Jee, B., Uttal, D., Gentner, D., Manduca, C., Shipley, T., & Sageman, B. (2013). Finding faults: analogical comparison supports spatial concept learning in geoscience. *Cognitive Processing*, 14(2), 175-187.

Cogn Process (2013) 14:175–187
DOI 10.1007/s10339-013-0551-7

RESEARCH REPORT

Finding faults: analogical comparison supports spatial concept learning in geoscience

Benjamin D. Jee · David H. Uttal · Dedre Gentner · Cathy Manduca · Thomas F. Shipley · Bradley Sageman

Received: 30 November 2012 / Accepted: 5 February 2013 / Published online: 24 February 2013

Abstract A central issue in education is how to support the spatial thinking involved in learning science, technology, engineering, and mathematics (STEM). We investigated whether and how the cognitive process of analogical comparison supports learning of a basic spatial concept in geoscience, *fault*. Because of the high variability in the appearance of faults, it may be difficult for students to learn the category-relevant spatial structure. There is abundant evidence that comparing analogous examples can help students gain insight into important category-defining features (Gentner in Cogn Sci 34(5):752–775, 2010). Further, comparing high-similarity pairs can be especially effective at revealing key differences (Sagi et al. 2012). Across three experiments, we tested whether comparison of visually similar contrasting examples would help students learn the fault concept. Our main findings were that participants performed better at identifying faults when they (1) compared contrasting (fault/no fault) cases versus viewing each case separately (Experiment 1), (2) compared similar as opposed to dissimilar contrasting cases early in learning (Experiment 2), and (3) viewed a contrasting pair of schematic block diagrams as opposed to a single block diagram of a fault as part of an instructional text (Experiment 3). These results suggest that comparison of visually similar contrasting cases helped distinguish category-relevant from category-irrelevant features for participants. When such comparisons occurred early in learning, participants were more likely to form an accurate conceptual representation. Thus, analogical comparison of images may provide one powerful way to enhance spatial learning in geoscience and other STEM disciplines.

Keywords Analogy · Similarity · Science learning · Conceptual representation

Introduction

Increasing participation and performance in the STEM disciplines—science, technology, engineering, and math—is a major educational initiative in the United States. President Obama stated in a 2010 speech, "…Our nation's success depends on strengthening America's role as the world's engine of discovery and innovation, and that leadership tomorrow depends on how we educate our students today—especially in science, technology, engineering, and math" ("Changing the Equation in STEM Education,"

This article is part of the special issue on ''Spatial Learning and Reasoning Processes'', guest-edited by Thomas F. Shipley, Dedre Gentner and Nora S. Newcombe. Handling editor of this manuscript: Nora S. Newcombe.

B. D. Jee (✉)
Department of Education, College of the Holy Cross,
1 College Street, Worcester, MA 01610, USA
e-mail: bjee@holycross.edu

D. H. Uttal · D. Gentner
Department of Psychology, Northwestern University,
Evanston, IL, USA

C. Manduca
Science Education Resource Center, Carleton College,
Northfield, MN, USA

T. F. Shipley
Department of Psychology, Temple University,
Philadelphia, PA, USA

B. Sageman
Department of Earth and Planetary Sciences,
Northwestern University, Evanston, IL, USA

2010). This educational issue, which will likely play a key role in shaping the future economy of the U.S. and other industrialized nations, relates to deeper issues in science learning, including the cognitive processes that contribute to scientific understanding.

A growing body of research has established that *spatial thinking* is vital to student success in STEM disciplines (Baenninger and Newcombe 1989; Downs and DeSouza 2006; Gobert and Clement 1999; Hegarty et al. 2009; Kali and Orion 1996; Liben et al. 2011; Sorby 2009; Uttal et al. 2012). Students in many STEM disciplines are required to interpret or imagine the spatial layout of objects and events, including things that are perceptually inaccessible because of their extremely small or large scale. In this article, we explore a way to potentially enhance students' learning of basic spatial concepts in geoscience through the cognitive process of analogical comparison.

Geoscience: a domain for exploring spatial learning in science

Geoscience deals with the dynamics and physical history of Earth, the substances of which it is composed, and the physical, chemical, and biological changes that Earth has undergone or is undergoing. As Kastens and Ishikawa (2006) discuss, geoscience learning involves a range of cognitive processes. One fundamental process is the acquisition of conceptual knowledge that enables the geoscientist to classify geological objects. There are many different systems of classification depending on the nature of the objects, for example, paleontologists use a system to classify fossils that is separate from the system that mineralogists use to classify mineral samples. These classification systems are useful because they enable geoscientists to make useful predictions, such as the age, composition, and origins of an object (see Solomon et al. (1999) for an extended discussion of the functions of conceptual knowledge).

