Routledge
Taylor & Francis Group

Effects of Variation and Prior Knowledge on Abstract Concept Learning

David W. Braithwaite
Carnegie Mellon University

Robert L. Goldstone
Indiana University

Learning abstract concepts through concrete examples may promote learning at the cost of inhibiting transfer. The present study investigated one approach to solving this problem: systematically varying superficial features of the examples. Participants learned to solve problems involving a mathematical concept by studying either superficially similar or varied examples. In Experiment 1, less knowledgeable participants learned better from similar examples, while more knowledgeable participants learned better from varied examples. In Experiment 2, prior to learning how to solve the problems, some participants received a pretraining aimed at increasing attention to the structural relations underlying the target concept. These participants, like the more knowledgeable participants in Experiment 1, learned better from varied examples. Thus, the utility of varied examples depends on prior knowledge and, in particular, ability to attend to relevant structure. Increasing this ability can prepare learners to learn more effectively from varied examples.

A large part of formal education is dedicated to abstract concepts—that is, concepts characterized by underlying structure rather than surface features, such as conservation of energy in physics, supply and demand in economics, or combinations and permutations in mathematics. Such concepts are powerful because they capture regularities among situations that bear little superficial resemblance to each other. For example, conservation of energy in physics applies equally well to springs, bouncing balls, or pendulums. However, abstract concepts can be difficult to understand when presented in abstract form without any concrete context (Cheng, Holyoak, Nisbett & Oliver, 1986; Nuñes, Schliemann & Carraher, 1993). Thus, educators often rely on concrete examples to facilitate learning (Eiriksdottir & Catrambone, 2011; Nathan, 2012). Indeed, several studies have shown facilitative effects of learning abstract concepts from concrete examples rather than from abstract descriptions alone (De Bock, Johan, van Dooren, Roelens, & Verschaffel, 2011; Gentner, Loewenstein & Thompson, 2004; Rawson, Thomas & Jacoby, 2015).

Correspondence should be address to David W. Braithwaite, Department of Psychology, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213. E-mail: baixiwei@gmail.com

However, it can be challenging for learners to focus on the underlying structure of concrete examples and to distinguish between surface features that are relevant and irrelevant with respect to this structure. Failure to do so can inhibit subsequent transfer to novel instances that share the same structure but not the same surface features (Bassok, 1996; Bassok & Holyoak, 1989; Belenky & Schalk, 2014; Harp & Mayer, 1998; Kaminski, Sloutsky, & Heckler, 2008, 2013; LeFevre & Dixon, 1986; see also Mayer & Griffith, 2008; Son & Goldstone, 2009) or cause inappropriate transfer to cases with similar surface features but different structure (e.g., Chang, 2006). How can learners enjoy the potential benefits of concrete examples while avoiding their potential negative effects on transfer?

The present study explores the effectiveness of *variation among examples* as a solution to this problem. Specifically, studying examples that are highly diverse with respect to their superficial features but share the same underlying structure may lead to superior transfer of knowledge. This approach relates to the selection of examples rather than the types of interventions in which they are embedded and so may be viewed as complementary to approaches such as combining examples with abstract descriptions (Cheng et al., 1986; Gentner et al., 2004; Rawson et al., 2014), fading from examples to abstract descriptions (Braithwaite & Goldstone, 2013; Fyfe, McNeil, Son & Goldstone, 2014; Goldstone & Son, 2005), or explicitly comparing examples (Catrambone & Holyoak, 1989; Gentner, Loewenstein & Thompson, 2003; Gick & Holyoak, 1983; Rittle-Johnson, Star & Durkin, 2009). As reviewed below, variation among examples offers potential benefits but also potential drawbacks, and the tradeoff between these may depend on factors such as prior knowledge. This study aims to clarify whether, and when, variation among examples can effectively promote transfer of abstract concepts to novel instances.

PROMISE OF VARIATION

In considering the potential effects of variation on transfer, we make two key assumptions. The first assumption is that transfer depends on the concept representations,[1] if any, that learners derive from the examples. Specifically, concept representations based on underlying structure rather than superficial features are more likely to support transfer because such representations capture what all instances of the target concept have in common (Belenky & Schalk, 2014; Day & Goldstone, 2012; Gentner, 1983; Gick & Holyoak, 1983; for alternative perspectives, see Lobato, 2006; Wagner, 2010). The second assumption is that concept representations are derived from examples in part by detecting commonalities among the examples. That is, learners are likely to incorporate features shared by multiple examples rather than those idiosyncratic to individual examples into their concept representations[2] (Chen & Mo, 2004; Gick & Holyoak, 1983; Kuehne,

---

[1]Concept representations are sometimes referred to as "schemas," and the process of deriving them from multiple examples as "schema abstraction" (Bernardo, 2001; Elio & Anderson, 1981, 1984; Gick & Holyoak, 1983; Reed, 1989). The term "concept representation" is preferred to schema here to avoid confusion with our somewhat different use of the term schema to describe our experimental stimuli. See note 4.

[2]Of course, studying examples of different concepts may draw attention to features that distinguish concepts, instead of or in addition to commonalities within each concept (Birnbaum, Kornell, Bjork, & Bjork, 2012; Chang, 2006; Goldstone, 1996; Kang & Pashler, 2012; Rosch & Mervis, 1975; Taylor & Rohrer, 2010; Vanderstoep & Seifert, 1993). Furthermore, studying both examples and nonexamples of a concept may highlight criterial features that determine whether or not the concept applies rather than highlighting commonalities among examples more generally (Chang, 2006; Gick & Paterson,

Forbus, Gentner & Quinn, 2000; Medin & Ross, 1989; Rosch & Mervis, 1975; Skorstad, Gentner & Medin, 1988).

Together, the above assumptions suggest a danger to studying superficially similar examples of a concept. Superficial features shared by multiple examples may become part of learners' concept representations, inhibiting transfer to novel instances that do not share the same features. There is some empirical support for this conclusion (Ben-Zeev & Star, 2001; Chang, Koedinger, & Lovett, 2003; see also Ross, 1984). In one study, after practicing using multiple mathematical algorithms to solve example problems with different superficial features, undergraduate students were tested on novel problems (Ben-Zeev & Star, 2001). Although the test problems could be solved using any of the learned algorithms, students tended to use each algorithm mainly to solve problems that superficially resembled the examples with which that algorithm had been practiced and not to solve superficially dissimilar problems.

By contrast, studying varied examples, which by definition have fewer superficial features in common, should help to highlight shared underlying structure, thereby promoting transfer. Studies have supported this prediction in a wide range of learning domains, including mathematics (Braithwaite & Goldstone, 2012; Chen & Mo, 2004; Paas & Van Merriënboer, 1994), statistics (Chang et al., 2003; Chang, 2006; Quilici & Mayer, 1996, 2002), science (Corbalan, Kester, & van Merriënboer, 2009; Day, Goldstone, & Hills, 2010), and language (Gómez, 2002; Onnis, Monaghan, Christiansen, & Chater, 2004; Perry, Samuelson, Malloy and Schiffer, 2010; Rost & McMurray, 2010). For example, Chen & Mo (2004, Experiment 3) trained participants to solve arithmetic word problems sharing a common general principle. When semantic details of the training examples were varied, participants were more able to transfer the principle to novel problems with the same underlying structure but different semantic details. Similarly, in a study of Chang et al. (2003), participants practiced using each of several types of statistical graphs to solve multiple example problems. The examples for a given graph type were either similar or varied with respect to superficial features, such as cover story and wording. Varied examples led to more accurate selection of graph types for novel problems on a subsequent test.

PERILS OF VARIATION

Despite the potential facilitative effects of variation on transfer, a possible drawback is that variation may make learning more difficult. First, deriving a concept representation from varied examples depends on attending to and encoding their shared underlying structure in the absence of shared superficial features. However, if the shared structure is not obvious or overt, the examples may appear so different that learners fail to notice their shared structure at all, let alone transfer it to novel instances (Gentner et al., 2003; Gick & Holyoak, 1983). Furthermore, variation with respect to superficial features may increase cognitive load during learning (Paas & Van Merriënboer, 1994; van Merriënboer & Sweller, 2005) or draw attention toward superficial features and away from more important structural features (Hammer, Bar-Hillel, Hertz, Weinshall, & Hochstein, 2008). Either of these effects might again inhibit concept learning. Consistent with

---

1992; Große & Renkl, 2007; Namy & Clepper, 2010). However, because the present study is concerned mainly with effects of variation among examples of a single concept, we focus on extraction of commonalities rather than detection of distinctive or criterial features.

these considerations, variation among examples has failed to yield improved transfer in several studies (Corbalan et al., 2009 system control condition; Gick & Holyoak, 1983, Experiment 4; Reed, 1989; Renkl, Stark, Gruber, & Mandl, 1998) and in others has yielded slower learning or poorer performance during study (Braithwaite & Goldstone, 2012; Chen & Mo, 2004; Posner & Keele, 1968).

Furthermore, according to the theory of *progressive alignment*, superficially similar examples can actually facilitate transfer in some circumstances (Gentner, 2010; Gentner, Anggoro, & Klibanoff, 2011; Gentner, Loewenstein, & Hung, 2007; Kotovsky & Gentner, 1996). This theory states that learners may initially attend to superficial similarities between examples, but discovery of superficial correspondences may draw attention to deeper structural correspondences, leading to the discovery of shared structure that would have gone unnoticed in the absence of the superficial similarities. Kotovsky and Gentner (1996) found that 4-year-old children were unable to match pairs of visual arrays with shared underlying structure (e.g., symmetry) but different superficial features (e.g., one array symmetric with respect to sizes of its elements, the other symmetric with respect to color of its elements). However, after practicing matching arrays that shared both structural and superficial features (e.g., both arrays symmetric with respect to size, or both symmetric with respect to color), children subsequently succeeded in matching arrays that shared structure alone. Apparently, practice with examples sharing both superficial features and underlying structure enabled subsequent recognition of the same underlying structure even across pairs of superficially dissimilar instances.

Several other studies have found suggestive evidence that initial study of similar examples can promote subsequent transfer (Elio & Anderson, 1984; Guo & Pang, 2011; Guo, Yang, & Ding, 2013; Rau, Aleven, & Rummel, 2010). For example, in a study of Rau et al., 2010, fifth and sixth grade students studied a variety of fractions problems using several graphical representations (e.g., pie charts, number lines). The problems were either blocked by representation type (e.g., several pie chart problems followed by several number line problems) or interleaved (e.g., switching frequently between pie chart and number line problems). Blocked sequencing led to superior performance on a subsequent assessment involving identification and ordering of fractions using novel graphical representations. Although Rau et al., 2010) manipulation strictly relates to example sequencing rather than variation per se, their results are consistent with the idea that studying several superficially similar examples (i.e., problems involving the same representational type) at a time can improve concept learning and transfer.

