Benefits of Variation Increase with Preparation

David W. Braithwaite (baixiwei@gmail.com)
Department of Psychology, Indiana University, 1101 E. 10th Street
Bloomington, IN 47405 USA

Robert L. Goldstone (rgoldsto@indiana.edu)
Department of Psychology, Indiana University, 1101 E. 10th Street
Bloomington, IN 47405 USA

Abstract

Abstract concepts are characterized by their underlying structure rather than superficial features. Variation in the examples used to teach abstract concepts can draw attention towards shared structure and away from superficial detail, but too much variation can inhibit learning. The present study tested the possibility that increasing attention to underlying structural relations could alleviate the latter difficulty and thereby increase the benefits of varied examples. Participants were trained with either varied or similar examples of a mathematical concept, and were then tested on their ability to apply the concept to new cases. Before training, some participants received pretraining aimed at increasing attention to the structural relations underlying the concept. The relative advantage of varied over similar examples was increased among participants who received the pretraining. Thus, preparation that promotes attention to the relations underlying abstract concepts can increase the benefits of learning from varied examples.

Keywords: variation; concept learning; conservative generalization; progressive alignment; prior knowledge; mathematics.

Background

Abstract concepts – that is, concepts characterized by underlying structure rather than by surface similarities among their members – are pervasive in human thought and language (Gentner & Kurtz, 2005). Such concepts play a particularly prominent role in formal education. Examples include functions in mathematics, forces in physics, supply and demand in economics, and natural selection in biology. The power of such concepts comes from their ability to capture deep commonalities across superficially dissimilar situations. Thus, a key goal for instruction is to enable learners to extend such concepts to a wide range of instances, which may differ substantially from studied examples.

Evidence suggests that learning concepts from highly varied examples can help to achieve this goal. Most directly, variation in studied examples increases the chance that a novel instance will resemble a studied example. Also, exposure to varied examples can increase the range of variation associated with the concept. Both of these considerations suggest that learning from varied, relative to similar, examples would allow concepts to be extended to a wider range of novel instances, a prediction confirmed by numerous studies of category learning (Hahn, Bailey, & Elvin, 2005; Posner & Keele, 1968).

An additional consideration more specific to abstract concepts is that similarity between examples can lead to overspecified concept representations, while variation can avoid this pitfall. Similar examples are likely to share superficial features in addition to their common, constitutive structure. Learners may then incorporate these superficial features as part of the learned concept, limiting their ability to extend the concept to novel instances that do not share the same features, a phenomenon known as *conservative generalization* (Medin & Ross, 1989). Varied examples, by contrast, are likely to share few or no superficial features. Concepts learned from such examples would include only the relevant underlying structure and, therefore, would generalize more easily to novel cases. Indeed, several studies have found that abstract concepts learned from varied rather than similar examples are more easily extended to novel instances (Chen & Mo, 2004; Day, Goldstone, & Hills, 2010).

However, variation among examples also has potential drawbacks. Too much variation might prevent learners from discerning what the examples have in common with each other, preventing them from learning the underlying concept at all. By contrast, if the examples are similar – that is, share some surface features in addition to underlying structure – then correspondences between their surface features may be easy for learners to notice. These correspondences can serve as scaffolding, allowing learners to notice their shared underlying structure, a phenomenon known as *progressive alignment* (Kotovsky & Gentner, 1996). Consistent with this view, some studies have found that abstract concepts are acquired more easily if similar, rather than varied, examples are presented initially (Elio & Anderson, 1984; Kotovsky & Gentner, 1996). These results dovetail with studies on perceptual category learning that have shown slower category learning from varied, relative to similar, examples (Hahn et al., 2005).

Thus, varied and similar examples may each offer potential benefits for abstract concept learning. The relative strengths of their benefits may depend on an additional factor: learners' predisposition, and ability, to attend to the structural relations present in the examples. On the one hand, learners who attend to structural relations might easily ignore superficial commonalities between examples. On this account, attention to relations would reduce the risk of conservative generalization, a potential drawback of similar examples, and would thereby increase the relative benefits of similar examples. On the other hand, learners who can easily attend to structural relations might be able to notice when different examples share a common structure, even if those examples differ with respect to superficial features. On this account, attention to structure would reduce a potential obstacle to learning from varied examples, and thereby increase the relative benefits of such examples.

