Journal of Experimental Psychology:
Learning, Memory, and Cognition
1991, Vol. 17, No. 3, 398-415

0278-7393/91/$3.00

Mathematical Problem Solving by Analogy

Laura R. Novick
Vanderbilt University

Keith J. Holyoak
University of California, Los Angeles

We report the results of 2 experiments and a verbal protocol study examining the component processes of solving mathematical word problems by analogy. College students first studied a problem and its solution, which provided a potential source for analogical transfer. Then they attempted to solve several analogous problems. For some problems, subjects received one of a variety of hints designed to reduce or eliminate the difficulty of some of the major processes hypothesized to be involved in analogical transfer. Our studies yielded 4 major findings. First, the process of *mapping* the features of the source and target problems and the process of *adapting* the source solution procedure for use in solving the target problem were clearly distinguished: (a) Successful mapping was found to be insufficient for successful transfer and (b) adaptation was found to be a major source of transfer difficulty. Second, we obtained direct evidence that schema induction is a natural consequence of analogical transfer. The schema was found to co-exist with the problems from which it was induced, and both the schema and the individual problems facilitated later transfer. Third, for our multiple-solution problems, the relation between analogical transfer and solution accuracy was mediated by the degree of time pressure exerted for the test problems. Finally, mathematical expertise was a significant predictor of analogical transfer, but general analogical reasoning ability was not. The implications of the results for models of analogical transfer and for instruction were considered.

In the past decade, there has been a surge of interest in the psychological study of analogical problem solving. Many studies have shown that solving an initial *source* problem can influence solution of a subsequent analogous *target* problem, as a result of exploiting systematic correspondences between the problems. At the same time, abundant evidence indicates that analogical transfer is highly fallible. Some analogies produce robust transfer (e.g., Holyoak & Koh, 1987), but others do not (e.g., Gick & Holyoak, 1980). Moreover, some individuals are reliably more successful than others (Novick, 1988a). The sources of variation in the use of analogies to solve problems are far from understood. The aim of the present paper is to investigate such variation in the domain of mathematical word problems. Because of their rigorous structure, these problems are particularly well suited for analyzing the formal basis of analogical transfer. Moreover, the study of mathematical problem solving has direct educational relevance.

This research was supported by National Science Foundation Grant BNS-8215316 awarded to Keith Holyoak, and by an NIMH Individual Postdoctoral Fellowship, two Vanderbilt University Research Council Grants, and a grant from the dean's office of Peabody College of Vanderbilt University awarded to Laura Novick. We thank Paula Messamer, Livian Perez, and Suzanne Dutka for their help in running subjects. Michael Whitlock deserves special thanks for assisting us in determining the reliability of our coding schemes. We also thank Miriam Bassok and several anonymous reviewers for their helpful comments on earlier versions of the manuscript.

Correspondence concerning this article should be addressed to Laura Novick, Department of Psychology and Human Development, Box 512 Peabody, Vanderbilt University, Nashville, Tennessee 37203.

The Role of Adaptation in Analogical Problem Solving

Task analyses of problem solving by analogy have focused on three processes: *retrieval*, *mapping*, and *learning*. While reading and attempting to understand the target problem, solvers will encode it in terms of various features. These features will provide memory retrieval cues that may enable the solver to access information about relevant problems encountered previously. To investigate the retrieval process, researchers have often used a hint/no-hint paradigm (Gick & Holyoak, 1980). The hint informs subjects that information presented earlier might help them solve the target problem. If performance is better with than without a hint, then retrieval is an important component of transfer, and one that solvers have difficulty executing on their own. That, in fact, is the modal finding (e.g., Catrambone & Holyoak, 1989; Gick & Holyoak, 1980, 1983; Keane, 1988; Reed, Ernst, & Banerji, 1974; Ross, 1987, 1989a; Spencer & Weisberg, 1986).

After retrieving a potential source problem, the solver must construct a mapping between it and the target problem. This step is considered the hallmark of analogy: "At the core of analogical thinking lies the process of *mapping*: the construction of orderly correspondences between the elements of a source analog and those of a target" (Holyoak & Thagard, 1989a, p. 295). Gentner's (1983) influential "structure-mapping" theory of analogy incorporates this idea in its name.

On the typical view, successful mapping enables the solver to transfer the solution procedure of the source problem to the target problem. We do not take issue with this claim for the *necessity* of mapping in analogical transfer. However, we do question the further assumption of the *sufficiency* of mapping for successful transfer. We will show that procedural transfer is not an automatic consequence of successful mapping. Rather, a potentially laborious *adaptation* process (a fourth component of analogy use) is required to construct an analogous procedure for the target problem (Gholson, Morgan, & Dattel, 1990; Novick, 1988b). This distinction between mapping and adaptation also has been made in the artificial-intelligence and proportional-analogy literatures (e.g., Burstein, 1986; Carbonell, 1983; Hofstadter & Mitchell, 1988; Sternberg, 1977); and it is implicit in Holyoak's (1985; Holland, Holyoak, Nisbett, & Thagard, 1986) discussion of morphisms.

Analogical Transfer and Schema Induction

The final component of analogy use is learning. Holyoak (1984, 1985; Holyoak & Thagard, 1989b) and Ross (1989b; Ross & Kennedy, 1990) have suggested that successful analogical transfer leads to the formation of an abstract schema for the class of problems represented by the source and target (also see Anderson & Thompson, 1989). The induced schema is hypothesized to facilitate subsequent problem solving with further analogues (Anderson & Thompson, 1989; Gick & Holyoak, 1983; Pirolli & Anderson, 1985; Ross, 1989b).

Current data are equivocal in their support for the schema induction hypothesis. Ross and Kennedy (1990) reasoned that if analogical transfer leads to schema induction, which facilitates later problem solving, then a manipulation that increases analogical transfer for target problems also enhances performance on subsequent analogous problems. As predicted, Ross and Kennedy found in several experiments that performance on the final test problems differed as a function of a transfer manipulation introduced earlier. Thus, their results provide indirect support for the schema induction hypothesis. Ross and Kennedy also presented some data that argue against two alternative explanations for their findings: (a) Better performance on the target problems may lead directly to better performance on the final problems, with no intervening role played by schema induction, and (b) the pattern of results for the final problems was due to differential memory for the example problems as a result of the transfer manipulation. However, in contrast to this positive support for a link between analogical transfer and schema induction, Reed (1989) found no evidence for schema induction using an experimental logic and methodology similar to that of Ross and Kennedy. Both sets of studies used mathematical word problems.

The only direct evidence for schema formation as a natural consequence of analogical transfer comes from think-aloud protocols for a single subject solving two algebra word problems, after having learned equations for solving isomorphic constant-acceleration problems (Bassok & Holyoak, 1989). Clearly, additional direct evidence for the schema induction hypothesis is needed. To measure schema induction directly, we asked subjects to describe the solution procedure common to the source and target problems immediately after solving the target problem(s). If evidence for schema induction as a consequence of transfer can be found, it will be important to replicate the suggestion in Ross and Kennedy's data that the effect of schema induction on later problem solving is independent of any direct effect of prior analogical transfer.

Analogical Problem Solving, Analogical Reasoning Ability, and Expertise

Being able to predict who will succeed at analogical transfer with mathematical problems is important for designing effective instructional programs. Novick (1988a) has shown that performance on the math section of the Scholastic Aptitude Test (MSAT), a measure of mathematical expertise,[1] predicts spontaneous analogical transfer. At least part of this relation between expertise and transfer is due to the effect of expertise on retrieval of the source problem. It is not known whether expertise will continue to be important when spontaneous retrieval of the source problem is not required. The results of Experiment 2 will enable us to address this issue because subjects were told the identity of the source problem (i.e., spontaneous retrieval was not required).

Regardless of the role of expertise, one would expect psychometric measures of general analogical reasoning ability to predict performance on analogical problem-solving tasks (Holyoak, 1984). In fact, Spearman (1923) advocated analogy problems on mental tests because he argued that analogical reasoning is pervasive in everyday life. Despite researchers' intuitions and despite the fact that ability tests predict school achievement, we are unaware of work examining whether specific components of ability tests predict performance in the specific real-world tasks they were designed to emulate. Therefore, in Experiment 2, we gave subjects a standardized test of analogical reasoning ability. We were interested in the relative predictive power of this general measure compared to a specific measure of math expertise. Because we studied analogical transfer in math, we expected that the specific measure would be a better predictor of transfer.

Overview of the Experiments

Nature of the Experimental Domain

Given the size and diversity of mathematics, we obviously had to limit our studies in several ways. First, we chose to study analogical problem solving in a situation in which solvers have imperfect knowledge of the source information, resulting from a single exposure to the problem that will be the source of analogical transfer. In this regard, our work is similar to that of Reed (1987, 1989) and Ross (1987, 1989a) but differs from that of Bassok and Holyoak (1989). Such imperfect knowledge is characteristic of many instructional situations.

---

[1] Novick (1988a) provided several justifications for using MSAT scores as a measure of mathematical expertise. We will only summarize the two most important points here. First, good performance on the math section of the SAT requires the same kinds of skills (primarily competence in arithmetic) that are needed to solve the problems used in the present studies. Second, scores on the math SAT increase with increased instruction in mathematics, as one would expect of a measure of expertise (College Entrance Examination Board, 1986; Messick, 1980).

Although our work is similar to that of Reed and Ross in this respect, it differs from their work in terms of solvers' prior relevant knowledge of the domain. Reed gave his subjects algebra word problems before they encountered such problems in their algebra class. Ross gave probability problems to statistically-naive subjects. In contrast, we used arithmetic story problems with subjects whose most advanced math class was typically some form of calculus. In Experiment 1, 55% of the subjects had taken at least first semester calculus, and only 17% had failed to progress as far as pre-calculus. In Experiment 2, these percentages were 83% and 11%. Thus although the particular procedure taught with the source problem was unknown to subjects (Novick, 1988a), most (if not all) subjects had encountered its components earlier in their mathematical education.

Finally, like Bassok and Holyoak, but unlike Reed and Ross, we chose problems that can be solved using several procedures besides the one learned for the source problem. This characteristic of our materials mimics many real-world problems (e.g., there are multiple ways to analyze most complex sets of data, even when attempting to answer a specific question). It also has important implications for how to measure analogical transfer.

Measuring Analogical Transfer

Analogical transfer occurs when specific, structural correspondences between objects and relations in the source and target are used to adapt the solution procedure learned for the source into an analogous procedure for solving the target. For problems that *must* be solved using a particular procedure (excluding the possibilities of guessing and estimation), transfer may be indicated by an increase in either use of that procedure or accuracy (e.g., the probability problems used by Ross, 1987; the tumor problem used by Gick & Holyoak, 1980, 1983). An increase in accuracy, however, is not a necessary sign of analogical transfer. For example, a student might construct the equation for an algebra word problem by analogy to a previous word problem but then fall back on general procedures for solving equations that were learned outside the context of word problems. If those procedures were not well learned, analogical use of the source to construct the equation for the target will not ensure correct solution of the target.