In summary, evidence is inconsistent regarding the benefits of learning from varied examples, with some studies showing such benefits, others failing to do so, and some even finding advantages for similar over varied examples. These results suggest that any simple prescription regarding variation may be problematic and point to the importance of specifying conditions under which variation is most likely to benefit learners.

VARIATION AND PRIOR KNOWLEDGE

One factor that may affect the relative benefits of varied and similar examples is learners' prior knowledge. Relative to learners with less domain knowledge, more knowledgeable learners are more sensitive to structural features relevant to the domain (Chi, Feltovich, & Glaser, 1981; Chi & VanLehn, 2012; see also Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Novick &

Holyoak, 1991). For example, compared to novices, physics experts are more likely to attend to and encode structural features (e.g., equivalence of total energy before and after a change in a physical system) in addition to superficial features (e.g., that fact that the system involves a spring) and to use structural features to categorize problems and draw inferences about them (Chi et al., 1981). Greater sensitivity to structure could enable more knowledgeable learners to detect underlying structure shared by multiple examples, even in the absence of superficial similarities. Thus, prior knowledge could alleviate the potential drawbacks of variation while reducing the potential benefits of similarity. Conversely, less knowledgeable learners might be less able to detect shared structure in the absence of superficial similarities, thus increasing the detrimental effects of variation and the potential benefits of similarity, as predicted by the theory of progressive alignment. In short, the relative benefits of varied examples may be greater for more knowledgeable learners and those of similar examples greater for less knowledgeable learners.

Several studies investigating dimensions of instructional design analogous to variation support the above predictions (Große & Renkl, 2007; Rau et al., 2010; Rittle-Johnson et al., 2009). More knowledgeable learners benefited more from studying both correct and incorrect problem-solving procedures (Große & Renkl, 2007) or comparing multiple procedures (Rittle-Johnson et al., 2009), while less knowledgeable learners benefited more from studying only correct procedures (Große & Renkl, 2007), individual examples without comparison (Rittle-Johnson et al., 2009), or examples blocked rather than interleaved by representational format (Rau et al., 2010). Similarly, sixth but not eighth graders benefited from initially low variation when studying examples of geometry concepts (Guo & Pang, 2011; Guo et al., 2013). Finally and most directly relevant, using a paradigm similar to that of the present study, Braithwaite and Goldstone (2012) found suggestive evidence of greater benefits of variation among more knowledgeable learners for transfer of a mathematical concept. However, interpretation of their results was rendered problematic because the interaction of variation and prior knowledge was only marginally significant, and because more knowledgeable learners showed little evidence of learning due to training. The present study is an attempt to replicate and extend these findings using an improved experimental paradigm.

A few studies appear to contradict the above pattern by finding greater benefits of variation among less, rather than more, knowledgeable learners (Day et al., 2010; Quilici & Mayer, 1996). The apparent discrepancies among the above studies may relate to the diversity of measures used to assess prior knowledge, which include self-reported domain knowledge (Große & Renkl, 2007), pretest score (Braithwaite & Goldstone, 2012; Große & Renkl, 2007; Rau et al., 2010), pretest strategy use (Rittle-Johnson et al., 2009), scores on standardized tests of mathematical ability (Quilici & Mayer, 1996), grade level (Guo & Pang, 2011; Guo et al., 2013), and membership in regular or accelerated classes (Day et al., 2010). On the other hand, as we argue in the General Discussion, it is possible for a single underlying mechanism to produce all of the above, apparently contradictory results even absent any influence from such methodological differences. For the moment, we note that while the evidence is mixed, the majority of studies point to greater benefits of variation among more knowledgeable learners.

TASK ANALYSIS

The present study investigated effects of variation and prior knowledge on learning to solve mathematics story problems involving the concept of "sampling with replacement" (SWR; Table 1).

TABLE 1
Sampling with Replacement (SWR) Problems

Schema	Example	Options	Selections	Solution
People Choosing Objects (PCO)	*A group of friends is eating at a restaurant. Each person chooses a meal from the menu. (It is possible for multiple people to choose the same meal.) In how many different ways can the friends choose their meals, if there are 5 friends and 6 meals?*	Meals	Friends	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. (It is possible to play the same key more than once in a sequence.) How many different melodies are possible, if there are 4 keys in the set and 7 notes in each melody?*	Keys in the set	Notes in each melody	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. (It is possible for multiple rooms to be painted the same color.) In how many different ways can she paint the rooms, if there are 8 rooms and 3 colors?*	Colors	Rooms	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?*	Categories	Paranormal events	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. (It is possible for multiple provinces to be assigned to the same child.) In how many different ways can the provinces be assigned, if there are 5 provinces and 7 children?*	Children	Provinces	7^5

Note. Text is abbreviated from the actual text used in the experiments. For problems whose text differs between Experiments 1 and 2, the Experiment 2 version is shown here.

Each problem describes a situation in which a certain number of *selections* is made from a fixed set of *options* and asks how many distinct outcomes are possible in this situation. The answers are given by a simple formula: $(options)^{(selections)}$, that is, the number of options raised to the power of the number of selections. Using this formula requires instantiating the structural roles of options and selections with specific elements from the problems and then filling in the formula with the numbers of options and selections. Several SWR problems, together with the correct instantiations of structural roles with problem elements and the resulting correct answers, are shown in Table 1.

The relationship between the problem elements determines which problem element fills the role of options and which that of selections. Specifically, if exactly one of element A is chosen for each of element B, then element A fills the role of options and element B that of selections. Multiple surface features of the problems can be used to infer this underlying relationship. For example, in the first problem in Table 1, the phrase "each person chooses a meal from the menu" implies that exactly one meal is chosen for each friend, even though the words "exactly one" are not used. Similarly, the statement that multiple people may choose the same meal implies that the number of people matched to a single meal is not necessarily exactly one. Importantly, the features available to support the mapping of elements to roles vary from problem to problem, making the instantiation of roles a "variable to constant" mapping task (Koedinger, Corbett, & Perfetti, 2012). For instance, one might guess from the first example in Table 1 that the direct object of the verb "choose" always fills the role of options, while a noun modified by the adjective "multiple" fills the role of selections, but neither of these words (or their near synonyms) appears consistently across all of the problems in Table 1. Thus, it is impossible consistently to solve the problems correctly by reliance on a few specific words or phrases.

An important challenge in assigning elements to roles is the need to focus on surface features that are relevant to the relationship just described, while ignoring irrelevant surface features. For example, the specific numbers involved, the relative sizes of these numbers, and the order in which the elements appear in the problem statement are all irrelevant to solving the problems. These features were randomized across problems to avoid potentially misleading correlations between them and the structural roles (Ben-Zeev & Star, 2001; Chang et al., 2003). Critically, semantic features of the problem elements, such as animacy, are also irrelevant. Thus, such features are at best unreliable, at worst misleading, cues for matching elements to roles (Markman & Stilwell, 2001). For instance, the first problem in Table 1 might suggest the intuition that options ("meals") are inanimate and selections ("friends") animate. However, this intuition would be unhelpful for the second, third, and fourth problems, in which both problem elements (e.g., "colors," "rooms") are inanimate, and misleading for the last problem, in which the options ("children") are animate and the selections ("provinces") inanimate.[3] Semantic features such as animacy strongly influence how people map between concrete situations and mathematical models (Bassok, Chase, & Martin, 1998; Bassok, Wu, & Olseth, 1995; Blessing & Ross, 1996; Fisher, Borchert, & Bassok, 2011; Martin & Bassok, 2005; Ross & Kilbane, 1997; Ross, 1987, 1989). Thus, participants were

---

[3]The common perception that a province will be chosen for each child, so that the provinces fill the role of options and the children that of selections, is incorrect because it rules out the possibility of individual children receiving multiple provinces, which is expressly allowed, and allows the possibility of some provinces remaining unassigned, which is expressly ruled out.

expected to rely on such features despite their formal irrelevance, and this reliance was expected to be a source of difficulty in solving the problems.

As illustrated above, solving an SWR problem using the formula requires *instantiating* the concept of SWR—that is, matching specific elements of the problem to general structure roles in the concept. More generally, using an abstract, structured concept to draw inferences about a concrete situation often requires not only identifying the relevant concept but also instantiating the structure of that concept. For example, applying knowledge about the formula relating distance, rate, and time requires identifying which quantities correspond to each of these roles (Nathan, Kintsch, & Young, 1992). Similarly, analyzing a situation using the concepts of Signal Detection Theory requires instantiating the roles of "signal" and "noise" with specific elements of the situation (Bassok & Holyoak,1989; Ross, 1987, 1989 for other examples, see Son & Goldstone, 2009). Many studies regarding variation, however, have focused mainly on effects of variation on ability to identify relevant concepts (Ben-Zeev, & Star, 2001; Chang et al., 2003; Chang, 2006; Day et al., 2010; Elio & Anderson, 1984; Kotovsky & Gentner, 1996; Posner & Keele, 1968; Quilici & Mayer, 1996, 2002), while few have examined the issue of instantiation. An important contribution of the present study is to investigate effects of variation specifically on subsequent instantiation of an abstract concept. SWR was deemed to be an appropriate basic case for such an investigation because it is minimally complex, in the sense of involving only a single relationship with only two roles.

HYPOTHESES AND EXPERIMENTS

The experiments described below were designed to test two hypotheses. First, based on the theoretical considerations and empirical evidence detailed above, in particular the findings of Braithwaite and Goldstone (2012), studying varied rather than similar examples was expected to yield greater benefits for learners with more prior knowledge of the target concept (SWR). Thus, an interaction between prior knowledge and variation among studied examples was predicted, such that more knowledgeable learners would show superior transfer after studying varied examples, while less knowledgeable learners would show no such benefit, or even a benefit of similar examples. This hypothesis was tested in Experiment 1.

The second hypothesis relates to a possible explanation for the results predicted by the first hypothesis. Specifically, more knowledgeable learners might be more able to learn from varied examples because they are more able to notice shared structural features between examples even in the absence of superficial similarity. If this explanation is correct, then directly training learners to attend to and encode relevant structural features should increase the relative benefits of subsequently studying varied examples, yielding effects analogous to those of greater prior knowledge. This hypothesis was tested in Experiment 2.

EXPERIMENT 1

Experiment 1 was designed to test whether participants' prior knowledge of SWR would affect the relative benefits of variation among examples on transfer of knowledge about SWR. Participants were trained to solve SWR problems using either varied or similar examples and then were tested

on novel SWR problems to assess transfer. Previous research has indicated strong influences of formally irrelevant semantic features on learning and transfer of abstract concepts in mathematics (e.g., Bassok et al., 1995). Thus, our manipulation of variation among examples focused on variation with respect to semantic features, while our transfer assessment included problems with novel semantic features.

To test for a possible moderating effect of prior knowledge on effects of variation, participants were asked to report their previous experience with SWR. Self-reported experience with target concepts has moderated effects of learning from examples in previous work (Rawson et al., 2015) and was employed in the present study because of its close relation to our theoretical reasons for predicting a moderating effect. In particular, previous experience with SWR would likely increase sensitivity to the structural relations underlying SWR, and such sensitivity was expected to increase the relative benefits of varied examples. By contrast, more general measures of domain knowledge, such as scores on standardized tests of mathematics, might be relatively poor indicators of sensitivity to the specific structural relations relevant to SWR.