The focus of the present research is the subfield of *structural geology*. Structural geologists use the structure of present-day rock formations—signs that rocks have been shifted, deformed, or otherwise reconfigured—to infer how forces operating underneath Earth's surface have changed an area over time (Marshak 2005). This process involves classifying geologic structures on Earth's surface, such as faults, folds, and fractures (Jee et al. 2010; Kastens et al. 2009; Libarkin and Brick 2002; Sibley 2009). Many of these structures are inherently spatial (Liben et al. 2011). A *fault*, for example, occurs when extensional or compressional forces produce a fracture in rock and the block of rock above the fracture is displaced relative to the block of rock below (e.g., Marshak 2005). Thus, a fault is defined by a spatial pattern: blocks of rock that are spatially separated along a fracture.

Learning to classify geological structures, such as faults, can be challenging for several reasons. Objects on Earth's surface, such as vegetation, can obscure a structure. Also, a structure's appearance can change over time as a result of events, such as erosion, that alter its features. These factors can make it difficult to perceptually analyze an *outcrop*—the parts of Earth's crust that are visible at the surface—into its relevant parts. Even when the parts are recognized, classification can be difficult because the rocks in different outcrops can vary in color, stratification, orientation, and other visual features. Thus, different instances of the same geological structure can vary greatly in their visual appearance. Consider the three images in Fig. 1. Each image displays a fault. In each case, there is a displacement in rock types on each side of the fracture, indicating that the blocks have moved relative to each other. Aside from this common fact of relative displacement, the examples are highly dissimilar.

As within-category variability increases, category learning becomes more difficult (e.g., Posner et al. 1967). Yet, if learners' attention is directed toward category-relevant features—in this case, the relative spatial locations of the blocks of rock on each side of the fracture—learning can be greatly facilitated (Biederman and Shiffrir 1987). In this research, we explored one potentially powerful process through which the relevant spatial structure of faults could be highlighted: analogical comparison.

Fig. 1 Examples of geological faults

Analogical comparison and spatial concept learning

Analogy is a type of similarity in which two examples share the same system of *relations*. The concrete features of the examples may be similar or dissimilar; analogical comparisons are concerned with whether the features relate to one another in the same way in each case (Gentner 1983). The comparison process acts to achieve a structural alignment, which places the elements of the two examples in correspondence based on their roles in the common relational system. For example, in the well-known analogy between the atom and the solar system, the sun and the nucleus are placed in correspondence because they are both at the center of their respective systems, objects revolve around them, etc. The process of determining the corresponding elements of the systems is referred to as *analogical mapping*. There is considerable evidence that adults prefer analogical mappings that are based on common relations as opposed to superficial similarities (Gentner 1983, 1988, 2010; Goldstone et al. 1991; Holyoak and Koh 1987; Jones and Love 2007; Keane 1996; Krawczyk et al. 2004; Markman and Gentner 1993).

Analogies are often used to relate a novel example to one that is already well understood; this is called a *projective* analogy. In projective analogy, the learner is meant to liken a new example to a familiar (or more well understood) example and to project inferences from the familiar to the novel example. Structural alignment can also be performed when both cases are unfamiliar. This is called a *mutual alignment* analogy (Gentner 2005; Jee et al. 2010). One use of mutual alignment analogy is to help the student learn the abstract commonalities between two cases (Gentner and Namy 1999; Kurtz et al. 2001). A student is more likely to represent the category-relevant features of an item when they first compare two examples, as opposed to processing a single example (Gentner and Namy 1999; Gick and Holyoak 1983) or two examples separately (Gentner et al. 2003; Orton et al. 2012).

In addition to highlighting structural commonalities, mutual alignment can facilitate the noticing of differences. When two examples are aligned, differences that are connected in the same way to the common system (alignable differences) stand out (Gentner and Markman 1994; Markman and Gentner 1993, 1996). Perhaps surprisingly, evidence shows that although it is easier to respond *that* two examples are different when they have many differences, it is easier to notice a particular difference when the examples are highly *similar* to one another. For example, Sagi et al. (2012) gave participants pairs of images and asked them to type a difference between them as quickly as possible. People were faster to type a difference for highly similar (and highly alignable) pairs than for dissimilar (and difficult to align) pairs (see Fig. 2).

