Inducing Mathematical Concepts from Specific Examples: The Role of Schema-Level Variation

David W. Braithwaite (dwbraith@indiana.edu)
Robert L. Goldstone (rgoldsto@indiana.edu)
Indiana University, 1101 E. 10th Street
Bloomington, IN 47405 USA

Abstract

Previous research suggests that comparing multiple specific examples of a general concept can promote knowledge transfer. The present study investigated whether this approach could be made more effective by systematic variation in the semantic content of the specific examples. Participants received instruction in a mathematical concept in the context of several examples, which instantiated either a single semantic schema (non-varied condition) or two different schemas (varied condition). Schema-level variation during instruction led to better knowledge transfer, as predicted. However, this advantage was limited to participants with relatively high performance before instruction. Variation also improved participants' ability to describe the target concept in abstract terms. Surprisingly, however, this ability was not associated with successful knowledge transfer.

Keywords: mathematics; analogy; comparison; schemas; instruction; transfer

Introduction

Part of the power of mathematics lies in its generality. The same mathematical formulae may be used to understand the growth of slime molds or the accumulation of interest from investments, the probabilities of hands in poker or outcomes of scientific experiments, and the oscillations of mechanical or electromagnetic systems. In order to fully realize this power, however, learners must be able to recognize and apply mathematical concepts in contexts different from those in which they were learned – that is, to transfer their mathematical knowledge from learned to novel contexts.

Learners' difficulties in achieving such transfer are well-documented (Novick & Holyoak, 1991; Ross, 1987). One reason may be that, when a general idea is learned in the context of specific examples, learners' concepts become tied to the details of the examples, inhibiting their ability to recall the concept or apply it correctly when faced with cases that do not share similar details (Ross, 1987). This difficulty may be especially strong when the examples are presented in a perceptually detailed format (Kaminski, Sloutsky, & Heckler, 2008), and is likely to be more serious for domain novices than experts (Novick & Holyoak, 1991).

One way to address this difficulty is to present mathematical ideas in abstract form, without specific examples. Such an approach has indeed been shown to promote transfer in some cases (Kaminski et al., 2008). However, in other cases, learners have experienced serious difficulties with abstractly-presented mathematics, despite being competent with the same mathematics encountered in familiar contexts (Nuñes, Schliemann, & Carraher, 1993). In such contexts, learners can apply intuitions from everyday life to help in understanding the mathematical ideas involved. Abstract presentation of mathematical ideas therefore risks sacrificing learning for the sake of transfer.

It may, then, be desirable for learners to encounter mathematical ideas in a way that leverages their intuitive understanding of specific examples, while also drawing attention to the abstract structure present in those examples. Research on analogy suggests that this goal might be achieved through presentation of multiple specific examples followed by comparison (Gentner, Loewenstein, & Thompson, 2003; Gick & Holyoak, 1983). Comparing examples encourages learners to align their corresponding elements, and thereby to notice their common relational structure. Awareness of this structure, in turn, can facilitate understanding of new cases with the same structure. Thus, learning mathematical ideas by studying and then comparing multiple examples may enable learners to gain intuitive accessibility without losing generality.

The question then arises as to how the examples which will instantiate a mathematical concept during learning are to be chosen. Central to this question is the issue of how much, and in what ways, the examples should differ from each other. If, as the above research suggests, learners induce concepts that incorporate commonalities among the examples, it seems desirable that the examples should share the mathematical structure in question, but should not share other extraneous details. Extraneous commonalities might be misunderstood as part of the concept to be learned, limiting learners' ability to generalize (Medin & Ross, 1989), and so defeating the purpose of using multiple examples in the first place. These observations suggest that extraneous aspects should be systematically varied across examples, while holding mathematical structure constant.