Method

Participants

For Experiment 1, 131 undergraduate students from Indiana University (14 male, 117 female; four aged under 18, 121 aged 18 to 21, and six aged over 21) participated in partial fulfillment of a requirement for an introductory psychology course. Forty-three participants were assigned to the similar examples condition and 88 to the two varied examples conditions (blocked: 44, interleaved: 44). Seventy-five participants were classified as having higher prior experience of SWR, and 56 as having lower prior experience, in the manner described under Procedure. The distribution of the different training conditions did not differ as a function of prior experience, $\chi^2(1) = 2.399, p = .121$.

Materials

Five semantic schemas,[4] each describing a different type of SWR situation, were created to serve as a basis for the experimental stimuli. The schemas were (a) people choosing objects (PCO), in which each of several people chooses once from several objects; (b) objects selected in sequence (OSS), in which a sequence of selections is made from several objects; (c) objects assigned to places (OAPlc), in which one of several objects is placed in each of several places; (d) categories assigned to events (CAE), in which each of several events is classified into one of several categories, and (e) objects assigned to people (OAPpl), in which each of several objects is assigned to one of several people. The schemas were designed to minimize semantic overlap among the types of elements filling the selections role in the different schemas (PCO: people,

---

[4]Past research has often employed the term "schema" to refer to relational concepts such as SWR. In the present study, the term schema is reserved for specific versions of abstract concepts defined by specific constraints on the types of elements that may fill their structural roles.

OSS: positions in temporal sequences, OAPlc: physical locations, CAE: events, OAPpl: inanimate objects). Further, OAPpl was designed to be particularly challenging for participants exposed to examples based on PCO, because these two schemas are semantically "cross-mapped" relative to each other, that is, they assign semantically similar elements to different roles. In particular, people fill the role of selections in PCO but of options in OAPpl. Cross-mapping can inhibit analogical transfer between concept instances (Gentner & Toupin, 1986; Ross, 1987, 1989; Ross & Kilbane, 1997). Thus, OAPpl problems provide a particularly stringent test of ability to instantiate roles based on underlying structure rather than superficial semantic features.

On the basis of each schema, a number of story problems were developed to serve as training examples and test problems. These problems were a modified and expanded version of the stimuli employed by Braithwaite and Goldstone (2012). One example problem belonging to each schema is shown in Table 1. The complete set of problems is available in the online Supplementary Information.

All training examples were based on the PCO and OSS schemas. For each of these schemas, four stories were developed for use as training examples, for a total of eight stories. The different stories for a given schema used different elements to fill the roles of options and selections. For example, one PCO story involved friends choosing meals at a restaurant (Table 1, first problem), while another involved consumers choosing pizza flavors in a marketing survey. Variation during training was manipulated by presenting examples based either on only one (similar examples condition) or both (varied examples condition) of the two schemas.

More specifically, the eight training example stories were used to create six training sets (i.e., two sets for each of three training conditions: similar, varied–blocked, and varied–interleaved). Each set contained four of the eight training example stories. (This number of examples was expected to be sufficient because previous studies of analogical transfer and mathematics concept learning have found successful transfer after study of similar numbers of examples or fewer, e.g., Bassok et al., 1995; Braithwaite & Goldstone, 2012; Chen & Mo, 2004; Gick & Holyoak, 1983; Novick & Holyoak, 1991.) In the similar condition, one set contained all four PCO, and the other all four OSS, stories. In the varied–blocked condition, one set contained two PCO followed by two OSS stories, while the other contained the other two OSS followed by the other two PCO stories. In the varied–interleaved condition, each set contained two PCO and two OSS stories in interleaved sequence, starting with PCO in one set and with OSS in the other. Each story appeared in exactly one of the sets for each condition, and each story always appeared in the same ordinal position. One of the six training sets was selected randomly for each participant.

All five schemas were used to develop additional stories for use as test problems. Thus, the test problems served as a strong test of transfer because they included both novel stories from the same schema(s) shown during training, as well as stories from novel schemas not shown during training. Two test stories were developed for each schema, for a total of 10 stories. These stories were divided into two test sets, each including one story for each schema. Within each test set, the stories were presented in a fixed order according to their schema, as follows: OAPlc-CAE-OAPpl-PCO-OSS. For each participant, one of the two test sets was selected randomly to serve as pretest, while the other served as posttest. The assignment of pretest and posttest problem sets was randomized across participants to reduce the chance that effects of training on posttest performance would depend on idiosyncrasies of particular test problems.

For each story in both the training and test sets, two problems were constructed—a "long" problem and a "short" problem—which differed in that the short problem was stated more briefly

A homeowner is going to repaint several rooms in her house. She chooses one color of paint for each of the rooms. (It is possible for multiple rooms to be painted the same color.)

In how many different ways can she paint the rooms, if there are 3 colors and 8 rooms?

○ 3^8

○ 8^3

Submit

FIGURE 1 Screenshot of a typical test trial (Experiment 1).

to avoid unnecessary repetition. The specific numbers involved and the order in which the key problem elements were mentioned in the problem statement were also varied between the long and short versions. The reason for including two problems for each story was to obtain a more stable measure of accuracy for each story and to reduce the chance of obtaining a perfect score by guessing. The long and short versions of each training example or test problem were always shown in immediate succession, with the long version shown first. Thus, each training set consisted of eight problems based on four stories, while each test set consisted of 10 problems based on five stories.

Procedure

At the beginning of the study, participants were shown a brief review of exponential notation. Participants then read either a "pretraining" passage describing how to solve SWR problems (without focusing on the specific task appearing throughout the rest of the experiment) or a passage regarding an unrelated mathematical topic. This manipulation was intended to simulate the effects of different levels of prior knowledge. However, because this manipulation produced neither significant effects on change scores nor significant interactions with variation, prior knowledge, or both, it is not described further here. Details regarding the pretraining are available in the online Supplementary Information.

After the pretraining, participants began the pretest. The problems from the assigned test problem set were presented one at a time. For each problem, participants had to choose between two responses of the form m^n and n^m, where m and n were the two numbers mentioned in the problem. Thus, their task was to determine which number should be the base and which the exponent. The order in which the noun phrases designating the key problem elements appeared in the problem statements (e.g., "meals," "friends") and the specific numbers attached to these noun phrases were determined quasirandomly for each trial so that these aspects would not provide reliable cues to the correct answers. A screenshot of a typical test trial is shown in Figure 1.

After completing the pretest, participants read a passage explaining how to solve SWR problems in the context of a specific SWR story. This passage explained how to solve several SWR problems using this story with varying numbers and then presented the general formula for SWR problems. The passage ended with the following summary: "All Sampling with Replacement

problems can be solved with this formula: **(OPTIONS)** [(TIMES)]. You just need to figure out what is the number of OPTIONS chosen from and what is the number of TIMES an option is chosen." The term "times" here refers to the structural role of selections. The full text of the training passage is provided in the online Supplementary Information.

Participants then began the training example problems. For each example, participants were first shown the long version problem and were asked two multiple choice questions requiring them to identify which problem elements corresponded to the number of options and times, respectively. They were then asked to choose the correct answer to the problem in the same way as in the pretest. Participants giving incorrect answers to any of the above questions received feedback and were required to give the correct answers before proceeding. The corresponding short version problem was then administered without the prompts to identify options and times (since the answers to these prompts would be the same as for the long version problem).

After completing the training examples, participants were shown a summary of the four training example stories and reminded of which problem elements corresponded to the "base" and "exponent" of the correct answers in each case. They were then asked to "describe, in as general a way as possible, how to decide which number should be the base and which number should be the exponent" and responded in open answer format. In contrast to previous studies employing similar measures to assess concept learning (Braithwaite & Goldstone, 2012; Chen & Mo, 2004; Gick & Holyoak, 1983), no effect of training on responses to this question were found, perhaps because the question could be answered perfectly (e.g., "the number of options should be the base and the number of times should be the exponent") simply by repeating key phrases from the passage shown at the beginning of training. Thus, analyses of these responses are not reported below.

Next, the posttest was administered in the same manner as in the pretest. After completing the posttest, participants answered the following question: "Before taking this survey, had you ever learned about Sampling with Replacement problems before?" This question was posed at the end of the study to ensure that participants knew the meaning of the term "Sampling with Replacement" and so would not underestimate their experience due to not understanding the term. Participants could respond "yes," "maybe," or "no." Those responding "yes" were classified as having more previous experience with SWR and all others as having less previous experience.

All parts of the study were presented via a computer interface. Participants were allowed to proceed at their own paces but could not go back to any part of the study after completing it. The study took an average of 14.8 min to complete (minimum: 5.9, maximum: 27.6).

Results

Reliability was assessed for each test set, combining cases in which the set was administered as pretest or posttest. Cronbach's α was low for both sets (.45, .55), apparently owing to data from the semantically cross-mapped OAPpl problems. Accuracy for OAPpl was negatively correlated with all other schemas, while accuracies for all other pairs of schemas were positively correlated, and removing data from OAPpl problems improved consistency for both test sets ($\alpha = .73, .65$). Nevertheless, the OAPpl data were included in subsequent analysis because we were interested in performance on a variety of different problems. Problem schema was included as a factor to allow for the possibility that schema moderated effects of training condition.
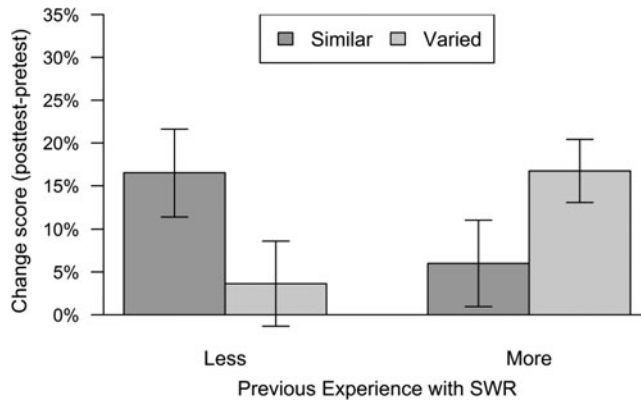
FIGURE 2 Change scores by previous experience and level of variation (Experiment 1). Error bars indicate standard errors.

Average accuracy (percentage correct) was 50.5% at pretest and 62.3% at posttest. Pretest accuracy was not significantly higher than chance (50.0%), $t(130) = .305$, $p = .761$, while posttest accuracy was significantly higher than chance, $t(130) = 6.986$, $p < .001$. The improvement in accuracy from pretest to posttest was significant, $t(130) = 5.00$, $p < .001$, $d = 0.58$.