In fact, accuracy levels may provide *no* evidence concerning the incidence of analogical transfer. This is particularly true for problems that can be solved by several procedures. Consider Bassok and Holyoak's (1989) work on isomorphic domains of physics and algebra. In one condition, subjects learned formulas for solving algebra problems and then received physics problems to solve. There was little difference in accuracy between these subjects and others who received the physics problems without prior algebra training because the latter subjects could use non-analogical methods to solve the problems. Nonetheless, there was a huge analogical transfer effect: 72% use of the algebraic formulas by trained subjects versus 0% by untrained subjects.

Similar considerations apply in using solution time as an index of analogical transfer. If the source solution procedure is both efficient and readily adaptable to fit the target problem, then solution times should be shorter for subjects who show successful analogical transfer than for subjects who correctly solve the target problem by a less efficient method. But if adapting the solution procedure is difficult, analogical transfer may not benefit solution time.

For multiple-solution problems, when analogical transfer does decrease solution time, the amount of time subjects are given to solve the problem may determine whether transfer also enhances solution accuracy. With little time pressure, so that most subjects have enough time to solve the problem regardless of the procedure they use, there is unlikely to be an increase in accuracy associated with an increase in procedural transfer. Under time pressure, however, the benefits of analogical transfer should be more apparent because subjects will not have enough time to devise and execute a new procedure.

Our studies concern analogical problem solving with multiple-solution problems, thus hypotheses about analogical transfer will be tested by analyzing the procedures subjects use. In addition, the relation between analogical transfer and solution accuracy will be considered by testing the hypothesized effects of differential time pressure described above. In Experiment 1 subjects were placed under relatively little time pressure, whereas in Experiment 2 they were under considerable time pressure.

Synopsis of the Experimental Procedure

Subjects first studied the source problem and its solution procedure. Then they received one (Experiment 1) or two (Experiment 2) analogous target problems to solve. In a between-subjects manipulation, different hints were provided with the target problem(s). These hints either retrieved the source problem for subjects or provided them with information concerning the mapping between the source and target problems. After solving the target problem(s), subjects received a task to elicit whatever schema they might have induced for the solution procedure shared by the source and target problems. Finally, subjects received what we will call a "generalization" problem. This problem could be solved using the same procedure that worked for the source and target, but the procedure required additional adaptations beyond those required to show transfer for the target problem(s). Performance on this problem enabled us to assess the independent contributions of prior analogical transfer and schema induction to later problem solving.

Differentiating Mapping and Adaptation

We will provide evidence for the importance of the adaptation process in analogical transfer and correspondingly for a distinction between mapping and adaptation. Although the boundary between these processes is fuzzy, we nonetheless believe it is possible to study these separate components of analogical transfer. This goal necessitated two methodological innovations. Because isolating the adaptation process requires that subjects know the mapping between the source and target problems, we extended the hint paradigm used to study retrieval. The logic of this paradigm is that if a particular process is important for transfer and difficult for solvers to execute on their own, then subjects who have the process performed for them by a hint should be more likely to show transfer than subjects who must execute the process themselves. Moreover, as the hint performs one or more of the early processes required for transfer, it enables one to study the later processes uncontaminated by the earlier processes.

We constructed two hints that provided different kinds of mapping information. These hints stated the conceptual and numerical correspondences, respectively, between the source and target problems. We predicted that the number-mapping hint would be more useful for adaptation than the concept-mapping hint because that hint gets subjects closer to the dividing line between the mapping and adaptation processes. The basis for this prediction will be discussed in more depth later.

Second, studying the role of adaptation required that we do more than simply code use of the procedure indicative of transfer. Thus, we also performed a detailed coding of the errors subjects made in attempting to transfer the source procedure to the target and generalization problems.

Experiment 1

This experiment provided an initial test of our hypotheses concerning (a) the mapping and adaptation components of analogical transfer and (b) the relation between transfer and schema induction. In addition, we tested our hypothesis regarding the relation between analogical transfer and accuracy in the absence of speed pressure.

To test the insufficiency of mapping and the difficulty of adaptation, we compared experimental conditions that differed in their process requirements for transfer: Some subjects were told the conceptual or numerical mapping between the source and target (mapping hint conditions), whereas other subjects were only told the identity of the source problem (retrieval hint condition). Because mapping is necessary for transfer, we expected both mapping hint conditions to produce higher transfer rates than the retrieval hint condition. Successful adaptation of the source procedure required that solvers perform the same kinds of operations on the numbers in the target problem that were performed on the corresponding numbers in the source problem. Thus, adaptation required both knowledge of and attention to the numerical mappings. Therefore, we expected transfer to be enhanced in the number-mapping condition relative to the concept-mapping condition. We also included a condition in which the target problem was presented without mentioning the source problem (no hint condition). Given the transfer results for other problems, this condition should produce the worst performance because subjects must execute all of the hypothesized processes themselves. To summarize, we predicted monotonically increasing transfer across the no, retrieval, concept-mapping, and number-mapping hint conditions.

The written solution protocols from both this experiment and the next experiment were coded for the types of errors subjects made in attempting to adapt the source procedure for use with the target and generalization problems. For the sake of brevity, we will present the error data from both experiments together in a separate section after Experiment 2.

Method

Subjects. The subjects were 75 UCLA undergraduates (42 females and 33 males) who participated in partial fulfillment of course requirements. They were randomly assigned to one of four conditions: no hint (16 subjects), retrieval hint (20 subjects), concept-mapping hint (19 subjects), and number-mapping hint (20 subjects). An additional 13 subjects participated, but their data were excluded from the analyses because the answers those subjects gave for the source problem indicated that they failed to understand its solution procedure.

Materials. Subjects received three analogous problems: "garden," "band," and "seashell." Table 1 summarizes the similarities and differences among these problems, and the complete text of the problems may be found in Appendix A. (Table 1 also describes the "bake-sale" problem that was used only in Experiment 2 and which

Table 1
Similarities and Differences Among the Source, Target, and Generalization Problems Used in Experiments 1 and 2

Property	Source problem (Exps. 1 & 2)	Target problem 1 (Exps. 1 & 2)	Target problem 2 (Exp. 2)	Generalization problem (Exps. 1 & 2)
Content domain	Vegetable garden	Marching band	Bake sale	Seashell collection
Goal: How many . . .	Plants	Band members	Cookies	Seashells
Divisors with the same remainder	10, 4, 5	12, 8, 3	16, 14, 8	5, 6, 9, 10
Constant remainder	2	1	6	4
When find out about constant remainder	At end of problem	After each divisor	After each divisor	After each divisor
Divisor with no (zero) remainder	6	5	9	7
Divisor with a new remainder	—	—	—	3 (r1)
Range constraint	Find fewest number	45–200 students	Find fewest number	80–550 seashells
Answer is based on this LCM multiple	Second	Sixth	Third	Fourth

we will discuss later.) The source problem (garden) involved determining how many plants a couple bought for their garden, given information about what happened when they considered buying different numbers of plants of each kind. The target problem (band) involved determining the number of students in a marching band, given how many students were left out of the formation when each of several numbers of students per row were tried. These problems were taken from Novick (1988a). Both can be solved most efficiently by finding the lowest common multiple (LCM) of the three divisors that leave a constant remainder, generating multiples of the LCM, adding the constant remainder to each multiple, and picking from that set the number that is evenly divisible by a fourth number and also fits any additional constraints given (the "LCM procedure"). The generalization problem (seashell) involved determining how many seashells a girl had, given what happened when she tried counting them by various numbers (adapted from Corcoran, Gaughan, Ladd, & Salem, 1981). This problem can be solved using a generalization of the LCM procedure. In particular, there were four divisors that left a constant remainder instead of three, and there was an additional divisor that left a different remainder (besides the divisor that left no remainder).

All of these problems also can be solved by examining multiples of any of the divisors given in the problems, instead of by examining multiples of the LCM. These other multiples procedures are less efficient than the LCM procedure, however, because more multiples must be tested. For example, subjects who solve the band problem by examining multiples of 12 or of 5 must consider 2–4 times as many multiples as subjects who use the LCM procedure. Subjects who fail to show transfer use these other procedures (Novick, 1988a).

Design and procedure. The experiment used a one-way, between-subjects design with four levels. The manipulated variable was the type of information included in a hint given to subjects when they received the target (band) problem.

In the first task, subjects solved four word problems using solution procedures that were described in detail, under the guise of evaluating the comprehensibility of the solution procedures for a later experiment. This directed-solution task was taken from Novick (1988a). The garden source problem was presented second. The remaining problems were unrelated to the LCM problems with respect to their cover stories and their solution procedures.[2] Subjects were given 7 min to solve each problem, and the next problem was not presented until after 7 min had elapsed.

After solving the four initial problems, subjects were told that the real experiment was about to begin. Then they were given 15 min to work on the band target problem. Previous work (Novick, 1988a; pilot studies) suggested that this was ample time for most subjects to solve the problem, thus satisfying our design constraint of little, if any, time pressure.

One group of subjects received no hint with the band problem. A second group of subjects received a retrieval hint; it served to retrieve the source problem:

> The garden problem you saw earlier is similar to this problem. So try to use the garden problem to help you solve the marching band problem. In particular, try to use what you learned about how to solve the garden problem to come up with the same kind of procedure for solving the marching band problem. There may be other ways to solve the marching band problem, but it's very important that you try to use what you learned from the garden problem to help you solve the band problem.

This hint was read out loud to subjects when they received the band problem. It was also printed below the problem so subjects could not forget it during their solution attempt.

The remaining subjects received either a concept- or a number-mapping hint. Like the retrieval hint, these hints were presented both verbally and in writing. Both hints provided specific information about the relation between the garden and band problems, which was inserted after the first sentence of the retrieval hint. The remaining sentences of the retrieval hint were given after the mapping information. The conceptual correspondences provided by the concept-mapping hint were:

> In particular, your goal in this problem is to arrange band members into rows or columns so that each row (or each column) has the same number of people in it, with no one left over. That's like the goal you had in the garden problem of grouping plants into different types so that there were the same number of plants of each type, with none left over. In the garden problem the major difficulty encountered was that once the Renshaws finally figured out how many plants they had room for in their garden, all of the arrangements they had thought of failed to accommodate 2 plants. There is a similar difficulty in the marching band problem. There, each formation the band director thought of failed to accommodate 1 person. So to summarize, the band members are like plants, the rows and columns of band members are like kinds of plants, and the number of band members per row or column is like the number of plants of each kind.

The number-mapping hint simply told subjects the correspondences between the numbers in the two problems, without giving any supporting conceptual structure: "In particular, the 12, 8, and 3 in the band problem are like the 10, 4, and 5 in the garden problem. Also, the 1 in this problem is like the 2 in the garden problem. Finally, the 5 in this problem is like the 6 in the garden problem."

After the 15 min were up or after subjects finished working on the problem, whichever came first, subjects were given 6 min to complete two tasks intended to elicit information about their schemas for LCM problems. One task was to "Please write down all the similarities you can think of between the marching band problem you just worked on and the garden problem you saw earlier." The second task was to "Please write down all the similarities you can think of between how the garden problem is solved and how the band problem is solved."

