provides a clear example of the way conceptual competence can guide the acquisition of skill. Both 3- and 4-year-olds could recognize and take advantage of the correspondence between the position of a specified item in a row and the position of a given number word in the count sequence, but only the older children could create a correspondence when it did not already exist.

The fact that even 3-year-olds could take advantage of correspondences was a surprise to us. To do so, the child has to have some implicit knowledge of the ordinal use of numbers, the use of a number to represent the position of an item in an ordering (rather than, or in additon to numerosity). There is nothing in the counting principles we have so far postulated that makes this possible. To be sure, the tags used must be ordered; but there is no ordinal representation principle, a principle that asserts that when the count-procedure used in counting an ordered set is order-preserving, then the number used to tag an item represents the position of that item in the ordering. How to explain this knowledge that is not a part of the initial counting principles (Botman, 1981)? One could say there is nothing to explain, that these findings merely indicate that the list of implicit principles that underlie early competence is more extensive than we thought. Alternatively, one could argue that here is a case where conceptual competence (a principle) develops out of procedural competence (a practice). Since, we believe that there must be such cases, we end by arguing in favor of this latter alternative.

It has already been noted that children who count left to right are usually at an advantage over those who skip around. They are less likely to miss or double-count items. This is a fact about utilization competence that influences procedural competence. Since smooth left-to-right or right-to-left counts are most likely to succeed, they are reinforced and become conventionalized. The effect of this is that the child has a great deal of experience that supports the learning of the correspondence between the position of items in ordered displays and the count-word with which they are tagged. In short, by practicing the standard counting procedure, it is possible to develop an appreciation of the capacity of numbers to represent things other than numerosity. We do not contend that this is the correct account of the ordinal number concept; just that it is an illustration of how procedural competence can lead to the acquisition of conceptual competence. We start by granting the child some implicit understanding of counting, which guides the acquisition of skill, which in turn begets further understanding. A scenario of this form shows how conceptual competence may itself develop epigenetically. Conceptual competence need not be entirely or even mostly preformed, but a preformed kernel is a prerequisite for the development of both procedural competence and further conceptual competence.

REFERENCES

Baroody, A. J. (1984). More precisely defining and measuring the order-irrelevance principle. *Journal of Experimental Child Psychology, 38,* 33–41.

Botman, H. (1981). *Cardinal and ordinal number skills in preschool children.* Unpublished manuscript, Duke University.

Briars, D., & Siegler, R. S. (1984). A featural analysis of preschoolers' counting knowledge. *Developmental Psychology, 20,* 607–618.

Brown, A. L. & Reeve, R. A. (in press). Bandwidths of competence: The role of supportive contexts in learning and development. In L. S. Liben & D. H. Feldman (Eds.), *Development and learning: conflict or congruence.* Hillsdale, NJ: Erlbaum.

Case, R. (1984). The process of stage transition: A neo-Piagetian view. In R. J. Sternberg Ed.), *Mechanisms of cognitive development.* San Francisco: Freeman.

Chomsky, N. (1957). *Syntactic structures.* The Hague: Mouton.

Donaldson, M. (1978). *Children's minds.* New York: Norton.

Fuson, K. C., & Hall, J. W. (1983). The acquisition of early number word meanings: A conceptual analysis and review. In H. P. Ginsburg (Ed.), *The development of mathematical thinking.* New York: Academic.

Gelman, R. (1982). Basic numerical abilities. In R. J. Sternberg (Ed.), *Advances in the psychology of intelligence* (Vol. 1). Hillsdale, NJ: Erlbaum.

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

Gelman, R., & Meck, E. (1983). Preschoolers' counting: Principle before skill. *Cognition, 13,* 343–359.

Gelman, R., & Meck, E. (in press). The notion of principle: The case of counting. In J. A. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics.* Hillsdale, NJ: Erlbaum.

Ginsburg, H. P., & Russell, R. L. (1981). Social class and racial influences on early mathematical thinking. *Monographs of the Society for Research in Child Development. 46* (6, Serial No. 193).

Greeno, J. G., Riley, M. S., & Gelman, R. (1984). Conceptual competence and children's counting. *Cognitive Psychology, 16,* 94–134.

Labov, W., & Fanshel, D. (1976). *Therapeutic discourse: Psychotherapy of conversations.* New York: Academic.

Schaeffer, B., Eggleston, V. H., & Scott, J. L. (1974). Number development on young children. *Cognitive Psychology, 6,* 357–79.

Smith, D. A., & Greeno, J. G. (1984). *Implicit understanding and competence: A theoretical analysis of procedural and utilization competence.* Unpublished manuscript, Learning, Research and Development Center, University of Pittsburgh.

Wilkinson, A. C. (1984). Children's partial knowledge of the cognitive skill of counting. *Cognitive Psychology, 16,* 28–64.