Change scores were calculated by subtracting pretest from posttest accuracy separately for each schema for each participant. Initial analysis of change scores comparing the varied–blocked and varied–interleaved conditions found no significant effects involving condition, $ps > .5$, so these two conditions were combined into a single "varied" condition for subsequent analysis. The main effects of assignment of the test problem sets to pretest and posttest and of selection of the schema (PCO or OSS) used for the first training example were not significant, nor were the interactions of these factors with variation or previous experience with SWR or both, $ps > .15$, so these factors were excluded from further analysis. Change scores were then submitted to a $2 \times 2 \times 5$ mixed ANOVA, with level of variation (similar or varied) and previous experience with SWR (less or more) as between-subjects factors and problem schema (PCO, OSS, OAPlc, CAE, or OAPpl) as a within-subjects factor.

While the main effects of variation and previous experience did not reach significance, $ps > .75$, a significant interaction between these two factors was found, $F(1, 127) = 5.58$, $p = .020$, $\eta_g^2 = .011$. As shown in Figure 2, more experienced participants improved more on posttest in the varied condition (16.7%) than in the similar condition (6.0%), while less experienced participants showed the opposite trend (varied: 3.6%, similar: 16.5%).

Additionally, the main effect of schema was significant, $F(4, 508) = 12.50$, $p < .001$, $\eta_g^2 = .069$. Participants tended to show greater posttest improvement for schemas that were shown during training (PCO: 26.7%, OSS: 17.9%) than for those shown only at test (OAPlc: 14.9%, CAE: 18.3%, OAPpl: -19.1%). For OAPpl problems, scores were lower on posttest than on pretest, presumably because these problems were semantically cross-mapped relative to the PCO problems shown (to most participants) during training. Schema did not interact significantly with any other factor, $ps > .10$.

TABLE 2
Change Scores by Previous Experience, Level of Variation, and Trained Versus Untrained Schema
(Experiment 1)

Schema Type	Less Previous Experience		More Previous Experience	
	Similar Examples	Varied Examples	Similar Examples	Varied Examples
Trained schemas	19.6%	19.7%	15.0%	25.9%
Untrained schemas	15.8%	−7.1%	3.8%	10.6%

Given the significant effect of problem schema, an important question is whether the effects of previous experience and variation were limited to the schemas shown during training or also extended to schemas shown only at test. To investigate this question, each schema was classified with respect to each participant as either trained or untrained, depending on whether that participant was shown any training examples based on that schema. Change scores were then calculated separately for trained and untrained schemas. As shown in Table 2, among more experienced participants, varied examples led to greater improvement than similar examples for both trained and untrained schemas, while among less experienced participants, similar examples led to an advantage for untrained, but not for trained, schemas. Thus, trained schemas alone did not drive the effects observed in our main analysis. Confirming this conclusion, when only data from untrained schemas were analyzed, the interaction between previous experience and variation was still significant, $F(1, 127) = 6.79$, $p = .010$, $\eta_g^2 = .051$, and was directionally identical to that observed for the entire dataset.

Discussion

Experiment 1 found a significant interaction of previous experience and level of variation such that learners with less experience of SWR learned better from similar examples, while those with more experience learned better from varied examples. Many previous studies have focused on the benefits either of variation (Chang et al., 2003; Chen & Mo, 2004; Corbalan et al., 2009; Day et al., 2010; Gick & Holyoak, 1983; Paas & Van Merriënboer, 1994; Quilici & Mayer, 2002) or of similarity across examples (Graham, Namy, Gentner, & Meagher, 2010; Guo & Pang, 2011; Kotovsky & Gentner, 1996; Rau et al., 2010). An important contribution of the present study is to show both types of benefit within the same study but for different types of learners.

Why did less experienced participants learn relatively more from similar, rather than varied, examples? This result is consistent with the theory of progressive alignment (Gentner, 2010; Gentner et al., 2007, 2011; Kotovsky & Gentner, 1996), according to which superficial similarities between examples can draw learners' attention to deeper structural correspondences and thereby to shared structure. For example, learners who studied one PCO example involving friends choosing meals (Table 1, first problem), then another involving consumers choosing pizza flavors, might easily draw correspondences between friends and consumers and between meals and pizza flavors based on semantic resemblance alone. However, drawing these correspondences might lead learners to notice a deeper resemblance between the relation of friends to meals and that of consumers to pizza flavors (i.e., in each case, exactly one of the latter is chosen by each

of the former). Having noticed this deeper relational resemblance, learners might subsequently recognize it even in the context of problems that do not superficially resemble the studied examples. This effect is likely to be strongest for learners who would otherwise experience the greatest difficulty in noticing shared underlying structure—that is, less knowledgeable learners.

A possible alternate explanation is that similar examples only facilitated relatively shallow learning. For example, participants who studied the PCO examples mentioned above might learn the shallow rule that "the people are the selections and the number of people should be the exponent in the correct answer." Learning this rule would improve performance on test problems belonging to the PCO schema, although not on problems belonging to other schemas. More generally, this explanation predicts a greater advantage of similar examples with respect to trained schemas compared to untrained schemas. However, in fact, the advantage of similar examples among less experienced participants was greatest with respect to untrained schemas. This result suggests that similar examples promoted deep rather than shallow learning in this group, as predicted by the theory of progressive alignment.

Why did the relative benefits of varied examples increase with greater previous experience of SWR? A possible account outlined in the Introduction involves the assumption that the more experienced participants were more sensitive to structural features in the examples, in the same way that experts are typically more sensitive than novices to the structural relations critical to their domains of expertise (Chi et al., 1981; Chi & VanLehn, 2012; see also Chi et al., 1989; Novick & Holyoak, 1991). This greater sensitivity to structural features could alleviate a key drawback of varied examples—the difficulty of seeing what such examples have in common—by increasing the salience of the structural features shared by the examples. This account leads to a novel prediction: Directly increasing learners' attention to structural features prior to training should produce effects analogous to those of previous experience, namely, improved learning from varied examples relative to similar examples. This prediction was tested in Experiment 2.

Finally, the possibility of inaccuracy in participants' self-reported experience with SWR deserves mention. Indeed, factors other than actual experience with the concept, such as understanding the meaning of the term "Sampling with Replacement," poor memory of past experience, or inflated opinions of one's knowledge may have influenced these reports. Even if the reports were accurate, experience with SWR may have been correlated with other, unobserved variables, so we cannot be certain that experience with SWR per se led to the observed interaction. To demonstrate a causal relationship requires directly manipulating, rather than observing, participants' prior knowledge of SWR. Just such a manipulation was performed in Experiment 2.

EXPERIMENT 2

The goal of Experiment 2 was to test the prediction that increasing learners' attention to structural features of examples would yield effects analogous to those of greater prior experience with SWR in Experiment 1 (i.e., increased benefits of varied, relative to similar, examples). Prior to training in solving SWR problems, experimental participants first received a "pretraining" intervention (distinct from the pretraining of Experiment 1, as described below) in which they practiced identifying critical structural relationships in SWR problems, while control participants received no such pretraining. Experimental and control participants both then received a training intervention similar to that of Experiment 1, using either similar or varied examples. Varied

(a)

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. (It is possible to play the same key more than once in a sequence.)

How many different melodies are possible, if there are 9 keys in the set and 6 notes in each melody?

Which picture is a better representation of the problem?
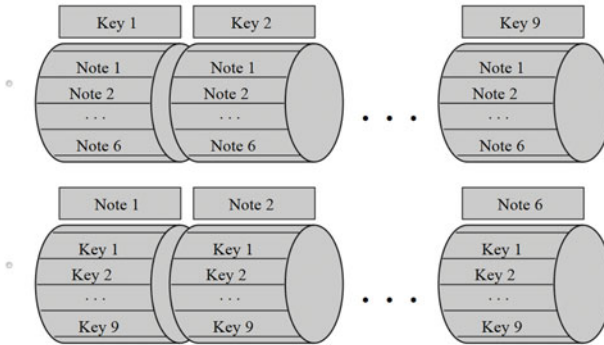
(b)

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. (It is possible to play the same key more than once in a sequence.)

How many different melodies are possible, if there are 9 keys in the set and 5 notes in each melody?

Which of these statements is correct?

○ For **EACH** of the keys in the set, **ONE** of the notes in each melody is chosen
○ For **EACH** of the notes in each melody, **ONE** of the keys in the set is chosen

FIGURE 3 Pretraining interface (Experiment 2). (A) Graphical, (B) Verbal.

examples were expected to show a greater advantage over similar examples in the experimental conditions compared to the control condition.

Two versions of the pretraining intervention were developed. The first version employed diagrams to represent graphically the relationship between elements of SWR problems (Figure 3A). Visual representations such as diagrams and graphs can be effective tools for highlighting underlying structure in concrete examples of concepts (Braithwaite & Goldstone, 2013; Catrambone, Craig, & Nersessian, 2006; Gick & Holyoak, 1983; Scheiter, Gerjets, & Schuh, 2010; Stern, Aprea & Ebner, 2003; Woodward et al., 2012). The second version aimed to highlight verbally such structure using relational language (e.g., "for each of the . . . , one of the . . . is chosen," Figure 3B). Like diagrams, relational words and phrases can increase learners' attention to structural relations (Gentner et al., 2011; Loewenstein & Gentner, 2005; Son, Doumas, & Goldstone, 2010; Son, Smith, Goldstone, & Leslie, 2012). We made no prediction regarding the relative efficacy of graphical versus verbal pretraining.

Method

Participants

For Experiment 2, 215 undergraduate students from Indiana University (82 male, 133 female; three aged under 18, 198 aged 18 to 21, and 14 aged over 21) participated in the experiment in partial fulfillment of a requirement for an introductory psychology course. Seventy-one participants were assigned to the control pretraining condition, 71 to the experimental–graphical condition, and 73 to the experimental–verbal condition. Within each pretraining condition, participants were assigned randomly to study either similar or varied examples (control condition: similar = 36, varied = 35; experimental–graphical condition: similar = 36, varied = 35; experimental–verbal condition: similar = 37, varied = 36). The graphical and verbal pretraining conditions will be referred to jointly as the "experimental" conditions.

Materials

The four PCO and four OSS stories used for training examples in Experiment 1 were used as training examples in Experiment 2, with minor changes in wording. Two new stories for each schema (PCO and OSS) were also added, yielding a total of six stories per schema, or 12 stories total. Each participant was shown only six of these examples, selected as follows. First, each participant was assigned randomly to receive either similar or varied examples. Those in the similar condition were shown all six examples of one schema and no examples of the other schema. Participants in the varied condition were shown three examples of each schema, presented in interleaved sequence. The blocked sequence used in Experiment 1 was not used in Experiment 2. Whether the first training example (and, for the similar condition, all subsequent examples) was based on PCO or OSS was counterbalanced within each condition. The specific examples shown and the order in which they were presented were determined randomly for each participant, subject to the above constraints. In the experimental conditions, the same set of examples was used for both pretraining and training, and the examples were presented in the same order in both cases. In contrast to Experiment 1, only long problem versions of the pretraining/training examples were shown.