After the 6 min had elapsed, subjects spent 20–30 min working on two filler tasks. Finally, they were given 20 min to work on the seashell generalization problem. No hints were given with this problem. Subjects participated individually or in groups of 2 to 4 in a single session that lasted approximately 2 hr. Math SAT scores were provided by UCLA for all subjects.

Results and Discussion

Transfer processes. Three levels of performance were scored for the band target problem: successful analogical transfer (score of 2), partial transfer (score of 1), and no transfer (score of 0). A score of two was awarded to subjects who correctly used the LCM procedure: Find the LCM of the divisors that leave the same remainder (LCM = 24), generate multiples of the LCM, add the constant remainder (1) to each multiple,

[2] The first problem concerned two trains traveling toward each other and a bird that flew back and forth between them. The goal was to figure out how far the bird traveled (adapted from Posner, 1973, pp. 150–151). The third problem was a Pythagorean theorem problem. The fourth problem required solvers to divide 15 canteens filled with varying amounts of water among 3 men without pouring water among the canteens. The men had to receive the same number of canteens and the same amount of water (adapted from Wickelgren, 1974, p. 97).

and then pick the number from this set that satisfies the remaining constants of the problem (answer = 145).[3] Partial credit was awarded to subjects who indicated at least three consecutive multiples of the LCM, or at least two consecutive multiples in addition to the LCM, or at least three multiples of a number incorrectly believed to be the LCM (e.g., 48).[4] All other subjects received a score of zero. Subjects' solutions were coded with a reliability of .93. The method used to assess reliabilities for all coding schemes is described in Appendix B.

We used Abelson and Tukey's (1963) monotonic trend contrast (weights of −6, −1, 1, 6, for increasing levels of hints) to test our prediction of monotonically increasing transfer rates across the no, retrieval, concept-mapping, and number-mapping hint conditions.[5] The contrast was reliable, $F(1, 71) = 9.54$, $p < .003$, $MS_e = 0.72$, with means of 0.50, 0.90, 0.84, and 1.40 for the four conditions, respectively. (The rates of successful transfer mirrored this pattern: 19%, 35%, 37%, 50%, respectively). The residual term was not reliable, $F(2, 71) < 1$.[6] The one exception to our prediction of increasing transfer rates across conditions was that the concept-mapping and retrieval hints led to comparable performance, $t(71) = -0.21$, $p > .80$. Evidently, once the analogue was retrieved, solvers typically were able to determine on their own the mapping of the concepts.[7]

Consistent with our hypothesis that the numerical mappings are crucial for adaptation, a comparison of the two mapping conditions favored subjects who received the number hint, $t(71) = 2.05$, $p < .05$. Nevertheless, knowing the numerical correspondences between the source and target was not sufficient for transfer. The fact that only 50% of the subjects in the number-mapping condition were successful at transfer strongly suggests that the adaptation process is a major source of difficulty, separate from the difficulty of the mapping process. This conclusion is supported further by the fact that an additional 40% of the subjects in the number-mapping condition received partial credit for transfer because their transfer attempts were not entirely successful.

An uninteresting alternative explanation for the finding that only half of the subjects in the number-mapping condition were successful at transfer is that the remaining subjects failed to remember the source procedure. Some data from a pilot study argue against this hypothesis. In that study, subjects who failed to show transfer after a retrieval hint were given a concept-mapping hint. Then they worked on the generalization problem and the schema task. At that point, 1.5–2 hr after they saw the garden problem, subjects who failed to ever show transfer on the target problem were given a copy of the garden problem and asked to recall its solution procedure. Of these 11 subjects, 9 (82%) correctly recalled the LCM procedure.

Analogical transfer and accuracy. We have suggested that for multiple-solution problems, the relation between transfer and accuracy should depend on the relation between (a) the time required to solve the problem using the procedure indicative of transfer versus other procedures and (b) the amount of time allotted for the solution attempt. This hypothesis was tested for the band target problem, which subjects solved under conditions of little or no time pressure.

Using the LCM procedure to solve the band problem requires relatively straightforward adaptation of the procedure learned for the garden problem. Moreover, this procedure is more efficient than the alternative procedures because fewer multiples must be checked. Therefore, we expected that for subjects who solved the problem, those who used the LCM procedure would have shorter solution times than those who used some other procedure. This hypothesis was tested with a 4 (hint condition) × 2 (LCM vs. other procedures) analysis of variance (ANOVA) for subjects who solved the problem ($N = 45$). As predicted, LCM users required less time than did users of other procedures (6.2 vs. 8.8 min, respectively), $F(1, 37) = 4.55$, $p < .04$, $MS_e = 12.73$. Neither the main

---

[3] Full credit also was given to subjects who used the LCM procedure but made arithmetic errors. Common errors were to get either 110 or 130 when adding 24 to 96. The next number in the series would then be either 134 or 154, leading to 135 or 155 as the answer (rather than 145). These subjects also were coded as having gotten the correct answer to the problem.

[4] The criterion of two consecutive multiples other than the LCM was needed to eliminate subjects who produced the following sequence of numbers as part of their answer: 24, 48, 96. Most of these subjects clearly were computing an expanding series of numbers (i.e., LCM = 24, 24 × 2 = 48, 48 × 2 = 96) rather than multiples of 24. This procedure is incorrect.

[5] If the data are expected to conform to a *linear* trend, the choice of the appropriate contrast is clear. If the data (or the theory) are less constrained, such that one can only predict a *monotonic* trend, then the choice of appropriate contrast weights would seem to be less clear. One would like to choose from the infinite set of monotonically increasing weights the contrast that will maximize the power of detecting a monotonic trend, whatever its form. Fortunately, Abelson and Tukey (1963) have proved that there is a unique solution to this problem, and the appropriate weights (what they call the maximin contrast) for various numbers of groups are tabled in their paper. In addition, they note that "the 'linear-2-4' contrast, constructed from the usual linear contrast by quadrupling c_1 and c_n, and doubling c_2 and c_{-1}, is a reasonable approximation to the maximin contrast for small or medium n . . . Knowing *only* simple rank order for the μ_j, good practice seems to indicate the use of 'maximin' or 'linear-2-4' contrasts in careful work" (pp. 1347–1348). We used the linear-2-4 contrast.

[6] To rule out the hypothesis that our results solely reflect the fact that mathematically-competent students perform well on math-related tasks, we reran all analyses for both experiments with performance on the MSAT as a covariate. In no case for either experiment did adding this variable remove or change any of the relationships among the other variables that we report. For clarity and ease of exposition, we report the results of the analyses without MSAT as a covariate.

[7] An alternative explanation for the comparable performance of subjects in the retrieval and concept-mapping hint conditions is that subjects never compute the concept mappings because they are unnecessary for successful transfer. The spontaneous comments of several subjects in a pilot study call into question the adequacy of this explanation. In that study, subjects who failed to show transfer on the band problem after a retrieval hint were told the mapping between the concepts in the garden and band problems. When given that information, several of those subjects spontaneously made comments to the effect of, "I already know that. Why don't you tell me something I haven't already figured out."

effect of condition nor the condition by procedure interaction were reliable, both $p > .15$. Another way to assess the effect of transfer on solution time is to test for a monotonic trend across conditions in the time spent working on the problem, using the data from all subjects. This test revealed decreasing times across conditions, mirroring the increasing transfer rates: means of 10.77, 9.27, 9.12, and 6.70 min, respectively, $F(1, 71) = 7.99$, $p < .01$, $MS_e = 18.23$. The residual term was not reliable, $F(2, 71) < 1$.

Because it took subjects about 9 min to solve the problem using procedures other than LCM, and because 15 min were allotted for the problem, all subjects should have had enough time to solve the problem regardless of the procedure used. This lack of time pressure means that accuracy on the band problem should not be related to analogical transfer. That is, there is unlikely to be an increase in accuracy across conditions mirroring the increase in transfer rate. This hypothesis was tested using the monotonic trend contrast. As predicted, accuracy did not differ reliably across conditions ($M = 0.60$), $F(1, 71) < 1$, $MS_e = 0.25$. The residual term also was not reliable, $F(2, 71) < 1$.

Analogical transfer and schema induction. In this section, we consider the antecedents and consequences of schema induction, which was assessed by analyzing responses to the solution schema question.[8] Subjects' responses were coded for the presence or absence of each of the four steps in the LCM procedure. These data were then collapsed into three categories labelled good, intermediate, and poor schemas. A fairly liberal categorization scheme was used because we expected subjects' ability to verbalize complex solution procedures to be less developed than their ability to execute such procedures. Because finding the LCM of several divisors is the defining feature of the LCM procedure, subjects who mentioned that step and at least one other step were coded as having produced a good solution schema. Subjects who mentioned more than one step but did not mention finding the LCM were placed in the intermediate category. All other subjects were assigned to the category of poor schemas. This coding scheme had a reliability of .86. Poor, intermediate, and good schemas were produced by 39%, 29%, and 32% of subjects, respectively.

The hypothesis that analogical transfer leads to schema induction has two implications. First, the quality of the schema induced should be positively related to the strength of analogical transfer. Second, schema quality should be related to transfer but not to successful solution by nonanalogical means. Supporting the first implication, schema quality was positively related to the strength of analogical transfer for the target problem, $r = .52$, $p < .001$. To test the second implication, we conducted a multiple-regression analysis in which target transfer and accuracy were entered simultaneously as predictors of schema quality. As expected, the effect of transfer was reliable, $\beta = .48$, $t(72) = 4.26$, $p < .001$, $MS_e = 0.53$, but that of accuracy was not, $\beta = .07$, $t(72) = 0.61$, $p > .50$.

Figure 1 shows the dependence of schema quality on target transfer but not accuracy. The percent of subjects producing poor, intermediate, and good schemas (shaded bars) is graphed as a function of four categories of target performance (from right to left on the abscissa): successful analogical transfer, partial transfer, successful solution in the absence of transfer, and unsuccessful solution combined with no transfer. It is clear from the graph that schema quality increases with increasing strength of transfer. Moreover, the two groups of subjects who failed to show transfer ("other solve" and "fail to solve") do not differ with respect to schema quality. Finally, 13 of the 16 subjects who showed partial transfer failed to solve the problem, yet their pattern of schema quality does not resemble that of subjects who both failed to solve the problem and failed to show transfer.

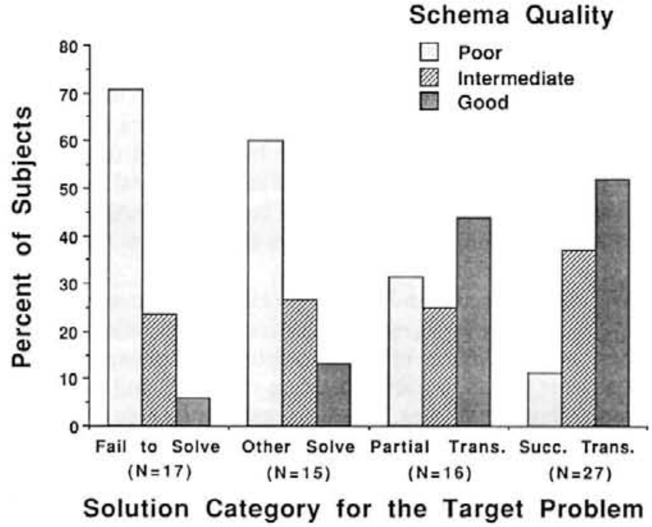

Figure 1. Percentage of subjects producing poor, intermediate, and good solution schemas as a function of performance on the target problem in Experiment 1.