In the current work, we applied research findings concerning the effects of mutual alignment to the problem of learning spatial concepts in structural geology, such as the fault. We hypothesize that viewing highly similar (and highly alignable) pairs that contrast in key features could highlight the category-relevant features of a fault—the displacement in spatial locations of the blocks of rock on each side of a fracture. If so, learning could be facilitated through comparing an image of a fault to a visually similar image that does not contain a fault. We tested this hypothesis across three experiments involving different applications of mutual alignment contrast.

Experiment 1

This experiment tested (a) whether viewing pairs of contrasting images—such that one contains a fault and the other does not—helps novice participants learn to classify faults and (b) whether high-similarity pairs are more effective for this purpose than low-similarity pairs. Participants were presented with the same set of images either separately (Separate condition) or in pairs (Paired condition) and had to identify the images that contained faults. Some of the pairs were highly alignable (high similarity) and others were difficult to align (low similarity). If comparison of contrasting images facilitates fault classification, then participants in the Paired condition should perform better than the Separate condition. Furthermore, performance of the Paired condition should be especially high for the similar pairs, because fault-related features should be more salient when the images are similar.

Methods

Participants

The participants were 56 undergraduates from Northwestern University (33 females and 23 males) with no prior course experience in geoscience. Participants were given course credit to participate in this experiment as part of an introductory psychology class.

Materials

A brief instructional text (see the "Appendix") was created with input from a geoscience professor. The text provided some basic background information about structural geology and contained a conceptual definition of a geological fault, information about cues to identify faults (i.e., displacement of the blocks of rock along a fracture), and a block diagram of a fault, similar to those commonly used in textbooks.

The images for the classification tasks were created using 18 digital photographs of geological faults. From