The test sets used in Experiment 1 were also used for the pretest and posttest in Experiment 2, with minor changes in wording. However, two new OAPpl stories were added at the end of each test set, for a total six stories in each set. Just as in Experiment 1, the test sets included a long version and a short version of each problem, for a total of 12 problems in each set. One set was selected randomly to serve as the pretest and the other as the posttest for each participant. The complete set of training examples and test problems employed in Experiment 2 is available in the online Supplementary Information.

Procedure

In all conditions, participants began by completing the pretest and ended by completing the posttest, with both pretest and posttest administered in the same way as in Experiment 1. Between the pretest and posttest, participants in the experimental conditions received a pretraining

intervention followed by a training intervention, while those in the control condition received only a training intervention.

During the pretraining, participants first read a brief exposition (available in the online Supplementary Information) that described the common structure shared by all SWR problems, without explaining how to solve such problems. The six training examples were then presented one at a time. For each example, participants chose between two descriptions of the problem structure (Figure 3), and were required to explain their answer. In the experimental–graphical condition (Figure 3A), 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. In the experimental–verbal condition (Figure 3B), 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. After each response, participants received feedback, including explanations of why incorrect answers were incorrect, and reviewed all six examples together after completing each one separately.

After the pretest and, if applicable, the pretraining, participants began the training. They first read an expository passage explaining how to solve SWR problems (available in the online Supplementary Information). This passage was similar to that employed in Experiment 1 except that different terms (i.e., "alternatives" and "selections" instead of "options" and "times") were used to refer to the structural roles of SWR. The six training examples were then presented one at a time, and for each one, participants chose between the two possible final answers just as in the test sections. Participants received feedback on each response, just as in the pretraining, and reviewed all examples together after completing them separately. At the end of the training, participants answered two open comprehension questions. However, because experimental condition had no effect on responses to these questions, the results are not reported below.

As in Experiment 1, the study was presented via a computer interface, and participants proceeded at their own paces but could not return to already-completed problems. The study took an average of 22.9 min to complete (minimum: 9.8, maximum: 44.5). The experiment took longer in the experimental conditions (25.7 min) than in the control condition (19.9 min).

Results

As in Experiment 1, internal consistency was relatively low for both test sets (Cronbach's $\alpha =$ .59, .58) due to data from OAPpl problems. Removing these data improved consistency for both sets (Cohen's $\alpha =$ .74, .68), but they were nevertheless included in subsequent analysis, with problem schema included as a factor, for the reasons given in Experiment 1.

Average accuracy (percentage correct) was 58.5% on the pretest and 67.4% on the posttest. In contrast to Experiment 1, both pretest and posttest accuracy were higher than chance, that is, 50%, $t(214) = 7.04$, $p < .001$ for pretest and $t(214) = 12.7$, $p < .001$ for posttest. However, as in the previous experiments, accuracy increased significantly from pretest to posttest, $t(214) = 6.41$, $p < .001$, $d = 0.47$.

Change scores were computed as in Experiment 1. Initial analysis found no significant main effects, or interactions with pretraining condition or level of variation, of which test sets served as pretest and posttest, $ps > .15$, so this factor was excluded from subsequent analyses. Training
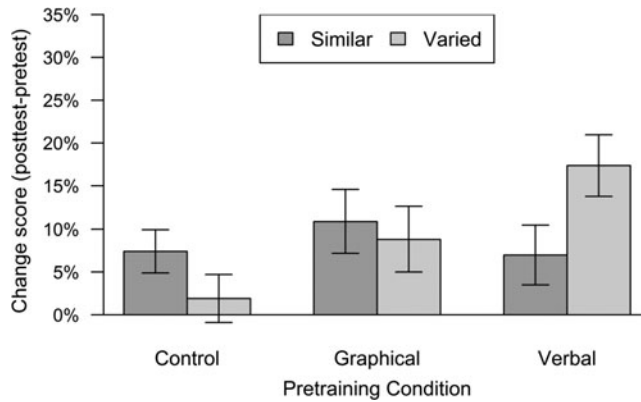
FIGURE 4 Change scores by pretraining condition and level of variation (Experiment 2). Error bars indicate standard errors.

version, that is, which schema was used for the first training example (and, in the similar examples condition, all other examples), did interact significantly with level of variation and so was included as factor (the primary finding of interest, i.e., the interaction of pretraining condition with variation, remained significant if training version was excluded from analysis). Thus, the final model used for analysis of the change scores was a $3 \times 2 \times 2 \times 5$ mixed ANOVA with pretraining condition (control, graphical, or verbal), level of variation (similar or varied), and training version as between-subjects factors, and problem schema (PCO, OSS, OAPlc, CAE, or OAPpl) as a within-subjects factor.

As in Experiment 1, the main effect of variation did not reach significance, $p > .9$. Additionally, despite a tendency for higher change scores in the experimental pretraining conditions (control: 4.7%, graphical: 9.9%, verbal: 12.1%), the main effect of pretraining was not significant either, $p > .15$. However, critically, a significant interaction of pretraining condition with level of variation was found, $F(2, 203) = 3.46$, $p = .033$, $\eta_g^2 = .007$. (This interaction was still significant if the graphical and verbal pretraining conditions were combined into a single "experimental" condition, $F(1, 207) = 4.09$, $p = .044$, $\eta_g^2 = .004$.) As shown in Figure 4, studying similar examples led to greater posttest improvement in the control condition (similar: 7.4%, varied: 1.9%), but this advantage was attenuated in the graphical pretraining condition (similar: 10.9%, varied: 8.8%), and reversed in the verbal pretraining condition, which showed a clear advantage for varied examples (similar: 7.0%, varied: 17.4%).

The main effect of problem schema was also significant, $F(4, 812) = 5.73, p < .001, \eta_g^2 = .022$. Replicating Experiment 1, participants improved on schemas shown during training (PCO: 16.0%, OSS: 13.3%) more than on those encountered only at test (OAPlc: 10.0%, OAPpl: −1.5%), with one exception (CAE: 17.4%). Furthermore, significant interactions were found between training version with level of variation, $F(1, 203) = 4.81$, $p < .029$, $\eta_g^2 = .005$, and problem schema, $F(4, 812) = 3.45, p = .008, \eta_g^2 = .013$. Because training version had a different meaning depending on level of variation—version determined the schema (PCO or OSS) of all training examples in the similar condition but only of the first training example in the varied condition—change scores in the similar and varied conditions were examined separately to understand the above interactions.

TABLE 3
Change Scores by Level of Variation, Training Version, and Problem Schema (Experiment 2)

Problem Schema	Similar Examples		Varied Examples	
	PCO Examples Only	OSS Examples Only	PCO Example First	OSS Example First
PCO	26.3%	0.0%	27.6%	23.5%
OSS	21.0%	12.5%	11.5%	15.8%
OAPlc	19.7%	7.2%	7.3%	13.3%
CAE	21.7%	18.4%	17.7%	14.3%
OAPpl	−17.8%	1.6%	−10.4%	−6.1%
Overall	14.2%	8.0%	10.7%	12.1%

As shown in Table 3, in the similar examples condition, participants improved more on whichever of the two training schemas (PCO, OSS) they were shown during training. Furthermore, they improved more overall after studying PCO (14.2%) than OSS (8.0%) examples. In the varied condition, by contrast, participants showed equal improvement on problems belonging to the two training schemas and equal improvement overall, regardless of which schema was used for their first training example. Finally, decrement in performance on OAPpl problems was observed among all participants who studied PCO examples but not among those who studied only OSS examples in the similar examples condition.

Discussion

An intervention aimed at increasing attention to underlying structure differentially increased learning from varied, relative to similar, examples. Commonalities between varied examples are relatively difficult to detect, and failure to detect such commonalities may prevent learners from deriving a general representation of the concept exemplified by the examples. Drawing attention to structure, we argue, alleviates this difficulty by highlighting underlying structure that is shared by the examples. Thus, a certain level of attention to structural features appears to be a prerequisite for variation effectively to promote learning and transfer.

The relative benefits of varied and similar examples depended on instructional condition in Experiment 2 in just the same way as they depended on previous experience with SWR in Experiment 1. This result is consistent with the view that attention to underlying structure was the "active ingredient" causing the effects of previous experience in Experiment 1. More generally, the results suggest that knowledge of a domain can enable learners to derive greater benefits from varied examples by virtue of increasing their sensitivity to relevant structural features.

Importantly, Experiment 2 demonstrates the feasibility of increasing attention to structure through direct intervention. Thus, even learners who lack the requisite prior knowledge may become prepared to learn effectively from varied examples. The idea of preparing learners to learn from instruction is not new. For example, Schwartz and his colleagues have argued that exploratory activities that focus attention on underlying structure can improve learning from subsequent formal instruction (Schwartz, Bransford, & Sears, 2005; Schwartz, Chase, & Bransford, 2012; Schwartz, Chase, Oppezzo, & Chin, 2011; Schwartz & Bransford, 1998; Schwartz & Martin,

2004). An important contribution of the present study, however, is to demonstrate that highly directed activities, such as the pretraining interventions of Experiment 2, can have a similar effect, provided that the structure to which learners are expected to attend is correctly and clearly identified (see also Mayer, 2005).

The effectiveness of verbal pretraining for improving the benefits of varied examples is, perhaps, not surprising. This pretraining employed relational language (the sentence template "for each of the ___, one of the ___ is chosen"), which several studies have found can increase attention to underlying structure (Gentner et al., 2011; Loewenstein & Gentner, 2005; Son et al., 2010, 2012). However, the utility of diagrams for promoting analogical transfer in several studies (Catrambone et al., 2006; Gick & Holyoak, 1983) suggests diagrams can sometimes have the same effect. Why, then, did the graphical pretraining not lead to an advantage of varied examples? One possibility is that the lock diagrams were hard to understand. However, accuracy on the training problems actually was higher in the graphical (87%) than in the verbal (78%) and control (73%) pretraining conditions. Thus, the diagrams were likely well understood and may even have facilitated performance. More plausibly, the diagrams may have served as a crutch rather than as a scaffold for learning, allowing some participants to solve the problems more easily without deeply processing their structure. Thus, the diagrams might not produce a lasting increase in attention to underlying structure and thus not increase the benefits of varied examples. This account fits into a larger body of research indicating that facilitating performance during study may not always improve learning outcomes (Bjork, Dunlosky, & Kornell, 2013; Oppenheimer, 2008).

Considered together with Experiment 1, Experiment 2 yielded strong evidence for the psychological reality of the schemas used to generate the training and test stimuli. In both experiments, participants tended to improve more on schemas shown during training (PCO, OSS) than on novel schemas shown only at test (with the exception of CAE). Similarly, in Experiment 2, participants in the similar examples condition improved more on whichever of the two training schemas they studied. Additionally, those who studied PCO examples during training (but not those who did not) showed a decrement in accuracy on OAPpl problems, which were semantically cross-mapped (Gentner & Toupin, 1986) relative to PCO. In summary, the degree of match or mismatch between training example schemas and test problem schemas exerted a strong influence on ability correctly to solve the test problems. On the one hand, these findings validate our use of the schemas as a basis for manipulating variation among examples during training. On the other hand, they suggest that these or similar schemas can be a useful tool to organize the systematic creation of varied examples in educational practice.