Research investigating interactions of prior knowledge and variation among examples on concept learning offers some evidence favoring the second possibility. In particular, several studies have found relatively greater benefits of similar examples for less knowledgeable learners, and of varied examples for more knowledgeable learners (Braithwaite & Goldstone, 2012, under review; Guo, Yang, & Ding, 2013). For example, Braithwaite and Goldstone (under review) found that university students who had previously studied a mathematical concept learned the concept better from varied than from similar examples, while those without such experience showed the opposite trend. Learners with more prior knowledge may resemble domain experts by paying more attention to relevant structural relations (Chi, Feltovich, & Glaser, 1981). Similarly, learners with less prior knowledge may resemble domain novices by paying relatively little attention to relevant structural relations. Thus, the results of Braithwaite and Goldstone (under review) can be interpreted as showing a positive relationship between attention to structure and the relative benefits of varied examples. Conversely, these results can also be interpreted as showing greater need for the scaffolding offered by superficial similarities between examples among learners less disposed, or able, to attend to structural relations. A similar account has been offered for the "expertise reversal effect," in which less knowledgeable learners show greater benefits of scaffolding during instruction (Kalyuga, Ayres, Chandler, & Sweller, 2003).

While the findings mentioned above (Braithwaite & Goldstone, 2012, under review; Guo et al., 2013) are consistent with the possibility that attention to structural relations increases the relative benefits of varied examples, this interpretation is open to debate because it relies on prior knowledge as a proxy for attention to relations. Additionally, prior knowledge was observed rather than manipulated in these studies, preventing any strong causal inference involving effects of that variable. The goal of the present study is to investigate the interactions of attention to structural relations and variation among examples more directly, by manipulating both within the same experiment. For the reasons described above, we predicted that manipulations intended to increase attention to structure would also increase the relative benefits of varied over similar examples.

Method

The concept employed for the study was the mathematical concept of Sampling with Replacement (SWR). SWR describes any situation in which, for **each** element of one set termed "selections," **one** element of another set termed "alternatives" is chosen. Examples of such situations are shown in Table 1.

Table 1. Examples of sampling with replacement (SWR) problems, with solutions. Example text is abbreviated.

Story Type	Example	Selections	Alternatives	Solution
People Choosing Objects (PCO)	*A group of friends is eating at a restaurant. Each person chooses a meal from the menu. In how many different ways can the friends choose their meals, if there are 5 friends and 6 meals?*	friends	meals	6^5
Objects Selected in Sequence (OSS)	*A piano student, when bored, plays random melodies on the piano. Each melody is the same number of notes long, and uses only keys from a fixed set of keys. How many different melodies are possible, if there are 4 keys in the set and 7 notes in each melody?*	notes in each melody	keys in the set	4^7
Objects Assigned to Places (OAPlc)	*A homeowner is going to repaint several rooms in her house. She chooses one color of paint for the living room, one for the dining room, one for the family room, and so on. In how many different ways can she paint the rooms, if there are 8 rooms and 3 colors?*	rooms	colors	3^8
Categories Assigned to Events (CAE)	*An FBI agent is investigating several paranormal events. She must write a report classifying each event into a category such as Possession, Haunting, Werewolf, etc. In how many different ways can she write her report, if there are 9 categories and 4 paranormal events?*	paranormal events	categories	9^4
Objects Assigned to People (OAPpl)	*An aging king plans to divide his lands among his heirs. Each province of the kingdom will be assigned to one of his many children. In how many different ways can the provinces be assigned, if there are 5 provinces and 7 children?*	provinces	children	7^5

Participants in the study were required to determine how many different outcomes were possible in a variety of different SWR situations, where an outcome constitutes a particular choice of one alternative for each of the selections. The number of outcomes in any SWR situation is given by a simple formula: $(\textit{number of alternatives})^{(\textit{number of selections})}$, which was given to participants. To apply this formula to a particular situation requires one to correctly identify which elements of the situation fill the roles of selections and alternatives. Table 1 shows the correct identification of selections and alternatives, and the consequent instantiation of the formula, for each of several examples of SWR. As the Table illustrates, there is little superficial similarity between the elements filling a given role in different examples of SWR. For this reason, SWR meets our definition of an abstract concept.

The study employed a pretest-pretraining-training-posttest design. During the pretest, participants solved a number of SWR problems. During the pretraining, participants learned about the structural relations underlying SWR problems, without learning how to solve them. In the training section, participants learned how actually to solve SWR problems. Participants in the two experimental conditions (graphical and verbal, detailed below) received both pretraining and training, while those in the control condition skipped the pretraining and received only training. In both experimental and control conditions, the level of variation between examples shown during training and, where applicable, pretraining was manipulated between subjects. Finally, after completing the training, participants solved more SWR problems as a posttest. Improvement in performance from pretest to posttest served as a measure of participants' learning from the instruction and practice they received.