Finally, we consider the consequences of schema induction. We hoped to replicate the suggestion in Ross and Kennedy's (1990) data that the effect of schema induction on later transfer is independent of any direct effect of prior transfer. To test this hypothesis, we examined the relation between transfer on the seashell generalization problem (scored as for the target problem) and both schema quality and target transfer. Seashell transfer was coded with a reliability of .97: 39% of the subjects showed successful transfer, 11% showed partial transfer, and 51% failed to show transfer.

Seashell transfer was reliably associated with both target transfer, $r = .30$, $p < .02$, and schema quality, $r = .43$, $p < .001$. The first relation, however, was largely redundant with the latter because when both target transfer and schema quality were entered simultaneously into a multiple regression, only the latter was a reliable predictor of seashell transfer: $\beta = .38$, $t(72) = 3.07$, $p < .01$, $MS_e = 0.74$, for schema quality, and $\beta = .10$, $t(72) = 0.81$, $p > .40$, for target transfer. The percentage of subjects showing (partial or complete) transfer

[8] Subjects' responses to the question about the similarities between the two problems (as opposed to the similarities between the solution procedures for the problems) were uninformative. Our impression was that subjects did not understand what we were trying to ask. We will therefore not present those data.

on the generalization problem increased with increasing schema quality: 20%, 57%, and 77%, respectively, for subjects who wrote poor, intermediate, and good schemas. In sum, not only does schema quality predict generalization-problem transfer independently of the variance attributable to target transfer (i.e., schema induction is not an epiphenomenon), but it is possible that the entire impact of initial transfer on subsequent transfer is mediated by the quality of the schema induced during the initial transfer episode.

Experiment 2

Experiment 2 had several objectives. First, we wished to provide additional evidence for our claim that numerical mapping is closer to adaptation than is conceptual mapping. Accordingly, all subjects received one of the two mapping hints. We also introduced two explicit mapping tasks to assess whether subjects had derived the appropriate conceptual or numerical mappings that they were not given explicitly. Performance on each mapping task was related to transfer success. To further illuminate the relation between mapping and adaptation, we report the results for a small sample of subjects who provided verbal protocols while solving the band problem.

A second goal of this experiment was to explore why transfer after a number-mapping hint in Experiment 1 was well below ceiling performance. We considered two explanations for this finding. To address the hypothesis that the below-ceiling performance was due to the difficulty of the adaptation process, we report on the nature of subjects' adaptation attempts for both experiments in a separate section after the Experiment 2 results. A second hypothesis is that successful transfer depends not only on knowing the appropriate correspondences between the source and target, but also on understanding the reasons behind them. Thus, we introduced an explanations manipulation in this experiment, which was crossed with the type of correspondences presented in the hint. For example, whereas subjects in the concept-hint condition were told that "The band members are like plants," subjects in the concept-hint-plus-explanations condition were told that "The two problems are both about putting objects into groups. The band members are like plants, because those are the objects being grouped in the problems."

Third, to further test the time-pressure hypothesis, subjects in this experiment were placed under considerable time pressure for the two target problems. Enough time was allotted for most LCM users to solve the problems, but not for most users of other procedures to do so. Given this speed pressure, the mapping manipulation should affect target accuracy as well as target transfer. That is, analogical transfer should increase the probability of correct solution because only those subjects who transfer the efficient LCM procedure will have enough time to solve the problems.

Fourth, to re-examine the relation between analogical transfer and schema induction in a slightly different context, we added a second target problem. Replication of the direct effect of schema induction on generalization-problem transfer in the face of two target problems that could be used instead would provide strong evidence for the importance of schema induction. We also wanted to determine whether target transfer would exert a direct effect on generalization-problem transfer when it was based on two problems rather than just one, because the absence of a direct effect of target transfer in Experiment 1 was somewhat surprising.

Finally, we examined expertise and analogical reasoning ability as predictors of analogical problem solving, a composite measure formed as the sum of transfer on the three test (target and generalization) problems. The individual-differences measures used as predictors were performance on the math and verbal sections of the SAT (specific measures of mathematical and verbal skill, respectively) and the verbal analogy section of the Differential Aptitude Tests (a general measure of analogical reasoning ability).

Method

Subjects. The subjects were 132 Vanderbilt University undergraduates (68 females and 64 males) who were paid $20 for their participation. Subjects were randomly assigned to one of four conditions: concept hint without explanations ($N = 34$), number hint without explanations ($N = 33$), concept hint with explanations ($N = 32$), and number hint with explanations ($N = 33$). An additional 14 subjects participated, but their data were excluded from the analyses: ten failed to understand the source problem's solution procedure, three did not feel well and left the experiment early, and one was not a native speaker of English (VSAT = 260).

Materials. In addition to the problems used previously, subjects also received the bake-sale target problem described in Table 1. It involved determining how many cookies a woman brought to a bake sale, given information about what happened when she tried putting various numbers of cookies in each bag. This problem is very similar in structure to the band problem (see Table 1).

Two mapping tasks were created for this experiment. For the number-mapping task, subjects were given the five numbers from the bake-sale problem (16, 14, 8, 6, and 9). Next to each they were to write the corresponding number from the garden (source) problem. For the concept-mapping task, subjects were given the four main concepts from the bake-sale problem, and they were to write the corresponding concept from the garden problem next to each. The four concepts were the same as those described in the concept hint: cookies, bags of cookies, number of cookies per bag, and number of cookies left for the last bag.

The solution-schema task was the same as that used in Experiment 1. Our measure of analogical reasoning ability was performance on the verbal analogy section of the Differential Aptitude Tests (Form W; Bennett, Seashore, & Wesman, 1982a). This multiple-choice test contains 50 five-alternative items, to be completed in 30 min. Each item presents the second and third terms of a proportional analogy (e.g., —— is to horses as worms are to ——). The answer alternatives present pairs of words to fill the first and fourth positions (e.g., hay/birds). The test was normed for 8th through 12th graders.

Design and procedure. This experiment used a 2 (conceptual vs. numerical mapping) × 2 (presence vs. absence of explanations for the mapping) between-subjects design. The mapping hints without explanations were essentially identical to those used in Experiment 1. The hints with explanations were also the same except that the mappings were given brief justifications. The exact wordings for the mapping correspondences and explanations for the band problem may be found in Appendix C. Hints for the bake-sale problem were constructed analogously.

The experiment began with the directed-solution task. The garden problem was the third of three problems. After this problem, subjects were given 20 min to work on a deductive-reasoning problem that was being used for another experiment. That problem was very different from all the materials used in this experiment and thus served as a filler problem.

The two target problems (band and bake sale) came after the deductive-reasoning problem. Because of the long delay between the target problems and the initial presentation of the source problem, subjects were given 4 min to review the garden problem and its solution procedure before working on the first target problem. (They were given clean copies of these materials, not the ones they had written on earlier.) Subjects were given 7 min to work on each target problem. This time was chosen by examining the distribution of solution times for the LCM solvers and other solvers in Experiment 1. In this amount of time, most LCM users were able to solve the band problem, but most users of other procedures were not. Each subject received the same type of hint for both target problems. As before, the hint was read out loud to subjects and also printed below the problem. All subjects received the band problem first and the bake-sale problem second.

After the bake-sale problem, subjects spent 3 min completing a mapping task for that problem. Subjects in the concept-hint conditions received the number-mapping task, whereas those in the number-hint conditions received the concept-mapping task. Then, after a 10-min break, subjects spent 4 min writing their solution schema. Next, they were given 9 min to work on the seashell problem (without a hint). After a 5-min break, subjects spent 30 min taking the verbal analogies test. SAT scores were provided by Vanderbilt University for all but 3 subjects (who declined to release their scores).[9] Subjects participated individually or in groups of 2 to 8.

Results and Discussion

Mapping and adaptation. Because the two target problems were very similar, and because the same hint was given for each, their transfer scores and accuracy scores were added to form composite measures of target transfer and accuracy. Preliminary analyses confirmed that the two problems behaved similarly in all analyses. Bake-sale transfer was coded with a reliability of .92.

To analyze the transfer data, we performed a 2 × 2 ANOVA with mapping hint (concept vs. number) and presence/absence of explanations as the independent variables. Replicating Experiment 1, the number-mapping hint led to greater transfer than did the concept-mapping hint (means of 2.30 and 1.35 out of 4, respectively), $F(1, 128) = 13.92$, $p < .001$, $MS_e = 2.20$. The absolute level of performance also supports our earlier conclusion that successful mapping is insufficient for transfer: 32% of the subjects who received a number-mapping hint did not succeed at transfer for even one target problem, even though they were told to use the source solution procedure. Neither the main effect of explanations nor the mapping by explanations interaction was reliable, both $p > .15$. The unexpected absence of an explanations effect suggests that subjects derived on their own the justifications we provided or that the justifications did not facilitate further mapping or adaptation.

The mapping tasks shed some light on the cause of the observed mapping-hint difference. One point was awarded for each of the four concepts that was mapped correctly in the concept-mapping task. For the number-mapping task, the five numbers fell into three categories: Divisors that left a remainder (three numbers), the common remainder, and the divisor that left no remainder. One point was awarded for each correctly-mapped category. (For the first category, subjects almost universally mapped all three numbers correctly.) Subjects did quite well on both tasks: The number-hint subjects had a mean of 78% correct on the concept-mapping task (s = 30%), and the concept-hint subjects had a mean of 90% correct for the number mappings (s = 23%). We could discern no systematicity in the errors. (The explanations manipulation did not affect performance.)

Because most concept-hint subjects knew the mapping between the numbers in the source and (second) target problems, the major benefit of providing those correspondences seems to have been to aid mapping and/or adaptation in some more indirect way, possibly by highlighting their importance or by helping subjects keep that information in mind while attempting adaptation. Nevertheless, there was a positive relation between ability to derive the numerical mapping on one's own and transfer on the target problems, $r = .25$, $p < .05$, $N = 66$. In contrast, performance on the concept-mapping task was unrelated to transfer, $r = .06$. These data provide additional evidence that the number mappings are more directly related to transfer success than are the conceptual mappings.

To gather further evidence concerning the importance of number mappings in the transfer of mathematical procedures, 8 UCLA undergraduates were asked to think out loud while using the garden problem to solve the band problem. When they experienced difficulty, the experimenter tried to ask questions that would help them figure out how to use the garden problem without explicitly telling them what to do. All but 2 subjects eventually were successful at transfer. The tape-recorded protocols were transcribed and studied for evidence of conceptual and numerical mapping.