GENERAL DISCUSSION

Summary of Findings

The present findings imply that the question of how variation during learning affects subsequent transfer may have no simple answer and suggest the importance of a related, more nuanced question: How do the effects of variation depend on learner knowledge? In Experiment 1, learners with less prior knowledge showed better transfer after studying similar examples, while those with more prior knowledge showed better transfer after studying varied examples. In Experiment

2, an intervention intended to increase a particular aspect of prior knowledge—awareness of and attention to relevant structural features—yielded directionally the same effects as greater prior knowledge in Experiment 1. Thus, it appears that the relative benefits of varied and similar examples are not fixed but instead depend on prior knowledge, and in particular, on sensitivity to structural features.

It is tempting to conclude that the benefits of varied relative to similar examples simply increase as prior knowledge increases. Indeed, as reviewed in the Introduction, several studies provide evidence consistent with this conclusion (Braithwaite & Goldstone, 2012; Große & Renkl, 2007; Guo & Pang, 2011; Guo et al., 2013; Rau et al., 2010; Rittle-Johnson et al., 2009). However, it is important to note that an interaction in the opposite direction—that is, relatively greater benefits of varied examples for less, not more, knowledgeable learners—is also plausible a priori. One of the key potential benefits of variation is that it highlights underlying structure and avoids drawing attention to superficial commonalities among examples, and this effect might be most beneficial for learners who have the most difficulty focusing on structure to begin with—namely, those with less prior knowledge. A few studies have obtained evidence supporting this possibility (Day et al., 2010; Quilici & Mayer, 1996). For example, among students in a regular level class, Day et al., 2010 found benefits of varying the topic domain used to illustrate the concepts of positive and negative feedback loops, while no such advantage appeared among the presumably more capable students in an accelerated class.

These apparently discrepant results suggest the need for a unifying theoretical account able to explain not only the observed interactions of prior knowledge and variation, but also why the pattern of such interactions might itself vary depending on circumstances. We now present such an account. The account is first described in qualitative terms and then formalized as a mathematical model. We demonstrate that the mathematical model is indeed able to accommodate all of the qualitative patterns of results in the studies mentioned above but is still sufficiently constrained so as to make predictions for future studies.

Theoretical Account

The key assumptions of our theoretical account are as follows:

1. For a learner to acquire the proper structural basis for a concept from examples, the learner must detect that there is a commonality between the examples and, having succeeded in this, must extract the structural basis underlying that commonality.
2. Superficial similarities among examples facilitate detection of commonalities, while variation facilitates extraction of underlying structure (Gick & Holyoak, 1983; Kotovsky & Gentner, 1996). In an earlier study using a similar paradigm to the present one (Braithwaite & Goldstone, 2012; see also Chen & Mo, 2004), participants who studied varied examples were subsequently less likely to describe the target concept (SWR) correctly, consistent with greater difficulty detecting commonalities among examples, but were more likely to refer to abstract structural features when describing it, consistent with greater ease of extracting underlying structure.
3. Both of the effects described in Assumption 2 depend on learners' structural sensitivity, that is, their ability to attend to and encode structural features. Prior knowledge influences

the effects of variation by increasing structural sensitivity (Chi et al., 1981; Chi & VanLehn, 2012).

$$P(l|s, v) = P(c|s, v) * P(e|c, s, v) \tag{1}$$

$$P(c|s, v) = \sigma\left(e^{-av} + sK, w, x\right) \tag{2}$$

$$P(e|c, s, v) = \sigma\left(\frac{sK}{e^{-av} + sK}, y, z\right) \tag{3}$$

A mathematical model reflecting the above assumptions is shown in Equations 1–3. Equation 1 states that the probability $P(l|s, v)$ of successfully learning a concept for a learner with structural sensitivity s studying examples whose level of superficial variation is v is equal to the product of the probability $P(c|s, v)$ of detecting commonalities among the examples, and the probability $P(e|c, s, v)$ of extracting their shared structure conditional on having detected commonalities. These two probabilities are multiplicatively combined because they are conjunctively required for structural learning to occur (Assumption 1).

Equation 2 states that the probability of detecting commonalities is a logistic function of the sum of their perceived superficial similarity e^{-av} and structural similarity sK, while Equation 3 states that the probability of extracting shared structure conditional on detection of commonalities is a logistic function of the ratio of structural to total similarity. The parameters of the logistic functions (center parameters w and y and scale parameters x and z) are free parameters of the model. Perceived superficial similarity e^{-av} is an exponentially decaying function of superficial variation v, reflecting a common assumption in psychological models (Shepard, 1987); the rate of exponential decay (a) is a free parameter of the model. Because of this relation, increasing v results in decreasing the value of Equation 2 and increasing the value of Equation 3. That is, making a set of examples more variable decreases the likelihood of noticing commonalities among them but increases the likelihood of identifying the structural basis underlying such commonalities if they are noticed (Assumption 2). Perceived structural similarity is the product of learners' structural sensitivity s (Assumption 3) and a constant value K, reflecting that all examples are assumed to have the same shared structure regardless of level of superficial variation. The constant K is a free parameter of the model.

Figure 5 shows the probabilities of successful learning predicted by the model as a function of superficial variation (v) for several different levels of structural sensitivity (s). Several important observations are immediately evident. First, for each level of structural sensitivity, there is an ideal level of variation, such that the probability of generalization peaks at the ideal level and is lower on either side of it. Second, the ideal level of variation is higher when structural sensitivity is higher. Third, higher structural sensitivity leads to weaker effects of variation, reflected by "flatter" curves.

Patterns consistent with each of the results discussed previously can be found in the model predictions. First, comparing structural sensitivity levels 0.2 and 0.3 for variation levels 0.15 and 0.25 (Figure 5, points A, B, C, and D) reveals a pattern consistent with the findings of the present study; that is, better learning outcomes for low variation for lower structural sensitivity and for high variation for higher structural sensitivity. Second, comparing structural sensitivity levels 0.2 and 0.6 for the same variation levels, 0.15 and 0.25 (Figure 5, points A, B, E, and F), reveals a pattern consistent with the findings of Guo and Pang (2011) and Rau et al., 2010; that is, better outcomes
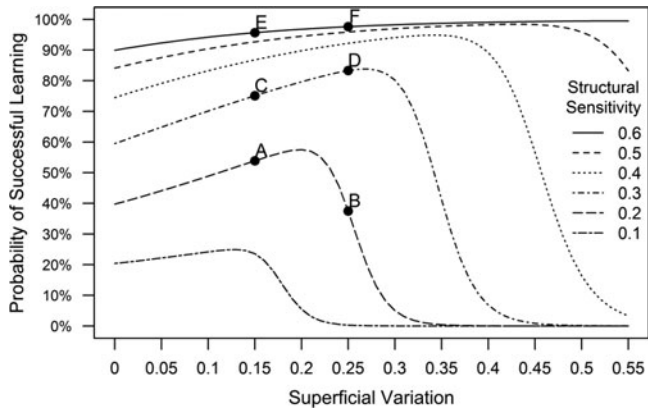
FIGURE 5 Probability of successful learning as a function of superficial variation and learners' structural sensitivity, as predicted by the model described in Equations 1–3, using the following settings for the model's free parameters: $a = 2$, $K = 1$, $w = .80$, $x = .02$, $y = .20$, $z = .08$.

for low variation at the lower level of structural sensitivity but virtually no effect of variation at the higher level. Finally, comparing structural sensitivity levels 0.3 and 0.6 again for the same variation levels, 0.15 and 0.3 (Figure 5, points C, D, E, and F), shows a pattern similar to those of Quilici and Mayer (1996) and Day et al., 2010; that is, better outcomes for high, rather than low, variation at the lower level of structural sensitivity and little effect of variation at the higher level.

We did not attempt to fit this model to the empirical results reviewed above due to the diversity of dependent measures, measures of prior knowledge, and manipulations of variation. However, the model demonstrates that these results can be organized within a common framework amenable to precise description and consistent with past research regarding effects of superficial similarity, variation, and learner knowledge. The model adds to existing accounts in three respects. First, it formally articulates two factors that trade off when determining the effects of variability for an individual learner: the likelihood of detecting commonalities among examples (benefitting from superficial similarity) and the likelihood of extracting a structural concept representation (benefitting from variation). Second, the model predicts that, for any given learner, optimal learning occurs at an ideal level of variation. Thus, for any level of learner knowledge, some levels of variation are too low and others too high, resulting in poorer learning than the ideal level. Third, the model predicts that the ideal level of variation increases with learner knowledge. Thus, for example, if increasing variation from a specific lower level to a specific higher level improves learning among less knowledgeable learners, the same increase in variation will also improve learning among more knowledgeable learners, although possibly to a lesser degree.

Related Theoretical Accounts

Although the above account connects most closely to research on concept learning and analogical transfer, our findings may also be viewed as an instance of the *expertise reversal effect*

(Kalyuga, 2007; Kalyuga, Ayres, Chandler, & Sweller, 2003), an effect typically interpreted using the framework of cognitive load theory (Sweller, 1988; van Merriënboer & Sweller, 2005). This effect occurs when instructional supports benefit less knowledgeable learners but do not benefit, or even harm, more knowledgeable ones (for a recent example see Fyfe, Rittle-Johnson, & DeCaro, 2012). Less knowledgeable learners are assumed to lack the knowledge structures needed to organize information from instruction, while instructional supports can facilitate such organization, reducing cognitive load. More knowledgeable learners already possess the requisite knowledge structures, while the need to integrate them with external supports can create extraneous cognitive load. In the present study, superficial similarity between examples may be viewed as a form of support, insofar as such similarity should facilitate drawing correspondences between examples. Thus, our observed interaction between similarity and prior knowledge may be viewed as an instance of the expertise reversal effect, and our findings as consistent with the larger literature regarding this effect.

Viewing similarity between examples as a type of support implies that the question of what level of similarity (or variation) promotes the best learning outcomes is a special case of a more general issue in instructional design—the *assistance dilemma* (Koedinger & Aleven, 2007; Koedinger, Pavlik, Mclaren, & Aleven, 2008). The larger question is how much assistance should be offered to learners during study. While broad qualitative responses such as "provide more assistance (i.e., greater similarity between examples) for less knowledgeable learners" may be useful as guidelines, an important long-term goal is the creation of quantitative models that can help to determine specific optimal levels of assistance for specific learning situations. Existing models have addressed this goal with respect to some dimensions of instructional design, such as temporal spacing of vocabulary review (Pavlik & Anderson, 2008) and balance of worked examples and problem solving (Koedinger et al., 2008). To our knowledge, no such model currently exists with respect to the issue of similarity and variation among examples. The model described in the previous section can serve as a first step in this direction.