The present study investigates the effects on mathematical concept learning of a particular type of variation among examples: variation at the level of "semantic schemas." This term here refers to structures more general than specific examples but less general than mathematical structure. Consider the three combinatorics problems shown in Figure 1. Problems (a) and (b) share a schema, termed "Objects Selected in Sequence" (OSS), in which a sequence of selections is made from a fixed set of options. Problem (c), by contrast, belongs to a different schema, termed "People Choosing Options" (PCO), in which several people each choose once from a fixed set of options.

(a)	(b)	(c)
A piano student, when bored, plays random sequences of notes on the piano, using sequences of a fixed length, and choosing from a fixed set of notes. How many different sequences are possible, if there are 5 possible notes and the sequences are 6 notes long?	A website generates user passwords by selecting a certain number of characters randomly from a fixed set of characters. How many different passwords are possible, if the passwords are 6 characters long and there are 5 permissible characters?	A marketing research company conducts a taste test survey. Several consumers are each asked to choose their favorite from among several pizza flavors. How many different results of the survey are possible, if there are 6 consumers and 5 pizza flavors?

Figure 1. Three combinatorics problems.

Of course, all three problems share the same mathematical structure (discussed further in the Methods section), and the differences between them would likely not seem important to a mathematics expert. For mathematics novices, however, semantic schemas are known to exert a strong influence on the mathematical interpretation of contextualized problems. For example, Bassok, Wu, and Olseth (1995) found that learners were more likely to solve correctly problems in which schematic and mathematical roles were matched consistently with their default expectations than problems in which such matches were inconsistent. In light of the preceding discussion, learning about a mathematical structure via several examples based on the same schema might lead learners to induce concepts tied to that particular schema, and thus to perform poorly on problems involving other schemas. Conversely, systematic variation of the schemas encountered during learning should lead to induction of more general concepts and thus to more successful transfer to novel problems.

This hypothesis was investigated in the present study. Combinatorics problems were used as the domain for study and transfer for several reasons. First, the discovery of better methods for learning and teaching combinatorics would have considerable practical value due to the foundational role of combinatorics in applied mathematics – in particular, probability and statistics. Second, mathematics learners are known to have considerable difficulty correctly applying combinatorics methods to novel problems (Bassok et al., 1995; Ross, 1987). Finally, semantic schemas are known to play a role in the mathematical interpretation of combinatorics problems (Bassok et al., 1995).

Methods

Participants

Participants were 109 Indiana University undergraduate students, who participated in partial fulfillment of a course requirement.

Materials

Sixteen story problems were constructed as stimuli. All of the problems had the same mathematical structure: Sampling with Replacement (SWR), in which multiple selections are made from a fixed set. The number of possible joint outcomes in such a case is given by the expression m^n, where m is the number of elements of the set and n is the number of selections, or sampling events.

The sixteen problems belonged to four different schema categories. The first two categories were those already illustrated above: PCO and OSS (OSS: Figure 1a-b, PCO: Figure 1c). Problems in these categories were used as learning examples. The other two categories were Options Assigned to Places (OAPlc) and Objects Assigned to People (OAPpl), illustrated below (Figures 2a and 2b respectively). OAPlc and OAPpl problems served as pretest and transfer problems. Note that in the learning examples (OSS and PCO) and OAPlc problems, people are either doing the choosing or are not mentioned at all. In OAPpl, by contrast, people are being chosen instead of choosing. Due to this role reversal relative to the learning examples, transfer to OAPpl problems was expected to be particularly difficult, as found in previous research (Ross, 1987).

(a)	(b)
A homeowner is going to repaint several rooms in her house. She chooses one color of paint for each of the rooms. In how many different ways can she paint the rooms, if there are 3 colors and 5 rooms?	A prize drawing is held at a small office party, and each of several prizes is awarded to one of the employees. In how many different ways can the prizes be awarded, if there are 6 prizes and 4 employees?

Figure 2. Combinatorics problems from the (a) OAPlc and (b) OAPpl categories.