Participants

Participants were N=215 Indiana University undergraduate students who participated in partial fulfilment of a course requirement. Participants were assigned randomly to one of three pretraining conditions: verbal pretraining (N=73), graphical pretraining (N=71), or control, i.e. no pretraining (N=71). Within each pretraining condition, participants were assigned randomly to receive either varied or similar examples (verbal pretraining: N=37 similar and N=36 varied, graphical pretraining: N=36 similar and N=35 varied, control: N=36 similar and N=35 varied).

Materials

12 SWR story problems were created for use in the pretest and posttest. The problems were based on five story types, which differed in terms of the types of story elements that filled the roles of alternatives and selections. The types were: People Choosing Objects (PCO), Objects Selected in Sequence (OSS), Objects Assigned to Places (OAPlc), Categories Assigned to Events (CAE), and Objects Assigned to People (OAPpl). One example of each story type is shown in Table 1. Two problems were created for each story type except OAPpl, for which four problems were created, yielding a total of 12 problems. These problems were divided into two sets of 6 problems each, each set containing two OAPpl problems and one problem of each other type. For each participant, one problem set was selected randomly to serve as pretest, with the other serving as posttest.

12 additional problems were created for use as pretraining and training examples. These problems were all based on the PCO and OSS story types, and included 6 problems of each type. Each participant was exposed to only 6 of the 12 examples, selected as follows. First, participants were assigned randomly to receive either similar or varied examples. Participants who received similar examples were shown all 6 examples of one type and no examples of the other type. Participants who received varied examples were shown 3 randomly-selected examples of each type in alternating sequence. The story type of the first example (and, for the similar examples case, for all subsequent examples) was determined randomly for each participant as either PCO or OSS.

Because only PCO and OSS problems were employed during training, while participants were tested on problems of all 5 types, the test afforded a measure of knowledge transfer, i.e. ability to solve SWR problems embedded in situations quite different from those studied during training. For participants who studied PCO problems during training, the OAPpl problems afforded a particularly stringent measure of transfer, because the PCO and OAPpl story types were cross-mapped (Gentner & Toupin, 1986) relative to each other. In particular, in PCO problems, the role of selections is filled by people, but in OAPpl problems, people instead fill the role of alternatives. Thus, we expected OAPpl problems to be particularly challenging for participants who were exposed to PCO problems during training.

Procedure

The experiment was conducted through a web-based computer interface[1], with each participant working in a cubicle at a separate computer. Participants first completed the pretest. For each problem in the pretest, two variations were presented which differed mainly with respect to the specific numbers in the problem; these variations were used to decrease the likelihood of obtaining a perfect score by guessing. Participants had to answer each problem by choosing between two responses, one of which was "x^y" and the other "y^x," where x and y were the two numbers mentioned in the problem statement. Participants answered the problems one at a time in a fixed order, and no feedback was given.

After the pretest, participants in the experimental conditions began the pretraining section. First, they read a brief exposition describing the common structure shared by all

[1] A demonstration version of the experiment may be viewed at https://perceptsconcepts.psych.indiana.edu/experiments/dwb/scmvar06/demo.html.

SWR problems. They were then presented with the 6 pretraining/training examples, selected as described under Materials, one at a time. For each example, participants chose between two descriptions of the problem structure, and were required to explain their answer (Figure 1). In the verbal pretraining condition, the descriptions involved a sentence template of the form "For **EACH** of the ___, **ONE** of the ___ is chosen," and participants had to choose which problem element belonged in each blank (Figure 1A). In the graphical pretraining condition, the descriptions involved a diagram resembling a combination lock, and participants had to choose which problem element corresponded to the lock tumblers and which to the different options on each tumbler (Figure 1B). Feedback was given after each response, and after completing each of the 6 examples separately, participants reviewed them all together.