Interestingly, subjects only occasionally verbalized the conceptual correspondences between the garden and band problems. There were only five instances of conceptual mappings (from the protocols of 4 subjects), and three of these cases also involved numerical mappings: for example, "... They have two extra spaces and then they have, for the band they have one extra space." It appears that the process of mapping the concepts occurred so quickly that usually it was not verbalized (cf. Ericsson & Simon, 1980).

In contrast, evidence for numerical mapping was abundant. All but 1 of the 8 subjects (the one who transferred most quickly) mentioned numerical correspondences, with 2–6 instances per subject ($M = 3.6$, $s = 1.8$). These mappings *always* occurred in the course of attempting to adapt the source solution to the target problem: for example, "In the garden problem they had [pause] initially they had, umm, they had the 20 they would have had something like the 24. And then they, I think they doubled it. So it would have been like [pause] 48." These data further support our claim that the numerical mapping is more directly related to adaptation than is the conceptual mapping.

---

[9] Vanderbilt requires students to take either the SAT or the ACT. Five subjects took the ACT only. SAT scores were estimated for them based on their ACT scores.

Analogical transfer and accuracy. Because subjects were placed under considerable time pressure for the target problems, we predicted that, contrary to Experiment 1, analogical transfer would benefit accuracy. Specifically, the effects of the manipulated variables on accuracy should mirror the effects of those variables on analogical transfer. The results were as predicted: Target accuracy was higher after a number-mapping than a concept-mapping hint, with means of 1.09 and 0.76 (out of 2), respectively, $F(1, 128) = 6.62$, $p < .02$, $MS_e = 0.55$. Neither the main effect of explanations nor the mapping by explanations interaction was reliable, both $p > .25$.

Analogical transfer and schema induction. As in Experiment 1, we tested the hypothesis that schema quality is positively associated with strength of analogical transfer but not associated with correct solution by nonanalogical means. Poor, intermediate, and good schemas were produced by 27%, 30%, and 43% of subjects, respectively. Analogical transfer and accuracy on the target problems were entered simultaneously into a multiple regression as predictors of schema quality. Replicating Experiment 1, target transfer was a reliable predictor, $\beta = .42$, $t(129) = 3.09$, $p < .001$, $MS_e = 0.61$, but accuracy was not, $\beta = -.14$, $t(129) = -1.04$, $p > .25$.

On the seashell generalization problem, 51% of the subjects showed successful transfer, 11% showed partial transfer, and 39% showed no transfer. As before, we expected that schema quality would make a unique contribution to transfer on this problem, over and above any contribution of transfer on the target problems. The results of a simultaneous multiple regression analysis indicated that both variables were reliable predictors of generalization-problem transfer: $\beta = .18$, $t(129) = 2.34$, $p < .02$, $MS_e = 0.61$, for schema quality, and $\beta = .49$, $t(129) = 6.37$, $p < .001$, for target transfer. Thus, unlike in Experiment 1, schema quality was not the sole predictor of systematic variance in transfer accuracy. Nonetheless, schema quality clearly had an independent impact on transfer to the generalization problem. The independent contribution of target-problem transfer may reflect the availability of an additional analogue to aid mapping and/or adaptation.

A graph of mean seashell transfer scores as a function of target transfer and schema quality clarifies the relation among these variables (see Figure 2): Either successful transfer on at least one target problem (scores of at least 2) or a good solution schema was required for at least partial transfer on the generalization problem (score of 1), with little added benefit for good performance on both predictor variables.[10] That is, schema quality primarily had its effect at low levels of target transfer, whereas target transfer was most important for subjects with poor or intermediate schemas.

Individual differences. Because 3 subjects declined to release their SAT scores, the individual-differences analyses are based on only 129 subjects. Means, standard deviations, score ranges, and intercorrelations for the verbal analogies test, VSAT, and MSAT are shown in Table 2. Performance on each of the tests was similar across conditions. The mean verbal analogy score falls at about the 90th percentile on the 12th-grade norms (Bennett, Seashore, & Wesman, 1982b). As expected, scores on the three standardized tests are moderately correlated (range of .39–.47). Table 2 also shows data for our

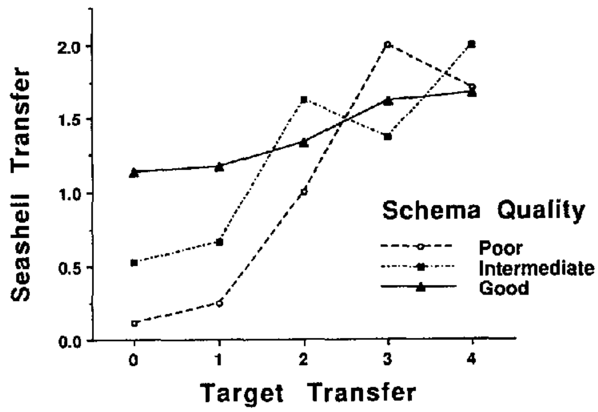

Figure 2. Mean transfer scores on the seashell generalization problem as a function of transfer on the target problems and schema quality in Experiment 2.

measure of analogical problem solving, which was computed as the sum of analogical transfer on the band, bake-sale, and seashell problems (scored 0, 1, 2 for each problem). The Cronbach's alpha reliability of this measure is .73. This measure of mathematical analogy skill has a reliable correlation only with performance on the MSAT, $r = .33$, $p < .001$.

To more clearly assess the relation between (a) analogical reasoning ability and verbal and mathematical expertise and (b) analogical transfer in math, the three test scores were entered into a simultaneous multiple regression as predictors of transfer. The two manipulated variables (type of mapping hint and presence/absence of explanations) also were included in the equation. Only math SAT and type of mapping hint reliably predicted transfer, $\beta = .33$, $t(123) = 3.49$, $p < .001$, $MS_e = 3.97$, and $\beta = .25$, $t(123) = 3.00$, $p < .001$, respectively (verbal SAT: $\beta = -.15$, $t(123) = -1.54$, $p > .12$; verbal analogy test: $\beta = .11$, $t(123) = 1.13$, $p > .25$; and explanations: $\beta = -.15$, $t(123) = -1.78$, $p > .08$). In sum, the best predictors of solving math word problems by analogy appear to be expertise in math and knowledge of the correspondences required for successful procedure adaptation.

The Importance of Adaptation in the Transfer of Mathematical Procedures

Conceptual Analysis and Predictions

Although it might seem rather trivial to "do the same things" with the numbers in the band and bake-sale problems as were done with the numbers in the garden problem, once the numerical correspondences have been identified, our data indicate otherwise. To provide a more formal understanding of the importance and difficulty of adaptation in the transfer of mathematical procedures than we have so far, we will discuss three general types of adaptations that seem

[10] We should note that the data points in Figure 2 are based on between 2 and 19 subjects. The most discrepant points in terms of the relations expressed (i.e., the two highest points) are based on only two to three subjects.

Table 2
Descriptive Data for the General and Specific Ability Measures and for the Measure of Analogical Transfer for Math Word Problems (Experiment 2)

	Intercorrelations			Descriptive Measures		
	Analogies	VSAT	MSAT	*M*	*SD*	Range
Verbal Analogies	—	—	—	44.7	4.2	28–50
Verbal SAT	.47*	—	—	572	77	420–780
Math SAT	.39*	.43*	—	625	76	410–780
Analogical Transfer	.15	.00	.33*	2.9	2.2	0–6

Note. Starred correlations are statistically reliable at $p < .001$. All other correlations are not reliably different from zero ($p < .05$).

relevant for our problems. This is not intended to be an exhaustive taxonomy of adaptations, because research using any particular set of problems cannot hope to illuminate all aspects of this process. The three categories, as well as specific examples relevant to our problems, are presented in Table 3.

The simplest type of adaptation consists of substituting numbers from the test problem into the operators learned for the source problem. For example, to solve the band problem, subjects must substitute "1" for "2" in the add-remainder operator. Substitution provides two potential errors for our problems: (a) finding the LCM of the wrong numbers and (b) adding a number other than the constant remainder to the LCM multiples. We do not expect this type of adaptation to greatly impede transfer for any of our three test problems (target and generalization) because it is not a major source of difficulty for algebra word problems (Reed, Dempster, & Ettinger, 1985).

Table 3
Types of Procedure Adaptations Required for Successful Transfer of a Mathematical Procedure

1. Substitute numbers from the test problem into the source operators.
   (a) Find the LCM of the wrong numbers (error).
   (b) Add the wrong number to the LCM multiples in place of the constant remainder (error).
2. Postulate new test-problem elements not described in that problem.

	Error Expected if Subjects:	
Source Operator	Fail to Adapt Operator	Adapt Operator Incorrectly
(a) Find-LCM	—	Wrong number as the LCM
(b) Compute-multiples	LCM (LCM + r) only	Multiples of LCM + r, expanding series
(c) Add-remainder	Fail to add remainder	Subtract remainder
(d) Select-answer	All relevant info but ...	—

3. Generalize source procedure in ways that preserve the essential structure of the procedure.
   (a) Extend the number of multiples examined for the test problems beyond number needed to solve the garden source problem.
   (b) Take account of the altered (range) solution constraint for the band problem.
   (c) Take account of an additional (fourth) argument for the find-LCM operator for the seashell problem.
   (d) Take account of the extra divisor in the seashell problem with the unique, nonzero remainder.

A second type of adaptation involves postulating new test-problem elements that were not described in that problem and hence could not be mapped to the source. Once these new elements are created, they can be mapped directly onto elements in the source problem or solution. This is where we expect most of the adaptation difficulty to occur for our problems, because all steps of the LCM procedure except the last involve the creation of new elements. Table 3 lists the four steps of the LCM procedure (find-LCM, compute-multiples, add-remainder, and select-answer) and the type of error that would be expected if subjects failed to adapt or incorrectly adapted each operator. Subjects who fail to execute the first step are coded as not having attempted adaptation. Numerous pitfalls await those who do attempt adaptation. They may fail to compute multiples of the LCM, stopping their adaptation attempt either after finding the LCM or after adding the constant remainder to the LCM ("LCM (LCM+r) only"). They may compute the multiples but "fail to add remainder." Or, they may fail to select an answer from among the list of remainder-corrected multiples generated, even though the correct answer is in that list of numbers ("all relevant info but . . ."). These errors reflect failures to attempt adaptation of the operators.

Subjects also may attempt adaptation but do so incorrectly. They may identify the "wrong number as the LCM" (e.g., 48 for the band problem). They may incorrectly compute multiples of the LCM, for example, by computing multiples of the LCM plus the remainder (e.g., multiples of 25 for the band problem; "multiples of LCM+r") or by computing an expanding series of number (e.g., 24, 48, 96 instead of 24, 48, 72 for the band problem; "expanding series"). Finally, they may subtract rather than add the constant remainder ("subtract remainder").