Each problem category contained two pairs of problems, for a total of four problems. The problems within a pair involved the same back story but different numbers, while the two pairs within each category involved different back stories (and different numbers from each other). The order in which the two critical numbers, i.e. the size of the sampled set and the number of sampling events, were presented was varied among questions so that it could not serve as a cue to match the numbers to their respective roles.

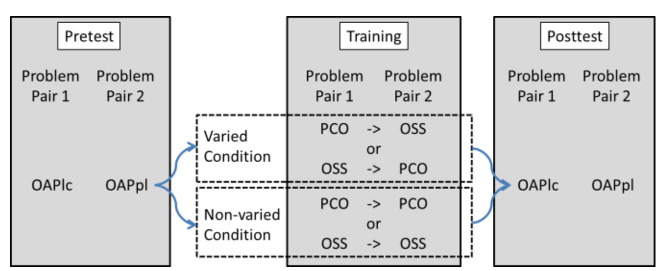

Figure 3. Summary of experimental design.

The experiment employed a pretest-training-posttest design, summarized in Figure 3. The pretest consisted of one OAPlc problem pair and one OAPpl problem pair, for four problems altogether. The posttest consisted of the other OAPlc problem pair followed by the other OAPpl problem pair. Thus, all eight OAPlc and OAPpl problems appeared in either the pretest or the posttest.

The training consisted of worked solutions to four problems drawn from the PCO and OSS categories. Participants were assigned randomly to one of two training condition. In the varied condition, participants were shown one pair of problems from each category, either PCO followed by OSS or vice versa (these two possible orders were balanced across participants). In the non-varied condition, participants were shown two pairs of problems from the same category, either both PCO or both OSS (again, the two possibilities were balanced across participants). If a certain problem category was shown in a given position (either first pair or second pair), it was always the same problem pair regardless of condition. For example, if PCO problems were shown first in the varied condition, they were the same problems that were shown first in the non-varied condition. An important consequence of this design is that each training problem was shown equally often across the two conditions.

Procedure

Participants were randomly assigned to receive one set of OAPlc / OAPpl problems as pretest. The pretest problems were displayed to participants on a computer monitor together with a virtual calculator, which participants were encouraged to use as needed. Only one problem appeared on the screen at a time. Two spaces were provided below each problem: one in which to show work, and another in which to write the final answer. Participants were required to show their work and enter some number as their final answer before they could proceed to the next question.

After the pretest, answers were scored for correctness, and participants were classified as high pretest performers if they answered at least 50% of the pretest problems correctly and low pretest performers otherwise. They were then assigned randomly to one of the two training conditions with the constraint that, at each level of pretest performance, the number of participants in each condition was balanced. This manipulation was intended to reduce differences in pretest scores between training conditions.

The training problems corresponding to participant's training conditions were then presented in the same way as the pretest problems. However, after completing each problem, participants were shown the correct answer together with a brief explanation of how the answer was calculated and why this calculation was appropriate. These explanations utilized exponential notation but did not show the general expression m^n. Instead, they only showed specific versions of this expression instantiated with the numbers used in the problem. The explanation for a given problem did not differ between training conditions.

After completing each pair of training problems, participants were asked to choose from a list of options the correct method of solving problems like those just seen, independent of the specific numbers involved. For example, the correct answer to this question after the problems involving pizza flavors (Figure 1c above) was "Multiply the number of pizza flavors by itself as many times as there are consumers." Participants who chose incorrectly were not allowed to proceed until they chose the correct answer.

After answering the above question for the *second* pair of training problems (only), participants were asked to choose from a list of options the correct mapping between elements of the preceding two problem pairs. For example, the correct answer to this question if the preceding problem pairs involved a website generating passwords and consumers tasting pizza flavors (Figure 1b and 1c) was "The length of the note sequences corresponds to the number of consumers, and the number of possible notes corresponds to the number of pizza flavors." The purpose of this question was to encourage participants to think about the shared structure of the training problem pairs. After answering this question, participants were asked to describe, in free-response format, a general method for solving problems like those just seen. No feedback was given for either of these questions.