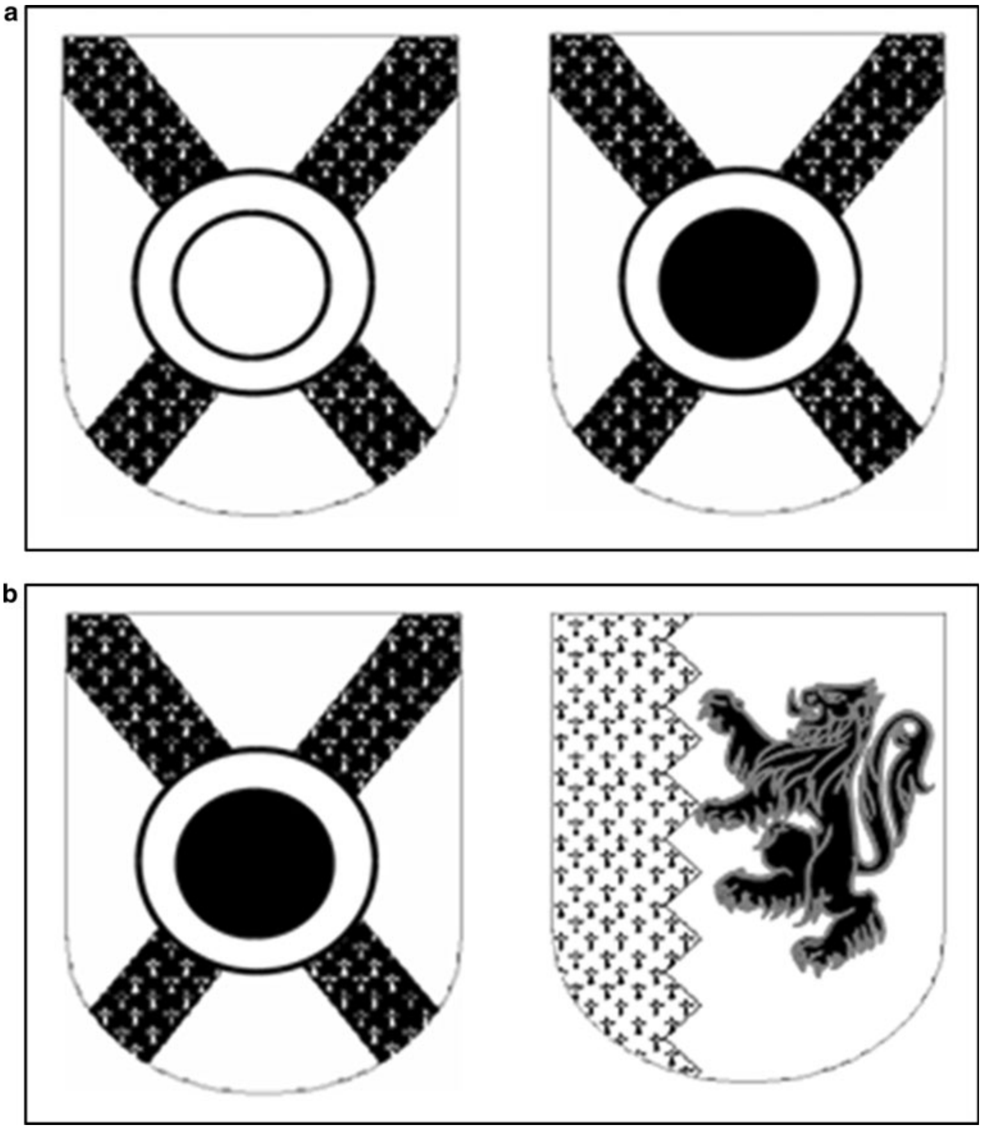


Fig. 2 Examples of **a** high-similarity and **b** low-similarity pairs from Sagi et al. (2012). Participants were faster to type a difference for the high-similarity pairs

each of these large photographs, we created two smaller images by cropping in on an area that showed the fault and an area that did not. This yielded two similar images (see Fig. 3, top row); the key difference between them was that one contained a fault and the other did not. The classification task consisted of 36 images in total—18 pairs of highly alignable images, half with faults and half without.

Procedure

All participants were tested individually in a quiet room. The instructional text and classification tasks were presented on a computer running E-Prime software (Psychology Software Tools, pstnet.com). The participant was asked to read the instructions carefully and could ask questions before the classification task began. After reading through the instructional text, participants were randomly assigned to either the Separate or Paired classification condition. (The proportion of men and women in each condition was approximately equal).

In the Separate classification task, participants received the 36 images (18 fault, 18 no fault) individually. The participant classified each image by pressing a button on the keyboard, either “1” for *fault* or “3” for *not a fault*. These category labels were presented at the bottom of the computer screen, one next to the other, along with a reminder about the designated keys. Each participant received all 36 images in random order. There was no feedback following the participant’s decision.

In the Paired classification task, participants received the 36 images in 18 pairs. In each pair, one of the images contained a fault and the other did not. In each trial, the two images were presented on opposite sides of the screen, one on the left and one on the right. Across the 18 trials, the fault image was located on the left and right equally often. The participant was asked to identify the image in each pair that