A third type of adaptation involves generalizing the source procedure in ways that nevertheless preserve the essential structure of the procedure. Four such adaptations are required for our problems (also see Table 1). First, subjects must extend the number of LCM multiples examined beyond the two needed for the garden problem. This adaptation is likely to be most difficult for the band problem, because it is solved first and it requires the greatest extension. Its answer is based on the sixth multiple, compared to the third and fourth multiples for the bake-sale and seashell problems, respectively. Second, the solution constraint in the garden problem of finding the smallest possible number must be adapted for the band problem to finding the number that falls within a given range. This adaptation is important because there is a number smaller than the correct answer that would work except that it is outside the indicated range (namely 25). The seashell problem also has the range constraint, but both constraints yield the same answer. The final two adaptations are relevant only for the seashell problem. In that problem, subjects must adapt the find-LCM operator to work on four numbers rather than the three required for the other problems. In addition, subjects must determine what to do with the extra divisor that leaves a nonzero remainder that is different from the nonzero remainder left by the other divisors that leave a remainder.

Error Data

Based on the conceptual analysis presented in Table 3, we coded the types of adaptation errors subjects made for the target and generalization problems in both experiments. If anything, these codings may underestimate the difficulty of adaptation. For the 8 protocol subjects, we independently coded adaptation errors for the band problem based on the oral and written protocols. We identified 13 errors from the oral protocols. Ten (77%) of these also were evident in the written protocols. No errors were coded exclusively from the written protocols.

It is clear that the adaptation process was a major source of difficulty for transfer. Collapsing across the various errors, 23% of the subjects in Experiment 2 who eventually showed successful transfer for the band problem initially made adaptation errors, as did 36% of the subjects who were never successful. (The remaining subjects gave no evidence of attempting adaptation of the LCM procedure.) Averaged across five test problems in two experiments, these percentages are 24% and 44%, respectively. The last three lines of Table 4 show the percentages of subjects for each problem in each experiment who attempted adaptation, who made errors given that they attempted adaptation, and who were successful given that they attempted adaptation.

The body of Table 4 shows for each problem (and each experiment) the percentage of the errors coded that fell into each category. Table 4 is organized and labeled in accordance with the conceptual analysis presented in Table 3. The distributions of errors for the successful and unsuccessful subjects for each problem were similar, as were the distributions for subjects who received different hints. Therefore, the data from all subjects who made errors for a particular problem are considered together. Although most subjects made only a single error, a few subjects for each problem made multiple errors. Therefore, the numbers in the table represent the percentage of all errors for each problem that were coded in each category.

As expected, subjects had little difficulty substituting the test-problem numbers into the source operators. Substitution errors for the find-LCM operator occurred too infrequently (if at all) to count. Such errors for the add-remainder operator occurred with nontrivial frequency for only one problem in one experiment. Averaged across the five sets of data (three problems in Experiment 2 and two problems in Experiment 1), only 3% of the errors involved incorrect substitution.

In contrast to the ease of substitution, the postulation of new problem elements—the real heart of adaptation for our problems—was very difficult, accounting for 86% of all adaptation errors. On average, 31% of subjects' errors reflected failure to attempt adaptation of one of the required operators, particularly compute-multiples and add-remainder. The majority of the errors (55%) consisted of incorrect attempts to adapt the operators, particularly find-LCM and compute-multiples. The greater frequency of incorrect operator adaptations compared to failures to attempt adaptation suggests that students understand the general goal of analogical transfer, which is to adapt a solution procedure. As indicated in Table 3, two types of errors were coded for incorrect adaptation of the compute-multiples operator. Computing multiples of LCM + r was more common than computing an expanding series, accounting for approximately 75% of the total errors in this category. Examining the data from the perspective of the operators, compute-multiples was the most difficult to adapt successfully, as it was associated with 37% of all errors.

Although we described four types of generalizations in the third category of adaptations (see Table 3), only the first was a significant source of difficulty for subjects. Extending the number of multiples examined for the band problem accounted for 21% of the errors for that problem (on average). Recall that the answer to the garden problem was based on the second multiple. In the description of the solution to that problem that subjects received, four multiples were listed (see Appendix A). Of the subjects who failed to generate enough multiples, 88% generated either three or four multiples. Only a few subjects had difficulty with this adaptation for the bake-sale and seashell problems, whose solutions were based on the third and fourth multiples, respectively. Adapting the solution constraint from find-fewest to select-from-range, which was necessary only for the band problem, was not as difficult for

Table 4
Percentage of All Adaptation Errors Falling Into Each Error Category for Each Problem in Experiment 2

Adaptation error	Problem: Band target	Bake-sale target	Seashell generalization	*M*
1. Substitute test-problem numbers into source operators				
(b) + wrong remainder	0.0 (0.0)	2.7	11.3 (0.0)	2.8
2.1. Postulate new elements: Fail to adapt source operator				
(b) LCM or LCM + r only	10.2 (10.8)	20.5	7.5 (19.0)	13.6
(c) Fail to add remainder	18.4 (10.8)	8.2	13.2 (14.3)	13.0
(d) All relevant info but. . .	4.1 (0.0)	8.2	7.5 (0.0)	4.0
Total	32.7 (21.6)	36.9	28.2 (33.3)	30.5
2.2. Postulate new elements: Adapt source operator incorrectly				
(a) Wrong number as LCM	8.2 (13.5)	38.4	28.3 (28.6)	23.4
(b) Incorrect multiples	38.8 (24.3)	16.5	22.7 (14.3)	23.3
(c) Subtract remainder	6.1 (5.4)	1.4	5.7 (23.8)	8.5
Total	53.1 (43.2)	56.3	56.7 (66.7)	55.2
3. Generalize source procedure in structure-preserving ways				
(a) Not enough multiples	10.2 (32.4)	4.1	3.8 (0.0)	10.1
(b) Forget range constraint	4.1 (2.7)	—	— (—)	1.4
Total	14.3 (35.1)	4.1	3.8 (0.0)	11.5
% of Ss attempting adaptation	65.2 (68.0)	71.2	74.2 (56.0)	66.9
% of those Ss who made errors	45.3 (64.7)	68.1	46.9 (42.9)	53.6
% who succeeded	69.8 (52.9)	40.4	69.4 (69.1)	60.3

Note. Percentages for Experiment 1 are in parentheses.

subjects. For the two experiments combined, only 3 of 207 subjects gave 25 as the answer to the band problem. In examining subjects' solutions to the seashell problem, we could find no solution attempts that seemed to indicate difficulty in taking account of the additional divisor with the same remainder as the other divisors or the extra divisor with the unique, nonzero remainder. It is likely, however, that these adaptations contributed to the greater amount of time required to use the LCM procedure to solve the seashell problem: In Experiment 1, in which there was no time pressure, subjects required 8.9 min to transfer the LCM procedure to the seashell problem, compared to 6.2 min for the band problem.

One obvious remaining question concerns predicting which subjects will make which errors. For example, one might expect performance on the MSAT to be reliably (negatively) associated with one or more types of adaptation errors. Surprisingly, we found no replicable effects of mathematical expertise on the nature of subjects' errors. In the interest of space, we will refrain from reporting these analyses. We will return to this issue near the end of the General Discussion.

General Discussion

The research reported here explored several issues concerning the component processes of analogical problem solving, the role of schema induction in analogical transfer, the impact of time pressure on the relation between analogical transfer and accuracy, and the basis for individual differences in analogical problem solving with mathematical word problems. We will review the major findings of our two experiments and verbal-protocol study and discuss their implications for theories of analogical transfer and for instruction.

The Mapping and Adaptation Components of Analogical Problem Solving

A primary goal of this research was to distinguish the mapping and adaptation processes of analogical transfer. Previous work has tended to treat these processes as unitary, implying that successful mapping more or less guarantees successful transfer. In contrast, we argued that mapping and procedure adaptation are separate, although related, processes. In particular, knowledge of the numerical correspondences between the source and target problems would seem to be required for successful adaptation of a mathematical procedure because the numbers serve as arguments of the solution operators. Although concept mappings may help define the numerical mappings, only the latter are directly required to transfer the solution procedure. In sum, we agree that mapping is necessary for transfer but disagree that it is sufficient. Even with a successful mapping, difficulty in adapting the source solution procedure to work for the target problem may impede transfer.

The relation between mapping and adaptation. Our data support our hypothesis that number mapping is more closely associated with successful adaptation than is concept mapping. First, in both experiments, subjects who were told the numerical mapping between the source and target (number-mapping hint) were more successful at transfer than were those who received a concept-mapping hint. Second, for the (concept-hint) subjects in Experiment 2 who completed the number-mapping task, mapping success was positively correlated with transfer. In contrast, success on the concept-mapping task (completed by the number-hint subjects) was unrelated to transfer. Finally, 7 of 8 subjects who provided verbal protocols while solving the target problem by analogy to the source problem explicitly mentioned number mappings ($M = 3.6$ per subject), and these correspondences were *always* stated in an effort to adapt the source procedure.

Because most of the concept-hint subjects in Experiment 2 were able to derive the numerical correspondences on their own, it appears that the benefit of the number-mapping hint was more indirect. It may have helped subjects keep the numerical correspondences in mind while adapting the analogous solution; it may have increased subjects' confidence in the accuracy of the correspondences they derived and hence increased the probability that they would persevere in their adaptation attempts; or it may have focused subjects' attention on those aspects of the mapping most crucial for adaptation (and, therefore, successful transfer). It is possible, of course, that some of the concept-hint subjects in Experiment 2 failed to compute the numerical mapping during solution of the target problems, doing so only afterwards in response to our request for that information. Although we cannot rule out this possibility, we tried to minimize its likelihood by limiting the time allotted to complete the mapping task.

Adaptation. It is clear that numerical mapping and adaptation are related but quite distinct. If adaptation were simply an automatic consequence of successful mapping, then subjects who knew the mapping of the numbers should have been able to "do the same things" with the numbers in the target problem(s) as were done with the numbers in the source problem. Yet in both experiments, the transfer performance of subjects who were told the numerical mapping between the source and target problems was far from ceiling: 50% were unsuccessful in Experiment 1, and 32% were unsuccessful in Experiment 2.

Our codings of the nature of subjects' analogical solution attempts provide more direct support for our hypothesis that adaptation is a major source of difficulty in analogical problem solving with math word problems. Averaged across the five target and generalization problems in the two experiments combined, 24% of the subjects who eventually showed successful transfer initially made errors. Overall, 54% of the subjects who attempted adaptation made errors.

To better understand the adaptation process, subjects' transfer errors were grouped into three categories representing distinct types of adaptations important for mathematical analogies. The first type of adaptation consists of substituting numbers from the test problem into the operators used to solve the source problem. This type of adaptation was not a major source of transfer difficulty, as only 3% of all adaptation errors involved incorrect substitution (also see Reed et al., 1985).

The second type of adaptation involves postulating new test-problem elements that were not described in that problem and hence could not be mapped to the source problem. Once created, these new elements must be mapped onto elements of the source problem or solution. As expected, most (86%) of the adaptation errors fell into this category. To illuminate the nature of subjects' difficulties, we cross-classified the errors by the particular source operator to be adapted (find-LCM, compute-multiples, add-remainder, select-answer) and by whether subjects failed to adapt the operator or adapted the operator incorrectly. Subjects were more likely to incorrectly adapt the solution operators than to fail to attempt adaptation of the operators (55% vs. 31% of all errors fell into these two categories, respectively), suggesting that they understood the general goal of procedural transfer. Of the four operators required to execute the LCM procedure, subjects clearly had the most difficulty with compute-multiplies, as it accounted for 37% of all errors (across problems and experiments). The find-LCM, add-remainder, and select-answer operators accounted for 23%, 22%, and 4% of all errors, respectively. Given these data, it is no wonder the oral protocols collected from our 8 protocol subjects consisted largely of laborious, error-prone attempts to adapt the source solution to fit the target problem.