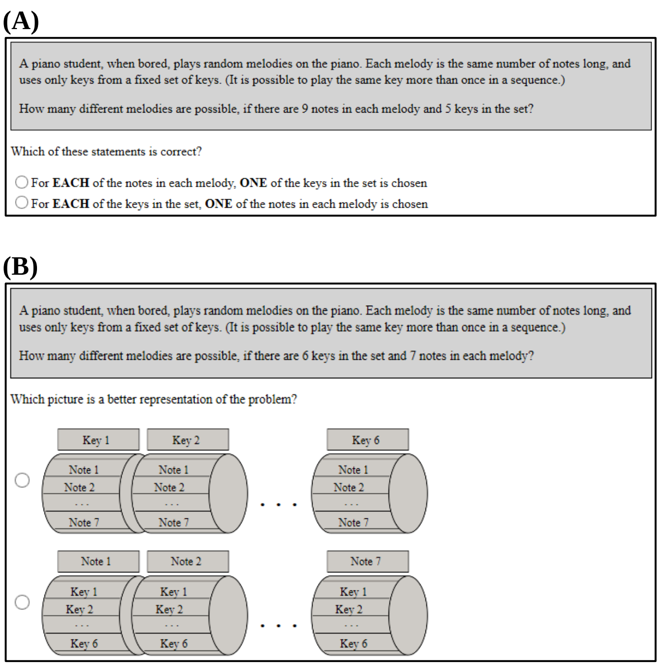

Figure 1: Pretraining interface. (A) Verbal, (B) Graphical.

After the pretraining section – or, for participants in the control condition, after the pretest – participants began the training section. They first read a passage explaining how to solve SWR problems. This passage presented the general formula for SWR problems, and emphasized that to instantiate the formula, one must identify the problem elements playing the roles of selections and alternatives. Participants were then presented with the 6 example problems one at a time. For each problem, participants chose between two responses of the form "x^y," just as in the pretest. Feedback was given after each response, and after completing each of the 6 examples separately, participants reviewed them all together. Finally, participants were asked to describe, in open-response format, the general method of solving SWR problems, and how to determine which problem elements played the roles of selections and alternatives.

After the training, participants completed the posttest, which was administered in the same way as the pretest.

Results

Training Performance

Although we were mainly interested in effects of pretraining and training on test performance, performance during training was also analyzed. Average accuracy for the training examples was 79.1%. A 3 (pretraining condition: control, graphical, or verbal)×2 (level of variation: similar or varied) ANOVA found a significant effect of pretraining condition on training accuracy, $F(2,209)=6.34$, $p=.002$. Average training accuracy by pretraining condition is shown in Figure 2. Accuracy was higher in the graphical (86.9%) than in the verbal (77.8%) and control (72.5%) conditions. Pairwise *t*-tests with a Holm correction for multiple comparisons confirmed that graphical differed significantly from control, $p=.002$, and marginally significantly from verbal, $p=.053$. Accuracy in the verbal and control conditions did not differ significantly, $p=.187$. No effect of variation or interaction of variation with pretraining condition was found, *p*s>.35.

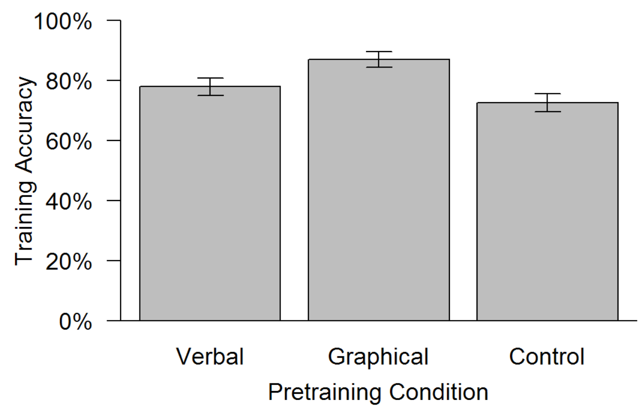

Figure 2: Training accuracy by pretraining condition. Here and elsewhere, error bars represent standard errors.

Test Performance

Average accuracy was 58.5% on the pretest and 67.4% on the posttest. Both pretest and posttest scores were significantly higher than chance, i.e. 50.0%, $t(214)=7.04$, $p<.001$ for the pretest and $t(214)=12.68$, $p<.001$ for the posttest.

Accuracy scores for the pretest and posttest were submitted to a mixed ANOVA with test section (pretest or posttest) and story type (PCO, OSS, OAPlc, CAE, or OAPpl) as within-subjects factors, and pretraining condition, level of variation, and story type of the first training example (PCO or OSS) as between-subjects factors. A significant effect of test section confirmed that test scores improved from pretest to posttest, $F(1,203)=59.58$, $p<.001$.