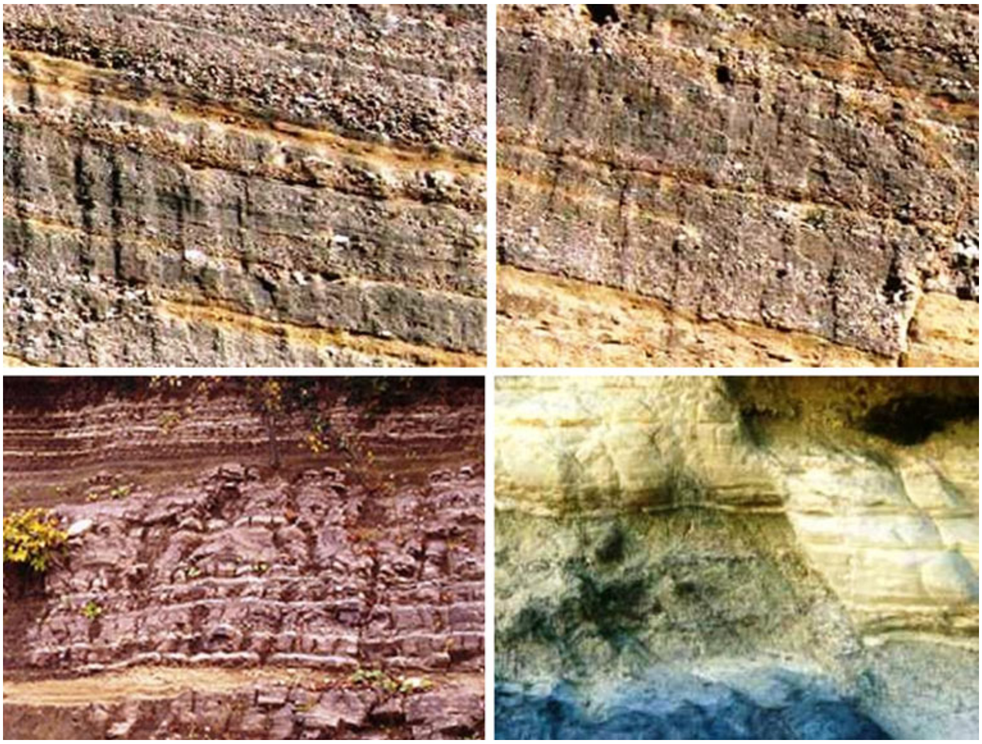


Fig. 3 *Top row:* Example of a high-similarity pair from the classification task. The fault is located in the lower right corner of the image on the right (notice how the lighter layer at the bottom of the image is higher on the right side of the fault than the left—the rocks have been displaced). *Bottom row:* Example of a low-similarity pair from the classification task. The fault is located in the center of the image on the right and slopes downward to the right

contained a fault by pressing a button on the keyboard, either “1” for *the image on the left* or “3” for *the image on the right*. These category labels were presented at the bottom of the computer screen, one next to the other, along with a reminder about the designated keys. Across the trials, half of the pairs were high in similarity (i.e., two images cropped from the same larger image) and half were low in similarity (two images cropped from different larger images; see Fig. 3, bottom row). Each participant received the same pairs in a different random order over the course of the task. Within the dissimilar pairs, the pairing of fault and no-fault images was also randomly determined. There was no feedback following the participant’s decision. The main dependent measure was accuracy at identifying faults and non-faults.

To verify that the participants had no prior geoscience course experience (at the high school or university level), they also completed a short survey that asked how many geology/Earth sciences courses they had taken in high school and in university, which course(s) they had taken (if any), and where.

Results

A significance level of .05 was employed for all statistical analyses. Performance on the classification task was quantified using d-prime. We also conducted the analyses including gender as an independent variable; however, there was no significant main effect or interaction involving gender. We therefore present the analyses without gender included. (Gender was not a significant factor in Experiments 2 and 3 either, and thus, we do not report gender results throughout the paper).

The main question concerns the performance of participants who could compare images during classification (Paired condition) versus those who could not (Separate condition). On average, participants in the Paired condition ($M = 1.59$, SD $= 0.82$) performed significantly better than participants in the Separate condition ($M = 1.06$, SD $= 0.73$), $t(54) = 2.56$, $p < .05$, $d = 0.14$–1.22 (95 % CI). For the paired condition, we also examined performance on the similar and dissimilar pairs that were distributed throughout the classification task. On average, participants in the Paired condition performed better when the pairs were similar ($M = 1.67$, SD $= 0.69$) than when they were dissimilar ($M = 1.27$, SD $= 0.90$), $t(27) = 3.07$, $p < .05$, $d = 0.01$–1.05 (95 % CI).

There was a marginally significant difference in response time between the conditions. The Separate condition responded faster on average ($M = 4.38$ s, SD $= 2.28$ s) than the Paired condition ($M = 5.81$ s, SD $= 3.14$ s), $t(54) = 1.85$, $p = .07$, $d = -0.01$ to 1.05 (95 % CI). This does not imply, however, that participants in the Paired condition spent more time processing each image. If we consider response time in terms of time per image, the Paired condition was actually faster (spending under 3 s per image).

Discussion

The findings from Experiment 1 support the hypothesis that comparing alignable contrasting images facilitates fault classification. The results also suggest that similar pairs were more effective at highlighting the faults. When the participant could compare two highly alignable images that differed specifically in whether they contained a fault, they were most accurate at detecting the image that contained the fault.

Although we have argued for the beneficial effects of structural alignment processes, we note that there is another potential contributor to the high performance of participants in the Paired condition. The d-prime statistic provides an index of sensitivity while adjusting for bias. The Paired condition might have required *less* sensitivity to the distinctions between faults and no faults, because if one member of a pair was easy to classify then the participant would be likely to respond correctly, even if the other member of the pair was difficult to classify. In the Separate condition, participants had to respond to each item separately, so they lacked any opportunity to “piggyback” hard items onto easier ones. Thus, the Separate condition may have required greater sensitivity to the defining features of faults. This difference is not related to analogical processing, but remains a possible factor in the results. Nevertheless, the fact that (within the Paired condition) the similar pairs were classified more accurately than the less similar pairs strongly suggests that alignment was an important factor in classification.