Implications for Instruction

A practical implication of the present study is that learners must be prepared in order to profit from studying varied examples. One possible approach to such preparation is to begin with similar examples, and later switch to varied examples, on the grounds that initial study of similar examples could create the prior knowledge needed to benefit from varied examples (Day et al., 2010; Elio & Anderson, 1984; Kotovsky & Gentner, 1996). This approach was tested in another experiment not reported above. Participants in an experimental condition practiced with similar examples until accuracy reached a criterion, after which they practiced with varied examples for the remainder of training. However, this "adaptive variation" condition showed no advantage over training with either similar examples only or examples with a uniform (high) level of variation. This null result suggests that inductive learning from multiple similar examples may not be the most effective way to prepare learners to benefit from variation.

An alternative approach is to precede the study of concepts with explicit instruction and practice in encoding relevant structural features. The success of the experimental conditions of Experiment 2, especially the verbal pretraining condition, provides evidence for the effectiveness

of this approach. The use of explicit terminology referring to critical relations and roles (e.g., "for each of . . . one of . . . ," "alternatives," "selections") may have been critical to the success of the experimental pretraining (Gentner et al., 2011; Loewenstein & Gentner, 2005; Son et al., 2010, 2012). While well-developed terminology exists to refer to structural roles within mathematical formalisms—for example, "addend" and "augend," "term" and "factor," "domain" and "range"—the use of relational terminology to refer to structural relations and roles within specific, concrete situations has received less attention. Such terminology has the potential to highlight the underlying structure within situations and thus facilitate making connections to formalisms that represent such structure.

ACKNOWLEDGMENTS

FUNDING

SUPPLEMENTAL MATERIAL

Supplemental data for this article can be accessed on the publisher's website.

REFERENCES

Bassok, M. (1996). Using content to interpret structure: Effects on analogical transfer. *Current Directions in Psychological Science*, *5*(2), 54–58.

Bassok, M., Chase, V. M., & Martin, S. A. (1998). Adding apples and oranges: Alignment of semantic and formal knowledge. *Cognitive Psychology*, *35*(2), 99–134. http://dx.doi.org/10.1006/cogp.1998.0675

Bassok, M., & Holyoak, K. J. (1989). Interdomain transfer between isomorphic topics in algebra and physics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(1), 153–166. http://dx.doi.org/10.1037/h0090398

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

Belenky, D. M., & Schalk, L. (2014). The effects of idealized and grounded materials on learning, transfer, and interest: An organizing framework for categorizing external knowledge representations. *Educational Psychology Review*, *26*(1), 27–50. doi:http://dx.doi.org/10.1007/s10648-014-9251-9

Ben-Zeev, T., & Star, J. R. (2001). Spurious correlations in mathematical thinking. *Cognition and Instruction*, *19*(3), 253–275. http://dx.doi.org/10.1207/S1532690XCI1903_1

Bernardo, A. B. I. (2001). Analogical problem construction and transfer in mathematical problem solving. *Educational Psychology*, *21*(2), 137–150. doi: http://dx.doi.org/10.1080/01443410020043841

Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2012). Why interleaving enhances inductive learning: The roles of discrimination and retrieval. *Memory & Cognition*, *41*(3), 392–402. doi: http://dx.doi.org/10.3758/s13421-012-0272-7

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*, 417–44. doi:http://dx.doi.org/10.1146/annurev-psych-113011-143823

Blessing, S. B., & Ross, B. H. (1996). Content effects in problem categorization and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(3), 792–810.

Braithwaite, D. W., & Goldstone, R. L. (2012). Inducing mathematical concepts from specific examples: The role of schema-level variation. In N. Miyake, D. Peebles, & Ri. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the Cognitive Science Society* (pp. 138–143). Austin, TX: Cognitive Science Society.

Braithwaite, D. W., & Goldstone, R. L. (2013). Integrating formal and grounded representations in combinatorics learning. *Journal of Educational Psychology*, *105*(2), 666–682. doi: http://dx.doi.org/10.1037/a0032095

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. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17128610

Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(6), 1147–1156. doi:http://dx.doi.org/10.1037//0278-7393.15.6.1147

Chang, N. M. (2006). *Learning to discriminate and generalize through problem comparisons*. (Doctoral Dissertation). Carnegie Mellon University, Pittsburgh, PA. Retrieved from http://gse3.soe.berkeley.edu/faculty/NMing/docs/Chang2006_Dissertation.pdf

Chang, N. M., Koedinger, K. R., & Lovett, M. C. (2003). Learning spurious correlations instead of deeper relations. In R. Alterman & D. Kirsch (Eds.), *Proceedings of the 25th annual conference of the Cognitive Science Society* (pp. 228–233). Boston, MA: Cognitive Science Society. Retrieved from https://ncrve.berkeley.edu/faculty/NMing/docs/ChangKoedingerLovett2003_CogSci03.pdf

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. doi:http://dx.doi.org/10.1037/0278-7393.30.3.583

Cheng, P. W., Holyoak, K. J., Nisbett, R. E., & Oliver, L. M. (1986). Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology*, *18*(3), 293–328. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3742999

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13* (2), 145–182. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/0364021389900025

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. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S0364021381800298

Chi, M. T. H., & VanLehn, K. A. (2012). Seeing deep structure from the interactions of surface features. *Educational Psychologist*, *47*(3), 177–188. doi: http://dx.doi.org/10.1080/00461520.2012.695709

Corbalan, G., Kester, L., & van Merriënboer, J. J. G. (2009). Combining shared control with variability over surface features: Effects on transfer test performance and task involvement. *Computers in Human Behavior*, *25* (2), 290–298. doi:http://dx.doi.org/10.1016/j.chb.2008.12.009

Day, S. B., & Goldstone, R. L. (2012). The import of knowledge export: Connecting findings and theories of transfer of learning. *Educational Psychologist*, *47*(3), 153–176. doi: http://dx.doi.org/10.1080/00461520.2012.696438

Day, S. B., Goldstone, R. L., & Hills, T. (2010). *The effects of similarity and individual differences on comparison and transfer*. In Proceedings of the 32nd annual meeting of the cognitive science society (pp. 465–470). Portland, OR: Cognitive Science Society. Retrieved from http://csjarchive.cogsci.rpi.edu/proceedings/2010/papers/0078/paper0078.pdf

De Bock, D., Johan, D., van Dooren, W., Roelens, M., & Verschaffel, L. (2011). Abstract or concrete examples in learning mathematics? A replication and elaboration of Kaminski, Sloutsky, and Heckler's study. *Journal for Research in Mathematics Education*, *42*(2), 109–126.

Eiriksdottir, E., & Catrambone, R. (2011). Procedural instructions, principles, and examples: How to structure instructions for procedural tasks to enhance performance, learning, and transfer. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *53*(6), 749–770. doi: http://dx.doi.org/10.1177/0018720811419154

Elio, R., & Anderson, J. R. (1981). The effects of category generalizations and instance similarity on schema abstraction. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(6), 397–417. Retrieved from http://act-r.psy.cmu.edu/papers/183/journaleperimental.pdf

Elio, R., & Anderson, J. R. (1984). The effects of information order and learning mode on schema abstraction. *Memory & Cognition*, *12*(1), 20–30. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6708807

Fisher, K. J., Borchert, K., & Bassok, M. (2011). Following the standard form: Effects of equation format on algebraic modeling. *Memory & Cognition*, *39*(3), 502–515. http://dx.doi.org/10.3758/s13421-010-0031-6

Fyfe, E. R., McNeil, N. M., Son, J. Y., & Goldstone, R. L. (2014). Concreteness fading in mathematics and science instruction: A systematic review. *Educational Psychology Review*, *26*(1), 9–25. Retrieved from http://link.springer.com/article/10.1007/s10648-014-9249-3

Fyfe, E. R., Rittle-Johnson, B., & DeCaro, M. S. (2012). The effects of feedback during exploratory mathematics problem solving: Prior knowledge matters. *Journal of Educational Psychology*, *104*(4), 1094–1108.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*(2), 155–170. doi:http://dx.doi.org/10.1207/s15516709cog0702_3

Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science*, *34*(5), 752–775. doi:http://dx.doi.org/10.1111/j.1551-6709.2010.01114.x

Gentner, D., Anggoro, F. K., & Klibanoff, R. S. (2011). Structure mapping and relational language support children's learning of relational categories. *Child Development*, *82*(4), 1173–1188. doi:http://dx.doi.org/10.1111/j.1467-8624.2011.01599.x

Gentner, D., Loewenstein, J., & Hung, B. (2007). Comparison facilitates children's learning of names for parts. *Journal of Cognition and Development*, *8*(3), 285–307. doi: http://dx.doi.org/10.1080/15248370701446434

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*, *95*(2), 393–405. doi: http://dx.doi.org/10.1037/0022-0663.95.2.393

Gentner, D., Loewenstein, J., & Thompson, L. (2004). Analogical encoding: Facilitating knowledge transfer and integration. In *Proceedings of the 26th annual conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science*, *10*(3), 277–300. doi:http://dx.doi.org/10.1016/S0364-0213 (86)80019-2

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

Gick, M. L., & Paterson, K. (1992). Do contrasting examples facilitate schema acquisition and analogical transfer? *Canadian Journal of Psychology*, *46*(4), 539–550. Retrieved from http://psycnet.apa.org/journals/cep/46/4/539/

Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory & Cognition*, *24*(5), 608–28. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8870531

Goldstone, R. L., & Son, J. Y. (2005). The transfer of scientific principles using concrete and idealized simulations. *The Journal of the Learning Sciences*, *14*(1), 69–110.

Gómez, R. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431–437. Retrieved from http://pss.sagepub.com/content/13/5/431.short

Graham, S. A., Namy, L. L., Gentner, D., & Meagher, K. (2010). The role of comparison in preschoolers' novel object categorization. *Journal of Experimental Child Psychology*, *107*(3), 280–90. doi:http://dx.doi.org/10.1016/j.jecp.2010.04.017

Große, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning and Instruction*, *17*(6), 612–634. doi: http://dx.doi.org/10.1016/j.learninstruc.2007.09.008

Guo, J.-P., & Pang, M.-F. (2011). Learning a mathematical concept from comparing examples: The importance of variation and prior knowledge. *European Journal of Psychology of Education*, *26*(4), 495–525. doi:http://dx.doi.org/10.1007/s10212-011-0060-y

Guo, J.-P., Yang, L.-Y., & Ding, Y. (2013). Effects of example variability and prior knowledge in how students learn to solve equations. *European Journal of Psychology of Education*, *29*(1), 21–42. doi:http://dx.doi.org/10.1007/s10212-013-0185-2

Hammer, R., Bar-Hillel, A., Hertz, T., Weinshall, D., & Hochstein, S. (2008). Comparison processes in category learning: From theory to behavior. *Brain Research*, *1225*, 102–118. doi:http://dx.doi.org/10.1016/j.brainres.2008.04.079

Harp, S., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology*. Retrieved from http://psycnet.apa.org/journals/edu/90/3/414/

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*(4), 509–539. doi:http://dx.doi.org/10.1007/s10648-007-9054-3

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, *38*(1), 23–31. Retrieved from http://www.tandfonline.com/doi/abs/10.1207/S15326985EP3801_4

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2008). The advantage of abstract examples in learning math. *Science*, *320*(5875), 454–455.