Finally, participants were administered the posttest. The posttest utilized whichever set of OAPlc / OAPpl problems had not been presented during the pretest, and the procedure was in all ways the same as for the pretest.

Coding

For each problem, participants were assigned a score of 1 if their answer was correct and 0 otherwise.

Responses to the free-response question regarding a general solution method posed at the end of the training were coded on a 0-2 scale in each of two respects. For the first respect, Correctness, responses were assigned a score of 2 if they indicated that the number of elements in the sampled set should be raised to the power of the number of sampling events (or multiplied by itself as many times as the latter). Responses which implicated exponentiation but did not correctly identify the base and exponent were assigned a score of 1, and all other responses received a score of 0. The second respect, Abstractness, was intended to measure how well participants had generalized beyond the specific details of the learning examples. Responses were assigned a score of 2 if they referred to the two numbers using general words, such as "the options" (for the size of the sampled set) or "the number of times they are able to be chosen" (for the number of sampling events). Responses which used general words for one but not the other number were assigned a score of 1, and all other responses received a score of 0. All responses were coded by two independent coders, and all disagreements were resolved through discussion. In the analyses detailed below, scores of 0 and 1 were combined for both correctness and abstractness, so that responses were classified as either correct (2) or not correct (0 or 1) and abstract (2) or not abstract (0 or 1).

Results

Average pretest and posttest scores are shown in Figure 4. Participants demonstrated considerable improvement on posttest, but the amount of improvement varied by problem category. The data were entered into a 2 (test section: pretest or posttest) x 2 (problem category: OAPlc or OAPpl) within-subjects ANOVA. The main effects of both factors and the interaction between them were all significant (test section: $F(1,108)=69.8$, $p<.001$; problem category: $F(1,108)=14.6$, $p<.001$; interaction: $F(1,108)=16.4$, $p<.001$). Participants improved from pretest (0.216) to posttest (0.489), but this improvement was greater for OAPlc (0.225 to 0.638) than for OAPpl (0.206 to 0.339).

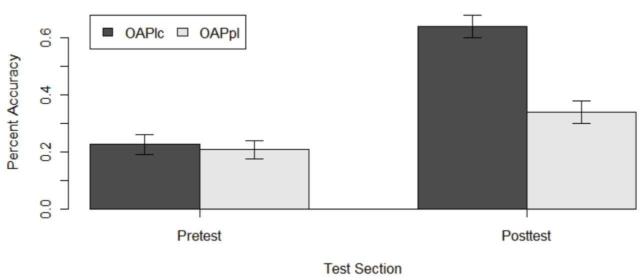

Figure 4. Pre and posttest accuracy by problem category[1].

Figure 5 shows average transfer scores, defined as the difference between posttest and pretest scores, for each training condition, among low and high pretest performers. Transfer scores were submitted to a 2x2x2 mixed ANOVA with training condition (varied vs. non-varied) and pretest performance (low or high) as between-subjects factors and problem category (OAPlc or OAPpl) as a within-subjects factor. The main effect of pretest performance was significant, $F(1,105)=66.6$, $p<.001$, indicating more improvement from pretest to posttest among low pretest performers (0.404) than high pretest performers (-0.056). Also, the effect of problem category was significant, $F(1,105)=12.3$, $p=.001$, indicating greater improvement on OAPlc (0.413) than on OAPpl (0.133). Problem category did not interact significantly with any of the other factors.

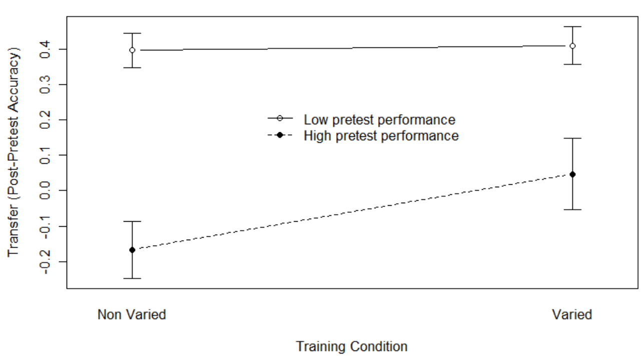

Figure 5. Transfer by condition and pretest performance.