Neither the main effect of pretraining condition nor that of training condition was significant, nor did either of these factors interact significantly with test section, *p*s>.15. Thus, none of the pretraining conditions, nor either of the training

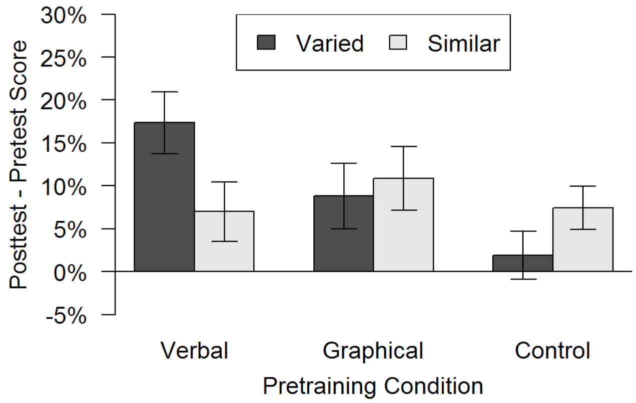

Figure 3: Change in test score from pretest to posttest, by pretraining and training condition.

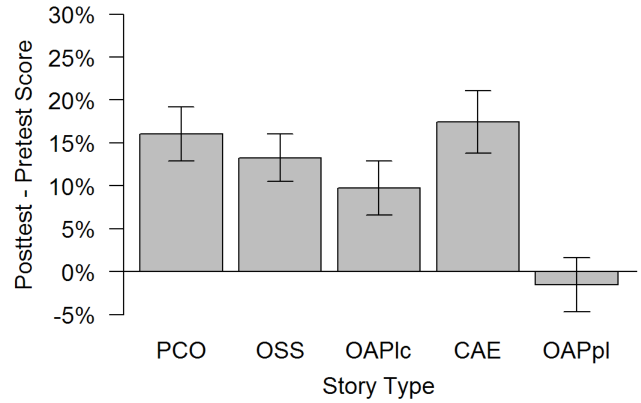

Figure 4. Change in test score from pretest to posttest, by story type.

conditions, led to greater overall improvement at posttest. However, critically, a significant 3-way interaction between pretraining condition, training condition, and test section was found, $F(2,203)=3.46$, $p=.033$. As shown in Figure 3, in the verbal pretraining condition, participants who received varied examples showed more improvement at posttest than those who received similar examples (varied: 17.4%, similar: 7.0%). However, this trend was absent, or even reversed, in the graphical (varied: 8.8%, similar: 10.9%) and control (varied: 1.9%, similar: 7.4%) conditions. To better understand this interaction, change in accuracy from pretest to posttest was compared between the similar and varied conditions separately for each of the pretraining conditions. These comparisons revealed a significant advantage of varied over similar examples in the verbal pretraining condition, $t(70.85)=2.07$, $p=.042$, and no difference between varied and similar examples in the other two pretraining conditions, *p*s>.10.

The main effect of story type and its interaction with test section were also significant, $F(4,836)=64.54$, $p<.001$ for the main effect and $F(4,836)=5.64$, $p<.001$ for the interaction. Pretest and posttest scores for each story type are shown in Figure 4. Scores improved at posttest on problems from both of the story types used during training (PCO: 16%, OSS: 13%) as well as two of the novel story types not seen in any version of the training (OAPlc: 10%, CAE: 17%). However, for the OAPpl problems, participants showed lower accuracy at posttest (-2%). This decrement in accuracy was driven by participants whose first training example belonged to the cross-mapped story type, PCO, as confirmed by a significant 3-way interaction of section and story type with story type of the first training example, F(4,812)=3.45, $p=.008$. Other significant effects (interactions of pretraining with story type, of these two factors with first training example, and of level of variation and first training example with test section) do not bear on the research questions and so are omitted for brevity.

Discussion

Participants who received verbal pretraining showed more improvement at posttest following training with varied rather than similar examples, while this effect was absent, and tended to reverse, in the control condition. This finding supports the hypothesis that greater attention to structural relations increases benefits of variation by enabling learners more easily to perceive structure shared by disparate examples. The alternate possibility mentioned in the Introduction, that such attention could improve the relative benefits of similar examples by helping learners to ignore superficial commonalities among such examples, was not supported.

This result may elucidate the interpretation of previous studies finding greater benefits of variation among learners with greater prior domain knowledge (Braithwaite & Goldstone, 2012, under review; Guo et al., 2013). In particular, the present findings are consistent with the possibility that such interactions of prior knowledge and variation are mediated by effects of prior knowledge on attention to structural relations, with more knowledgeable learners being better able to attend to such relations and therefore better able to profit from varied examples. Learners with less prior knowledge, conversely, can profit more from similar examples. Superficial similarities among examples can serve a scaffolding function, drawing attention to structural correspondences that would not otherwise be evident (Kotovsky & Gentner, 1996).