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2013). The cost of concreteness: The effect of nonessential information on analogical transfer. *Journal of Experimental Psychology: Applied*, *19*(1), 14–29. doi:http://dx.doi.org/10.1037/a0031931

Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology*, *26*(1), 97–103. doi:http://dx.doi.org/10.1002/acp.1801

Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, *19*(3), 239–264. doi:http://dx.doi.org/10.1007/s10648-007-9049-0

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The Knowledge-Learning-Instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science*, *36*(5), 757–98. doi:http://dx.doi.org/10.1111/j.1551-6709.2012.01245.x

Koedinger, K. R., Pavlik, P., Mclaren, B. M., & Aleven, V. (2008). Is it better to give than to receive? The assistance dilemma as a fundamental unsolved problem in the cognitive science of learning and instruction. In *Proceedings of the 30th annual conference of the Cognitive Science Society* (pp. 2155–2160). New York, NY: Erlbaum.

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

Kuehne, S. E., Forbus, K. D., Gentner, D., & Quinn, B. (2000). SEQL: Category learning as progressive abstraction using structure mapping. In *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society* (pp. 770–775). Mahwah, NJ: Erlbaum. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.38.4757&rep=rep1&type=pdf

LeFevre, J., & Dixon, P. (1986). Do written instructions need examples? *Cognition and Instruction*, *3*(1), 1–30. Retrieved from http://www.tandfonline.com/doi/abs/10.1207/s1532690xci0301_1

Lobato, J. (2006). Alternative perspectives on the transfer of learning: History, issues, and challenges for future research. *The Journal of the Learning Sciences*, *15*(4), 431–449. doi:http://dx.doi.org/10.1207/s15327809jls1504

Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology*, *50*(4), 315–353. Retrieved from http://www.sciencedirect.com/science/article/pii/S0010028504000751

Markman, A. B., & Stilwell, C. H. (2001). Role-governed categories. *Journal of Experimental & Theoretical Artificial Intelligence*, *13*(4), 329–358. doi:http://dx.doi.org/10.1080/09528130110100252

Martin, S. A., & Bassok, M. (2005). Effects of semantic cues on mathematical modeling: Evidence from word-problem solving and equation construction tasks. *Memory & Cognition*, *33*(3), 471–478.

Mayer, R. E. (2005). Principles for managing essential processing in multimedia learning: Segmenting, pretraining, and modality principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 169–182). New York, NY: Cambridge University Press.

Mayer, R. E., & Griffith, E. (2008). Increased interestingness of extraneous details in a multimedia science presentation leads to decreased learning. *Journal of Experimental Psychology: Applied*, *14*(4), 329–339. Retrieved from http://psycnet.apa.org/journals/xap/14/4/329/

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: Lawrence Erlbaum Associates. Retrieved from http://psycnet.apa.org/psycinfo/1989-98035-006

Namy, L. L., & Clepper, L. E. (2010). The differing roles of comparison and contrast in children's categorization. *Journal of Experimental Child Psychology*, *107*(3), 291–305. doi: http://dx.doi.org/10.1016/j.jecp.2010.05.013

Nathan, M. J. (2012). Rethinking formalisms in formal education. *Educational Psychologist*, *47*(2), 125–148. doi:http://dx.doi.org/10.1080/00461520.2012.667063

Nathan, M. J., Kintsch, W., & Young, E. (1992). A theory of algebra-word-problem comprehension and its implications for the design of learning environments. *Cognition and Instruction*, *9*(4), 329–389.

Novick, L. R., & Holyoak, K. J. (1991). Mathematical problem solving by analogy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(3), 398–415. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1829473

Nuñes, T., Schliemann, A. D., & Carraher, D. W. (1993). *Street mathematics and school mathematics*. Cambridge, UK: Cambridge University Press.

Onnis, L., Monaghan, P., Christiansen, M. H., & Chater, N. (2004). Variability is the spice of learning, and a crucial ingredient for detecting and generalizing in nonadjacent dependencies. *Proceedings of the 26th annual conference of the Cognitive Science Society* (pp. 1047–1052). Mahwah, NJ: Erlbaum.

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, *12*(6), 237–41. doi:http://dx.doi.org/10.1016/j.tics.2008.02.014

Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology*, *86*(1), 122–133.

Pavlik, P. I., & Anderson, J. R. (2008). Using a model to compute the optimal schedule of practice. *Journal of Experimental Psychology: Applied*, *14*(2), 101–117. http://dx.doi.org/10.1037/1076-898X.14.2.101

Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn locally, think globally: Exemplar variability supports higher-order generalization and word learning. *Psychological Science*, *21*(12), 1894–1902. doi:http://dx.doi.org/10.1177/0956797610389189

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*(3, pt 1), 353–363. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14959500

Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology*, *88*(1), 144–161. doi: http://dx.doi.org/10.1037//0022-0663.88.1.144

Quilici, J. L., & Mayer, R. E. (2002). Teaching students to recognize structural similarities between statistics word problems. *Applied Cognitive Psychology*, *16*(3), 325–342. doi:http://dx.doi.org/10.1002/acp.796

Rau, M. A., Aleven, V., & Rummel, N. (2010). Blocked versus interleaved practice with multiple representations in an intelligent tutoring system for fractions. In V. Aleven, J. Kay, & J. Mostow (Eds.), *Intelligent tutoring systems* (pp. 413–422). Berlin/Heidelberg, Germany: Springer.

Rawson, K. A., Thomas, R. C., & Jacoby, L. L. (2015). The power of examples: Illustrative examples enhance conceptual learning of declarative concepts. *Educational Psychology Review*, *27*(3), 483–504. http://dx.doi.org/10.1007/s10648-014-9273-3

Reed, S. K. (1989). Constraints on the abstraction of solutions. *Journal of Educational Psychology*, *81*(4), 532–540. doi:http://dx.doi.org/10.1037//0022-0663.81.4.532

Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from worked-out examples: The effects of example variability and elicited self-explanations. *Contemporary Educational Psychology*, *23*(1), 90–108. doi:http://dx.doi.org/10.1006/ceps.1997.0959

Rittle-Johnson, B., Star, J. R., & Durkin, K. (2009). The importance of prior knowledge when comparing examples: Influences on conceptual and procedural knowledge of equation solving. *Journal of Educational Psychology*, *101*(4), 836–852.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573–605.

Ross, B. H. (1984). Remindings and their effects in learning a cognitive skill. *Cognitive Psychology*, *16*(3), 371–416. doi:http://dx.doi.org/10.1016/0010-0285 (84)90014-8

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. doi:http://dx.doi.org/10.1037/0278-7393.13.4.629

Ross, B. H. (1989). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(3), 456–468.

Ross, B. H., & Kilbane, M. C. (1997). Effects of principle explanation and superficial similarity on analogical mapping in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(2), 427–440. doi:http://dx.doi.org/10.1037/0278-7393.23.2.427

Rost, G. C., & McMurray, B. (2010). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*(2), 339–349. doi:http://dx.doi.org/10.1111/j.1467-7687.2008.00786.x.Speaker

Scheiter, K., Gerjets, P., & Schuh, J. (2010). The acquisition of problem-solving skills in mathematics: How animations can aid understanding of structural problem features and solution procedures. *Instructional Science*, *38*(5), 487–502. doi:http://dx.doi.org/10.1007/s

Schwartz, D. L., & Bransford, J. (1998). A time for telling. *Cognition and Instruction*, *16*(4), 475–522. Retrieved from http://www.tandfonline.com/doi/abs/10.1207/s1532690xci1604_4

Schwartz, D. L., Bransford, J. D., & Sears, D. (2005). Efficiency and innovation in transfer. In J. P. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 1–51). Greenwich, CT: Information Age.

Schwartz, D. L., Chase, C. C., & Bransford, J. D. (2012). Resisting overzealous transfer: Coordinating previously successful routines with needs for new learning. *Educational Psychologist*, *47*(3), 204–214.

Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology*, *103*(4), 759–775. doi:http://dx.doi.org/10.1037/a0025140

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.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323. Retrieved from http://www.sciencemag.org/cgi/content/abstract/sci;237/4820/1317

Skorstad, J., Gentner, D., & Medin, D. L. (1988). Abstraction processes during concept learning: A structural view. In *Proceedings of the Tenth Annual Conference of the Cognitive Science Society* (pp. 419–425). Hillsdale, NJ: Erlbaum. Retrieved from http://www.psych.northwestern.edu/psych/people/faculty/gentner/papers/SkorstadGentnerMedin88.pdf

Son, J. Y., Doumas, L. A. A., & Goldstone, R. L. (2010). When do words promote analogical transfer? *Journal of Problem Solving*, *3*(1), 52–92.

Son, J. Y., & Goldstone, R. L. (2009). Contextualization in perspective. *Cognition and Instruction*, *27*(1), 51–89. doi:http://dx.doi.org/10.1080/07370000802584539

Son, J. Y., Smith, L. B., Goldstone, R. L., & Leslie, M. (2012). The importance of being interpreted: Grounded words and children's relational reasoning. *Frontiers in Psychology*, *3*. doi: http://dx.doi.org/10.3389/fpsyg.2012.00045

Stern, E., Aprea, C., & Ebner, H. G. (2003). Improving cross-content transfer in text processing by means of active graphical representation. *Learning and Instruction*, *13*(2), 191–203. doi: http://dx.doi.org/10.1016/S0959-4752(02)00020-8

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. Retrieved from http://onlinelibrary.wiley.com/doi/10.1207/s15516709cog1202_4/abstract

Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology*, *24*(6), 837–848. doi:http://dx.doi.org/10.1002/acp

Van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, *17*(2), 147–177. doi:http://dx.doi.org/10.1007/s10648-005-3951-0

Vanderstoep, S. W., & Seifert, C. M. (1993). Learning "how" versus learning "when": Improving transfer of problem-solving principles. *The Journal of the Learning Sciences*, *3*(1), 93–111.

Wagner, J. F. (2010). Transfer in pieces. *Cognition and Instruction*, *24*(1), 1–71. http://dx.doi.org/10.1207/s1532690xci2401

Woodward, J., Beckmann, S., Driscoll, M., Franke, M., Herzig, P., Jitendra, A., . . . Ogbuehi, P. (2012). Improving mathematical problem solving in grades 4 through *8*. In *IES practice guide. NCEE 2012-4055*. Washington, DC: National Center for Education Evaluation and Assistance, Institute of Education Sciences, U.S. Department of Education.