[1] Here and elsewhere, error bars indicate standard errors.

More importantly, the main effect of training condition was significant, $F(1,105)=4.0$, $p=.049$, indicating greater improvement in the varied (0.305) than in the non-varied (0.201) condition. However, this effect was qualified by a marginally significant condition by pretest performance interaction, $F(1,105)=3.1$, $p=.08$. Consequently, the same model (excluding the pretest performance factor) was applied separately to the data from low and high pretest performers. This analysis found a significant effect of training condition among high performers, $F(1,29)=.706$, $p=.022$, indicating higher transfer in the varied condition (0.047) than in the non-varied condition (-0.167), but no effect of training condition among low performers, $F(1,76)=.042$, $p=.838$ (varied: 0.410, non-varied: 0.397).

In addition to the effect of training condition on transfer, we were also interested in whether training condition affected participants' ability to induce a general method for solving SWR problems. The proportion of participants providing correct and abstract solution descriptions (i.e. receiving scores of 2 on the correctness and abstractness scales) within each training condition are shown in Figure 6. In the varied condition, 40% of participants' solutions were scored as correct, 62% as abstract, and 29% as both correct and abstract. In the non-varied condition, 56% of participants' solutions were scored as correct, 39% as abstract, and 20% as both correct and abstract.

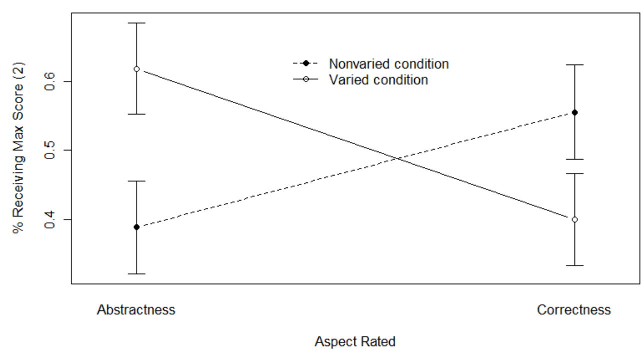

Figure 6. Percent generating correct or abstract general solutions by training condition.

The Breslow-Day test, a non-parametric test for stratified analysis of 2x2 tables, was applied to the frequencies of best (2) and other (0-1) scores within each training condition (varied or non-varied) for each aspect rated (correctness or abstractness). The relative frequencies of best vs. other scores between training conditions differed significantly according to aspect rated, $p=.004$. In other words, the effectiveness of varied relative to non-varied training was greater with respect to abstractness than with respect to correctness. To further clarify this effect, Pearson's Chi-square tests were applied to the contingency tables of best vs. other scores by training condition separately for each measurement respect. These analyses found that abstract solutions were more common in the varied than in the non-varied condition, p=.028, but the proportion of correct solutions did not differ by training condition, p=.152.

Were participants who provided solutions that were abstract, correct, or both more likely to perform well on posttest? Average posttest scores among participants displaying each combination of solution abstractness and correctness are shown in Figure 7. (Participants were approximately equally distributed over these combinations.) Scores were virtually identical for each of these combinations: 0.50 for both correct and abstract, 0.49 for neither abstract nor correct, 0.48 for abstract but not correct, and 0.48 for correct but not abstract. A mixed ANOVA applied to posttest scores with solution correctness (correct or not), solution abstractness (abstract or not), pretest performance, and training condition as between-subjects factors and problem category as a within-subjects factor found no significant main effects of solution correctness or abstractness, no significant interaction between them, and no significant interaction of either or both with any other factor. (None of these effects were significant when transfer rather than posttest scores were entered into the model.)