Verbal pretraining had the predicted effect of increasing the relative benefits of varied examples, while graphical pretraining did not. The effectiveness of verbal pretraining in this regard is, perhaps, not surprising. Several studies have found that exposure to relational terminology can increase attention to relational structure (Loewenstein & Gentner, 2005). However, the utility of diagrams for promoting analogical transfer in several prior studies (Catrambone, Craig, & Nersessian, 2006) suggests that under some circumstances, diagrams can have the same effect. Why, then, did graphical pretraining in the present study *not* lead to an advantage of varied over similar examples?

One possibility is that the combination lock diagrams were hard to understand or unsuitable for the task. However, the fact that graphical pretraining increased training accuracy relative to verbal pretraining or control argues against this possibility. A more plausible alternative is that the diagrams served as a crutch, rather than as a scaffold for learning. The diagrams may have been transparent enough that participants could easily perform the pretraining and training tasks without deeply processing the structure represented by the diagrams. Thus, the diagrams might facilitate performance during training but not produce a lasting increase in attention to underlying structure, and thus not increase the benefits of varied examples. While admittedly speculative, this account fits into a larger body of research indicating that facilitating task performance may not always promote the analytical mindset required for some kinds of learning, such as generalization and transfer of abstract concepts (Oppenheimer, 2008).

Our findings have practical implications for instruction in abstract concepts in fields such as mathematics and science. Differences in the effects of variation depending on learners' prior knowledge suggest that, when selecting illustrative examples, educators should take the differing needs of individual learners into consideration. However, preparing different sets of examples for different students might be impractical or undesirable for many reasons. The present results imply that it may be possible to prepare all learners to benefit from highly varied examples, regardless of prior knowledge, through instruction that focuses attention on relevant structural relations. Of course, implementation of such preparatory instruction requires time and resources. Whether the potential benefits would justify the costs is an important question to be addressed by future research.

The instructional implications of our findings echo those of proposals in educational psychology asserting that formal instruction can be more effective if learners are first prepared for it by learning about the critical features or dimensions involved in the domain (Schwartz & Martin, 2004). However, while these proposals emphasize discovery of these features or dimensions through exploration of contrasting cases, our approach was to provide an explicit framework which could increase the salience of the underlying structure. Whether appropriate instructional manipulations could lead learners to discover such structure for themselves, and whether such an approach would yield advantages over explicit scaffolding, are also promising areas for future research.

Acknowledgements

This research was supported in part by National Science Foundation REESE Grant No. 0910218 and Department of Education IES Grant No. R305A1100060.

References

Braithwaite, D. W., & Goldstone, R. L. (under review). Effects of variation and prior knowledge on abstract concept learning. *Cognition and Instruction*.

Braithwaite, D. W., & Goldstone, R. L. (2012). Inducing mathematical concepts from specific examples: The role of schema-level variation. *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 138–143). Austin, TX: Cognitive Science Society.

Catrambone, R., Craig, D. L., & Nersessian, N. J. (2006). The role of perceptually represented structure in analogical problem solving. *Memory & Cognition*, *34*(5), 1126–32.

Chen, Z., & Mo, L. (2004). Schema induction in problem solving: A multidimensional analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 583–600.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*(2), 121–152.

Day, S. B., Goldstone, R. L., & Hills, T. (2010). The effects of similarity and individual differences on comparison and transfer. *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society* (pp. 465–470). Portland, Oregon: Cognitive Science Society.

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., & Kurtz, K. J. (2005). Relational categories. In W. K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. W. Wolff (Eds.), *Categorization Inside and Outside the Laboratory*. Washington, DC: APA.

Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science*, *10*(3), 277–300.

Guo, J., Yang, L., & Ding, Y. (2013). Effects of example variability and prior knowledge in how students learn to solve equations. *European Journal of Psychology of Education*, *March*.

Hahn, U., Bailey, T. M., & Elvin, L. B. C. (2005). Effects of category diversity on learning, memory, and generalization. *Memory & Cognition*, *33*(2), 289–302.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, *38*(1), 23–31.

Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development*, *67*(6), 2797–2822.

Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology*.

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*. Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, *12*(6), 237–41.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*(3, pt 1), 353–363.

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction*, *22*(2), 129–184.