The third type of adaptation consists of generalizing the source procedure in ways that nevertheless preserve the essential structure of the procedure. For our materials, this translates into adaptations that preserve execution of the four operators identified earlier in essentially the same way (and same order) they were executed for the source problem. On average these adaptations were much less difficult than figuring out how to use the operators to generate the required new problem elements (the second category of adaptations), accounting for 11% of all errors. Their difficulty, however, was much more problem dependent than for the other adaptations. For example, based on the structures of our problems, the "extend-multiples" adaptation was predicted to be most problematic for the band target problem; and in fact it accounted for 21% of the errors for that problem, on par with the difficulty of adapting the operators themselves.

In sum, although the exact frequencies of particular types of adaptation errors undoubtedly will depend on the type of procedure being adapted, our results clearly indicate that adaptation is a very error-prone process and one that is conceptually quite distinct from the process of mapping. Adaptation is a major locus of transfer difficulty.

Analogical Transfer and Schema Induction

Previous researchers (e.g., Anderson & Thompson, 1989; Holyoak, 1984, 1985; Ross, 1989b; Ross & Kennedy, 1990) have suggested that analogical transfer leads to induction of a general schema encompassing the source and target problems. This hypothesis is important because schemas are an important theoretical construct in problem solving (e.g., Gick, 1986; Medin & Ross, 1989). They are thought to underlie experts' ability to categorize problems based on their structural features (e.g., Chi, Feltovich, & Glaser, 1981; Schoenfeld & Herrmann, 1982) and to remember large amounts of meaningful information in their domain of expertise after only a brief presentation (e.g., Chase & Simon, 1973; Egan & Schwartz, 1979). Thus, if analogical transfer leads to schema induction, it may also be important for the development of expertise (cf. Ross, 1989b). Several studies have shown that provision of multiple source analogues, coupled with explicit instructions to compare them, enhances schema induction and subsequent analogical transfer (Catrambone & Holyoak, 1989; Gick & Holyoak, 1983). The data on schema induction as a side-effect of solving problems by analogy are equivocal, however, as Ross and Kennedy (1990) found support for the hypothesized link, but Reed (1989) did not.

We attempted to clarify the relation between analogical transfer and schema induction by measuring the latter construct directly, rather than indirectly through its hypothesized effects. Our measure of schema quality was based on an analysis of subjects' descriptions of the solution procedure common to the source and target problems. In both experiments, not only was schema quality positively related to the strength of analogical transfer on the target problem(s), but it was uniquely related to transfer and not to correct solution by nonanalogical means. Although some subjects may have induced the schema when asked to do so rather than as a consequence of successful transfer, we tried to minimize this occurrence by limiting the time allotted for the schema induction task to several minutes less than the time required to show transfer for the target problems.

Our results further demonstrate that schema induction is an important consequence of analogical transfer. In both experiments, schema quality was a reliable predictor of transfer on the subsequent generalization problem, and its contribution was statistically independent of the effect of target transfer per se. Indeed, in Experiment 1 target transfer was not a reliable predictor of generalization-problem transfer after the effect of schema quality was taken into account. In Experiment 2, in which target transfer was based on performance on two problems, that variable did contribute reliably (and substantially) to the prediction of generalization-problem transfer. In that experiment, it appeared that either a good schema or transfer on at least one of the two target problems was required for making progress in attempting to adapt the LCM procedure for use with the generalization problem.

In sum, it appears that a major consequence of analogical transfer is the induction of more abstract knowledge about a class of problems, which in turn facilitates more flexible subsequent transfer. It is important to note, however, that when multiple example problems are available, the induced schema does not appear to supplant the individual exemplars. Rather, the abstract and specific forms of knowledge coexist, and both may be called upon to facilitate later problem solving. This finding supports arguments outlined by Medin and Ross (1989).

The Relation Between Analogical Transfer and Accuracy

In the introduction, we indicated that the relation between analogical transfer and accuracy ought to be complex for problems (like our own) that can be solved by any of several procedures (also see Bassok & Holyoak, 1989; Novick, 1988a). Degree of time pressure was hypothesized to be an important factor in mediating the relation between transfer and accuracy. If the solution procedure illustrated for the source problem can be adapted to fit the target problem, and it is relatively efficient, then analogical transfer should decrease solution time. But if other less efficient methods also are available, and subjects are not under time pressure, then no benefit is to be expected for accuracy. Both of these predictions were supported by the results of Experiment 1. Under time pressure, however, greater analogical transfer should be associated with higher solution rates because only solvers who use the source problem's efficient solution procedure will have enough time to solve the target problem. The results of Experiment 2 supported this prediction.

Analogical Reasoning Ability, Analogical Problem Solving, and Expertise

The final issue addressed in our research (Experiment 2) concerned predicting individual differences in success at analogical problem solving with math word problems (as defined by the sum of analogical transfer on the target and generalization problems). Two classes of individual-differences measures were considered as predictors: specific measures of expertise versus a general measure of analogical reasoning ability. Performance on the mathematics section of the SAT served as a measure of mathematical expertise for the domain tapped by our experimental problems. Similarly, verbal SAT scores reflected subjects' skill in the verbal domain. Analogical reasoning ability was assessed by performance on the verbal analogy section of the Differential Aptitude Tests. The results were clear: The measure of mathematical expertise was a reliable predictor of analogical transfer, but the measures of verbal skill and of general analogical reasoning ability were not. These results extend those of Novick (1988a) by indicating that mathematical expertise is an important predictor of transfer even when solvers are not required to retrieve the source problem themselves. In sum, the best predictors of analogical transfer for our problems were mathematical expertise and knowledge of the numerical correspondences required for successful procedure adaptation.

There are several theoretical reasons for expecting mathematical expertise to predict transfer of a mathematical procedure, as we observed. First, more expert subjects are more likely to represent the problems in ways that will provide potential retrieval cues (see Novick, 1988a) and aid in mapping. Furthermore, although (virtually) none of our subjects knew the complex LCM procedure prior to the experiment, it is likely that more mathematically sophisticated subjects had better mastery of its constituent operators. As a result, they would be better able to hold the relevant procedural knowledge in working memory while performing the integration necessary for adaptation. Our finding that performance on the MSAT was not reliably associated with particular adaptation errors, but was associated with overall success at adaptation, is consistent with this notion of expertise facilitating coordination and adaptation of a complex multistep procedure.

In contrast, it is less theoretically apparent why a measure of general analogical reasoning ability should fail to predict analogical problem solving, particularly given that a verbal measure was used to predict performance on verbal problems. Analogy problems were placed on mental tests because analogical reasoning is thought to be pervasive in everyday life (e.g., Spearman, 1923), and one analogical reasoning measure ought to predict another (e.g., Holyoak, 1984). Clearly, we cannot argue from our negative result that analogical reasoning ability as measured by psychometric tests is unrelated to analogical reasoning in the real world. Nonetheless, the observed predictive failure provides reason to reassess the relations among different measures of analogical reasoning.

There are several differences between the apparent requirements for success on the psychometric items versus analogical transfer: (a) necessity of retrieving the source domain, (b) types of knowledge required for good performance, (c) complexity of the mapping required, and (d) complexity of the adaptation process. Given our methods and results, the second and fourth differences would seem to be most important. Although general reasoning skills that can be applied to analogical problem solving in math may well exist, nontraditional types of psychometric tests may be required to tap the relevant aspects of analogical reasoning.

Methodological and Instructional Implications of the Present Findings

The separation between mapping and adaptation achieved in the present study has important implications for future investigations of analogical transfer. The ability to construct an analogous solution to target problems has often been treated as a direct measure of subjects' ability to map the source and target (e.g., Gick & Holyoak, 1980). The present results demonstrate, however, that when transfer requires adaptation of a complex multistep procedure, subjects may perform well on an explicit mapping test and yet be unable to derive the analogous solution. Thus the provision of an explicit test of subjects' knowledge of element correspondences would seem to be a desirable addition to future studies of analogical transfer.

Our findings also may have important implications for instruction. In particular, direct instruction in adapting a solution, perhaps focusing on the different types of adaptation noted above, might prove useful. That is, the process of adaptation can be conceptualized as a set of metaprocedures that potentially can be taught so as to improve students' ability to perform analogical transfer across a wide range of mathematical (and perhaps nonmathematical) problem types. In addition, providing partial information about the mappings between numbers crucial to successful adaptation might aid in teaching adaptation procedures for mathematical word problems. It is possible that other types of mapping hints (e.g., noting correspondences between numerical expressions or elements of diagrams) may prove useful for other types of mathematical (or nonmathematical) problems.

References

Abelson, R. P., & Tukey, J. W. (1963). Efficient utilization of non-numerical information in quantitative analysis: General theory and the case of simple order. *Annals of Mathematical Statistics, 34,* 1347–1369.

Anderson, J. R., & Thompson, R. (1989). Use of analogy in a production system architecture. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 267–297). Cambridge: Cambridge University Press.

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

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1982a). *Differential aptitude tests.* The Psychological Corporation. New York: Harcourt Brace Jovanovich.

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1982b). *Differential aptitude tests: Administrator's handbook.* The Psychological Corporation. New York: Harcourt Brace Jovanovich.

Burstein, M. H. (1986). A model of learning by incremental analogical reasoning and debugging. In R. Michalski, J. G. Carbonell & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (Vol. 2, pp. 351–370). Los Altos, CA: Morgan Kaufmann.

Carbonell, J. G. (1983). Learning by analogy: Formulating and generalizing plans from past experience. In R. S. Michalski, J. G. Carbonell & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (pp. 137–161). Palo Alto, CA: Tioga Publishing Co.

Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1147–1156.

Chase, W. G., & Simon H. A. (1973). Perception in chess. *Cognitive Psychology, 4,* 55–81.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5,* 121–152.

College Entrance Examination Board. (1986). *Taking the SAT.* Princeton, NJ: Author.

Corcoran, C. L., Gaughan, E. D., Ladd, N. E., & Salem, J. (1981). *Algebra: First course.* Glenview, IL: Scott, Foresman.

Egan, D. E., & Schwartz, B. J. (1979). Chunking in the recall of symbolic drawings. *Memory & Cognition, 7,* 149–158.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87,* 215–251.

Gentner, D. (1983). Structure-mapping: A theoretical framework. *Cognitive Science, 7,* 155–170.

Gholson, B., Morgan, D., & Dattel, A. R. (1990). The development of analogical problem solving: Strategic processes in schema acquisition and transfer. In D. F. Bjorklund (Ed.), *Children's strategies: Contemporary views of cognitive development.* Hillsdale, NJ: Erlbaum.

Gick, M. L. (1986). Problem-solving strategies. *Educational Psychologist, 21,* 99–120.