If similar pairs are easier to align (Gentner and Medina 1998), then receiving highly similar contrasting pairs before more dissimilar pairs could enhance learning compared to other sequences. These easier alignments could support the formation of an accurate conceptual representation, which could then be applied to more distant future examples—a sequence referred to as *progressive alignment* (Gentner and Medina 1998; Kotovsky and Gentner 1996; Thompson and Opfer 2010). Indeed, when children are initially shown highly similar images of objects that differ in a single part, they are more likely to learn the part than when they are shown dissimilar cases early in learning (Gentner et al. 2007, 2011).

In addition to the sequence of instruction, another question is how learning is affected by background experience in the domain of geoscience. Two very different outcomes seem plausible. On the one hand, it could be that participants with prior geoscience experience may be sufficiently familiar with the concept of fault that they will perform well regardless of whether they receive images in the progressive alignment sequence. In that case, we expect to see a greater effect of this manipulation for novices than for those with geoscience experience. However, the opposite prediction is also plausible; it could be that those with geoscience experience will be better able to take advantage of receiving alignable pairs. Research on expertise finds that prior knowledge supports the perception and identification of meaningful patterns (Chase and Simon 1973; Lesgold et al. 1988; Uttal and Cohen 2012). Thus, participants with experience in geoscience could possess domain-specific knowledge that facilitates the analysis of images into geologically relevant features and guides the selection of which differences to focus on in the comparison process. They might therefore benefit more than the novices from receiving the progressive alignment manipulation.

Experiment 2

Experiment 2 examined whether a progressive alignment sequence—viewing highly alignable pairs of contrasting images early in learning, followed by less readily alignable pairs—would help learners acquire the fault concept. We also explored whether background geoscience experience facilitates fault classification in general, for a particular sequence of instruction, or not at all. Because this experiment involved students with prior geoscience experience, we attempted to reduce the possibility of ceiling effects by removing some of the details from the instructions. The instructional text for Experiment 2 provided only a conceptual definition of a fault; it did not include the figure and instructions about cues for classification that were present in Experiment 1. These changes were intended to make the classification task more challenging for students who had prior geoscience knowledge.

Methods

Participants

The participants were 48 undergraduates from Northwestern University who were given course credit to participate in this experiment as part of an introductory psychology course. There were 24 participants with no prior geoscience experience (5 females and 19 males) and 24 who had taken at least one geoscience course at the high school or college level (9 female and 15 males). We recruited students with geoscience experience by surveying the potential subject pool at the beginning of the term.

Materials

The instructional text from Experiment 1 was modified by removing the block diagram of a fault and the information about using cues for displacement to identify faults. The background information concerning structural geology and the conceptual definition of geological fault remained the same. The same 36 images from Experiment 1 were used for the classification task.

Procedure

All participants were tested individually in a quiet room. The instructional text and classification tasks were presented on a computer running E-Prime software (Psychology Software Tools, pstnet.com). The participant was asked to read the instructions carefully and was allowed to ask questions before the classification task began. After reading through the instructional text, participants were randomly assigned to one of the two classification tasks. (The proportion of men and women in each condition was approximately equal).

The procedures for the classification tasks were similar to the Paired classification task from Experiment 1, except that the image pairs were ordered in two blocks of trials. In the Similar First classification task, participants received 9 similar pairs of images in the first block and 9 dissimilar pairs in the second. The Similar First classification task represents the progressive alignment sequence. In the Dissimilar First classification task, participants received 9 dissimilar pairs in the first block and 9 similar pairs in the second. Two versions of each task were created such that every image appeared in a similar and dissimilar pair equally often within conditions. However, each image was seen only once by each participant. The order of the pairs was randomized within each block of trials for each participant. There was no feedback during the task.

To verify their prior geoscience course experience, all participants completed the same brief survey of geology/Earth sciences experience that was used in Experiment 1.