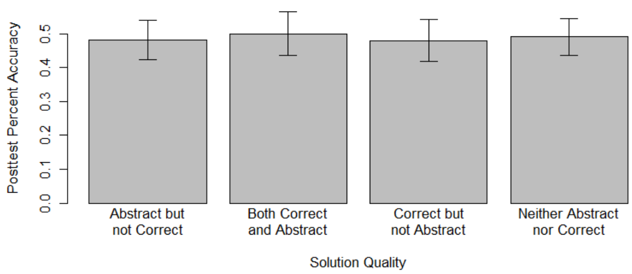

Figure 7. Average transfer scores by correctness and abstractness of generated solution and test problem pair.

Discussion

This experiment investigated whether exposure to multiple examples of an abstract mathematical concept followed by comparison among them would lead to better induction of the general concept when the semantic schemas of the examples were systematically varied during learning than when all examples were based on the same schema. As predicted, participants in the varied condition both induced more abstract solution methods for SWR problems, and showed greater improvement on a transfer test requiring them to apply such methods. These results suggest that schema-level variation of examples can be an effective way to promote transfer.

Caution is necessary in interpreting these results because the advantage of the varied over the non-varied condition in promoting transfer was almost entirely driven by high pretest performers. Low pretest performers did not benefit from the varied condition, although they were not hurt by it either. A possible reason is that the dissimilarity between examples in the varied condition made it difficult to notice their shared structure. This difficulty might be overcome by presenting several examples from the same schema, thus facilitating comparison and alignment of the examples, before introducing schema-level variation. Consistent with this view, Kotovsky and Gentner (1996) found that children initially presented with several examples sharing both abstract structure and superficial details were later able to notice shared structure even in the absence of superficial similarity. Similarly, Elio and Anderson (1984) found that category learning was better after a learning schedule beginning with low variation among exemplars and later progressing to more variation, as opposed to one beginning with and maintaining a high level of variability.

Interestingly, Elio and Anderson (1984) also found that when learners were specifically instructed to take an analytical approach to category learning, the effectiveness of training with initially high variability improved. Similarly, high pretest performers in the present study, who may have been better equipped to take an analytical approach to learning the SWR concept, derived greater benefits from varied relative to non-varied training. One account for this result is that good learners are more attentive to the features and relations that are relevant to domain principles. Consequently, good learners would be less likely to be distracted by – and more likely to benefit from – variation in extraneous features and relations. Considering this conclusion together with the previous one regarding weaker learners, the best instructional approach might be an adaptive one, beginning with examples drawn from a single schema and transitioning to schema-level variation once learners demonstrate understanding of the target concept in the context of the initial schema. This interesting possibility deserves further investigation.

However, the observed advantage of the varied training for high pretest performers must also be interpreted with caution. Transfer scores among high pretest performers were rather low, averaging around zero in the varied condition and below zero in the non-varied condition. One interpretation of these data is that varied training merely helped to avoid negative transfer, and did not actually benefit learners. On the other hand, high pretest performers might be expected to show regression to the mean on posttest, resulting in negative scores on our measure of transfer. In this case, the actual (slightly above zero) transfer scores in the varied condition would represent a positive effect of training. It is difficult to disambiguate between these possibilities due to the lack of a control condition in the present study. Also, the inclusion of particularly difficult transfer problems, i.e. those in the OAPpl category, may have obscured the presence of positive transfer by bringing down the overall average. The beneficial effects of schema-level variation might be better explored in future studies by using a wider range of relatively easy transfer problems.

In addition to their differing effects on transfer, the varied and non-varied training conditions also led to differing levels of success in describing general solutions for SWR problems. In particular, while participants in both conditions were equally able to describe correct solutions, those in the varied condition were better able to characterize the elements of those solutions in abstract, general terms.