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

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

Hofstadter, D. R., & Mitchell, M. (1988). Conceptual slippage and mapping: A report on the Copycat project. In *Proceedings of the Tenth Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. (1986). *Induction: Processes of inference, learning, and discovery.* Cambridge, MA: MIT Press.

Holyoak, K. J. (1984). Analogical thinking and human intelligence. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2, pp. 199–230). Hillsdale, NJ: Erlbaum.

Holyoak, K. J. (1985). The pragmatics of analogical transfer. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 19, pp. 59–87). San Diego, CA: Academic Press.

Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition, 15,* 332–340.

Holyoak, K. J., & Thagard, P. (1989a). Analogical mapping by constraint satisfaction. *Cognitive Science, 13,* 295–355.

Holyoak, K. J., & Thagard, P. (1989b). A computational model of analogical problem solving. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 242–266). Cambridge, England: Cambridge University Press.

Keane, M. (1988). *Analogical problem solving.* Chichester, England: Ellis Horwood.

Medin, D. L., & Ross, B. H. (1989). The specific character of abstract thought: Categorization, problem solving, and induction. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 5, pp. 189–223). Hillsdale, NJ: Erlbaum.

Messick, S. (1980). *The effectiveness of coaching for the SAT: Review and reanalysis of research from the fifties to the FTC.* Princeton, NJ: Educational Testing Service.

Novick, L. R. (1988a). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 510–520.

Novick, L. R. (1988b). Analogical transfer: Processes and individual differences. In D. H. Helman (Ed.), *Analogical reasoning* (pp. 125–145). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Pirolli, P. L., & Anderson, J. R. (1985). The role of learning from examples in the acquisition of recursive programming skills. *Canadian Journal of Psychology, 39,* 240–272.

Posner, M. I. (1973). *Cognition: An introduction.* Glenview, IL: Scott, Foresman and Company.

Reed, S. K. (1987). A structure-mapping model for word problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 124–139.

Reed, S. K. (1989). Constraints on the abstraction of solutions. *Journal of Educational Psychology, 81,* 532–540.

Reed, S. K., Dempster, A., & Ettinger, M. (1985). Usefulness of analogous solutions for solving algebra word problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 106–125.

Reed, S. K., Ernst, G. W., & Banerji, R. (1974). The role of analogy in transfer between similar problem states. *Cognitive Psychology, 6,* 436–450.

Ross, B. H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 629–639.

Ross, B. H. (1989a). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 456–468.

Ross, B. H. (1989b). Remindings in learning and instruction. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 438–469). Cambridge, England: Cambridge University Press.

Ross, B. H., & Kennedy, P. T. (1990). Generalizing from the use of earlier examples in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 42–55.

Schoenfeld, A. H., & Herrmann, D. J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 484–494.

Spearman, C. (1923). *The nature of 'intelligence' and the principles of cognition.* London: Macmillan.

Spencer, R. M., & Weisberg, R. W. (1986). Context-dependent effects on analogical transfer. *Memory & Cognition, 14,* 442–449.
Sternberg, R. J. (1977). *Intelligence, information processing, and analogical reasoning.* Hillsdale, NJ: Erlbaum.
Wickelgren, W. A. (1974). *How to solve problems.* San Francisco: Freeman.

Appendix A

Experimental Problems

Garden Source Problem

Problem

Mr. and Mrs. Renshaw were planning how to arrange vegetable plants in their new garden. They agreed on the total number of plants to buy, but not on how many of each kind to get. Mr. Renshaw wanted to have a few kinds of vegetables and ten of each kind. Mrs. Renshaw wanted more different kinds of vegetables, so she suggested having only four of each kind. Mr. Renshaw didn't like that because if some of the plants died, there wouldn't be very many left of each kind. So they agreed to have five of each vegetable. But then their daughter pointed out that there was room in the garden for two more plants, although then there wouldn't be the same number of each kind of vegetable. To remedy this, she suggested buying six of each vegetable. Everyone was satisfied with this plan. Given this information, what is the fewest number of vegetable plants the Renshaws could have in their garden?

Solution Procedure

Since at the beginning Mr. and Mrs. Renshaw agree on the total number of plants to buy, 10, 4, and 5 must all go evenly into that number, whatever it is. Thus the first thing to do is to find the smallest number that is evenly divisible by those 3 numbers, which is 20. So the original number of vegetable plants the Renshaws were thinking of buying could be any multiple of 20 (that is, 20 or 40 or 60 or 80 etc.). But then they decide to buy 2 additional plants, that they hadn't been planning to buy originally, so the total number of plants they actually end up buying must be 2 more than the multiples of 20 listed above (that is, 22 or 42 or 62 or 82 etc.). This means that 10, 4, and 5 will now no longer go evenly into the total number of plants. Finally, the problem states that they agree to buy 6 of each vegetable, so the total number of plants must be evenly divisible by 6. The smallest total number of plants that is evenly divisible by 6 is 42, so that's the answer.

Marching Band Target Problem

Members of the West High School Band were hard at work practicing for the annual Homecoming Parade. First they tried marching in rows of twelve, but Andrew was left by himself to bring up the rear. The band director was annoyed because it didn't look good to have one row with only a single person in it, and of course Andrew wasn't very pleased either. To get rid of this problem, the director told the band members to march in columns of eight. But Andrew was still left to march alone. Even when the band marched in rows of three, Andrew was left out. Finally, in exasperation, Andrew told the band director that they should march in rows of five in order to have all the rows filled. He was right. This time all the rows were filled and Andrew wasn't alone any more. Given that there were at least 45 musicians on the field but fewer than 200 musicians, how many students were there in the West High School Band?

Bake-Sale Target Problem (Experiment 2 Only)

Elena is packaging cookies for a bake sale. She made the cookies a few days ago and brought some to her office yesterday. Since she isn't sure how many cookies are left for the bake sale, Elena just starts putting them in bags, being careful to put the same number of cookies in each bag. First she tries putting sixteen cookies in each bag, but she ends up with only six cookies left for the last bag. So she takes them all out and starts over, putting fourteen cookies in each bag this time. But she ends up with only six cookies for the last bag again. On her third attempt, Elena tries putting only eight cookies in each bag, but again she's left with six cookies for the last bag. Just then the phone rings—it's her friend Cindy. After hearing about Elena's dilemma, Cindy suggests putting nine cookies in each bag. Lo and behold, there are nine cookies left at the end for the last bag. So Elena thanks Cindy for her help, puts a twist-tie on each bag, and leaves for the bake sale. What is the fewest number of cookies Elena could have brought to the bake sale?

Seashell Generalization Problem

Samantha's mother asked her how many sea shells she has in her collection. Samantha said she wasn't sure, but it was a lot—somewhere between 80 and 550. And she could count them by sevens without having any left over. However, if she counted them by threes, there was one shell left over. Things were even worse if she counted the shells by fives, by sixes, by nines, or by tens—there were always four shells left over. Samantha's mother promptly told her how many sea shells she had in her collection. What number did Samantha's mother come up with?

Appendix B

Reliability Analyses

Reliabilities were computed for the following coding schemes: (a) transfer on the band problem, (b) transfer on the bake-sale problem, (c) transfer on the seashell problem, (d) adaptation errors for all three problems, and (e) solution schema quality. Because the coding schemes were identical for the two experiments, reliabilities were computed for the combined set of data. The first author coded all of the data from all of the subjects from both experiments, and a second person independently coded the data from 20% of the subjects from each experiment. Coding was done blind to condition. The 41 "reliability" subjects were chosen randomly, and the second coder coded all data from those subjects (i.e., transfer and adaptation errors for all problems and schema quality). Because the bake-sale problem was used in Experiment 2 only, this procedure would have yielded reliability data for that problem from only 26 subjects. Therefore, the second coder also analyzed an additional seven subjects' bake-sale data, thereby bringing the total bake-sale data coded for reliability to 20% of that collected.

For the three LCM problems, the second coder coded each problem into one or more of the following categories: successful transfer, any of eight adaptation errors, and no attempt at transfer. This information was sufficient to derive the 0/1/2 coding of transfer, as described in the results section of Experiment 1. Reliabilities for the transfer scores were computed by correlating the scores obtained by the two coders, both separately for each problem and for all of the problems combined: $r = .94$ overall ($N = 115$); $r = .93$ for the band problem ($N = 41$); $r = .92$ for the bake-sale problem ($N = 33$), $r = .97$ for the seashell problem ($N = 41$).

Reliabilities also were computed for each of the individual adaptation error categories. Because most of the codes occurred very infrequently in the small sample of reliability data (2–19 occurrences in 115 problems, across categories, as coded by the first author), it was not possible to compute reliabilities separately for each of the three problems. Instead, the data from all problems ($N = 115$) were considered together in computing the correlations. The reliabilities (and percent agreement in parentheses) are: $r = .73$ (97% agreement) for LCM or LCM+r, $r = .79$ (98%) for fail to add remainder, $r = .70$ (99%) for all relevant info but . . . , $r = .82$ (95%) for wrong number as LCM, $r = .81$ (97%) for multiples of LCM+r, $r = .74$ (98%) for expanding series, $r = .93$ (99%) for subtract remainder or add incorrect number as the remainder, and $r = .72$ (97%) for not enough multiples.

For the solution schema task, each subject's written description was coded for the presence/absence of each of the four solution steps described in the results section of Experiment 1. These codings were then converted to numerical scores as described in the text. The reliability of the solution schema coding was determined by correlating the schema quality scores for the two coders: $r = .86$, $N = 41$.

Appendix C

Mapping Hints Used in Experiment 2

Concept-Mapping Hints (Explanations in Brackets)

Remember, the garden problem is similar to this problem. In particular, your goal in this problem is to arrange band members into rows or columns so that each row (or each column) has the same number of people in it, with no one left over. That's like the goal you had in the garden problem of grouping plants into different types so that there were the same number of plants of each type, with no extra spaces left in the garden. In the garden problem the major difficulty encountered was that once the Renshaws finally figured out how many plants they had room for in their garden, all of the arrangements they had thought of left the same number of extra spaces in the garden. There is a similar difficulty in the marching band problem. There, each formation the band director thought of resulted in the same number of people left out. So to summarize, [these two problems are both about putting objects into groups.] The band members are like plants, [because those are the objects being grouped in the two problems.] The rows and columns of band members are like kinds of plants, [because those are the groups in the two problems. Finally,] the number of band members per row or column is like the number of plants of each kind[, because in both problems those are the number of objects in each group].

Number-Mapping Hints (Explanations in Brackets)

Remember, the garden problem is similar to this problem. In particular, the 12, 8, and 3 in the band problem are like the 10, 4, and 5 in the garden problem[, because those are the divisors in each problem that have the same number of things left over]. Also, the 1 in this problem is like the 2 in the garden problem[, because those are the numbers of extra things that have to be accommodated in the two problems]. Finally, the 5 in this problem is like the 6 in the garden problem[, because those are the divisors in the two problems that have no things left over].

Received July 23, 1990
Revision received October 26, 1990
Accepted October 29, 1990 ■