Previous research has demonstrated that comparison between multiple analogous examples can lead participants to induce their shared abstract structure (Gentner et al., 2003; Gick & Holyoak, 1983). The present findings build on that principle by suggesting that if the examples in question share semantic content not intrinsic to the desired structure, learners may induce a more limited, less general concept than if such extraneous semantic content is systematically varied across learning examples. Moreover, not only superficial elements but also more abstract semantic structures, such as the schemas of the present study, can count as extraneous content in this context. This conclusion implies that instructional design in mathematics could benefit from attention to variation of semantic schemas across examples of a given concept.

Although the varied condition led both to more abstract described solutions and to better transfer performance, the former effect did not mediate the latter as expected. In fact, participants who succeeded in describing general solutions were not more likely than other participants actually to demonstrate successful transfer. This result is surprising in light of previous research, in which the quality of participants' generalizations following exposure to multiple examples of a concept *did* predict their ability to apply the concept to novel cases (Gick & Holyoak, 1983; Novick & Holyoak, 1991). Several explanations are possible for this dissociation of described solution methods and problem-solving performance.

First, participants may not have attempted to apply their described solutions during the transfer test, possibly due to failure to recall the solutions or failure to recognize their relevance. However, these possibilities seem unlikely given that the transfer test was administered immediately after participants described their general solutions, and that the problems in the transfer test were presented in the same format and with very similar wording to those in the training. Second, participants may have attempted to apply their solutions, but failed to do so successfully on either or both pairs of transfer problems. Such failure might have been due either to inability to map the elements of the transfer problems to the roles mentioned in their solutions, or to inability to apply the solution procedure despite having correctly mapped the corresponding elements. Both of these issues have been implicated in failures of analogical transfer in mathematics learning (Novick & Holyoak, 1991). Future research might disambiguate between these possibilities by, on the one hand, directly testing whether participants could map elements in the transfer problems to those in training problems, and on the other hand, testing the effects of providing such a mapping to participants.

Regardless of why posttest performance was not predicted by participants' ability to describe correct and general solution methods, it is clear that such ability was not the cause of the superior transfer observed in the varied over the non-varied condition. The question then arises: what *was* the cause for that advantage in transfer? Because this advantage was dissociated from explicit, articulable knowledge of how to solve the problems, it seems likely to relate to some form of implicit knowledge, e.g. improved perception / encoding of problems or improved procedural skill. Because the procedures required were essentially the same across problems and conditions, the perceptual explanation seems more likely. The varied condition may have encouraged learners to encode the elements of the problems in terms of their general roles in the mathematical structure of SWR, rather than in terms of their more specific roles in one or another semantic schema. Such improved encoding could, in turn, have facilitated application of the solution procedures learned during training to the transfer problems. This explanation is admittedly speculative, but offers a promising direction for future research.

Acknowledgments

This research was supported by National Science Foundation REESE grant 0910218.

References

Bassok, M., Wu, L.-ling, & Olseth, K. L. (1995). Judging a book by its cover: Interpretative effects of content on problem-solving transfer. *Memory & Cognition*, *23*(3), 354-367.

Elio, R., & Anderson, J. R. (1984). The effects of information order and learning mode on schema abstraction. *Memory & cognition*, *12*(1), 20-30.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*, *95*(2), 393-405.

Gick, M. L., & Holyoak, K. J. (1983). Schema Induction and Analogical Transfer. *Cognitive Psychology*, *15*, 1-38.

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2008). The Advantage of Abstract Examples in Learning Math. *Science*, *320*(April), 454-455.

Kotovsky, L., & Gentner, D. (1996). Comparison and Categorization in the Development of Relational Similarity. *Child Development*, *67*(6), 2797-2822.

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, N.J.: Lawrence Erlbaum.

Novick, L. R., & Holyoak, K. J. (1991). Mathematical problem solving by analogy. *Journal of experimental psychology: Learning, memory, and cognition*, *17*(3), 398-415.

Nuñes, T., Schliemann, A. D., & Carraher, D. W. (1993). *Street Mathematics and School Mathematics*. Cambridge University Press.

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*(4), 629-639.