Results

The results were analyzed using a 2 (Condition: Similar First vs. Dissimilar First) × 2 (Geoscience Experience: Some Experience vs. No Experience) × 2 (Pair Type: Similar vs. Dissimilar) mixed factorial ANOVA, with Pair Type as a within-subjects variable. The analysis revealed a main effect of Condition; the Similar First condition ($M =$ 1.17, SD = 0.84) performed significantly better than the Dissimilar First condition ($M = 0.80$, SD = 0.65), $F(1,44) = 4.64$, $p < .05$, $\eta_p^2 = .10$. There was an effect of Geoscience Experience; not surprisingly, students with geoscience experience ($M = 1.19$, SD = 0.71) performed better than those without ($M = 0.79$, SD = 0.65), $F(1,44) = 5.41$, $p < .05$, $\eta_p^2 = .11$. There was also an effect of Pair Type: as predicted, performance was higher on the Similar pairs ($M = 1.19$, SD = 0.78) than on the Dissimilar pairs ($M = 0.78$, SD = 0.85), $F(1,44) = 10.00$, $p < .05$, $\eta_p^2 = .19$. There was also a significant interaction between Condition and Geoscience Experience, $F(1,44) = 6.79$, $p < .05$, $\eta_p^2 = .13$. In the Similar First condition, participants with geoscience experience ($M = 1.61$, SD = 0.61) performed significantly better than those without experience ($M = 0.78$, SD = 0.74), $t(22) = 2.30$, $p < .05$, $d = 0.35$–2.10 (95 % CI). However, in the Dissimilar First condition, the two groups did not differ significantly ($M = 0.77$, SD = 0.67 for participants with geoscience experience and $M = 0.82$, SD = 0.65 for those without experience, $t(22) = 0.56$, *ns*, $d = -0.82$ to 0.67 (95 % CI)). These results are displayed in Fig. 4.

Both groups performed better on the similar pairs than on the dissimilar pairs. The Similar First condition performed better on the similar pairs they received in block 1 ($M = 1.40$, SD = 0.77) compared to the dissimilar pairs they received in block 2 ($M = 0.95$, SD = 0.91). The Dissimilar First condition showed the same pattern, with higher performance on the similar pairs from block 2 ($M = 0.97$, SD = 0.75) than on the dissimilar pairs from block 1 ($M = 0.62$, SD = 0.77). Strikingly, performance on the similar pairs was higher in the Similar First condition than in the Dissimilar First condition, despite the fact that the Dissimilar First participants received these pairs after having some experience with the task (with dissimilar items). This suggests that receiving Dissimilar pairs first may have led participants to focus on the wrong aspects of the scenes and diminished their ability to profit from the Similar pairs. In addition, although the Similar First participants did not maintain their high level of performance

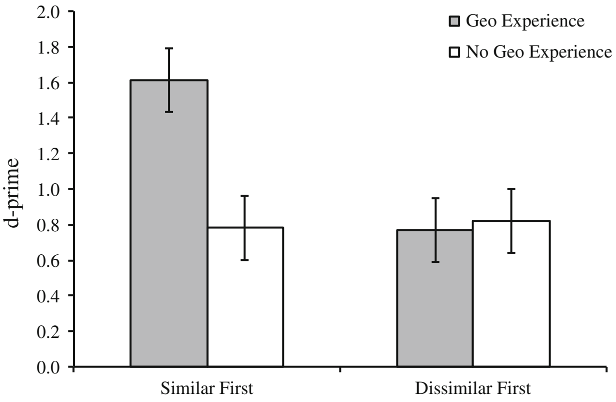


Fig. 4 Classification performance for participants with and without geoscience experience in the Similar First and Dissimilar First conditions (±SE)

from block 1 to 2, they also outperformed the Dissimilar First condition on the dissimilar pairs.

We also conducted a 2 (Condition) × 2 (Geoscience Experience) × 2 (Pair Type) mixed factorial ANOVA with response time as the dependent variable. There were no significant main effects or interactions, $F\text{s} < 1$, $p\text{s} > .35$.

Discussion

The results of Experiment 2 provide evidence that, as predicted, the progressive alignment sequence of contrasting images led to better overall fault classification performance. Viewing similar pairs early in learning not only facilitated performance on these items but may have also elevated performance on the dissimilar items. Participants who received the dissimilar items first performed worse overall—even on the similar pairs, which they received after having had a block of practice (with dissimilar pairs).

These findings could be attributable to the conceptual representations that participants formed in the different conditions. The alignment of similar contrasting images early in learning could have highlighted the features relevant to the fault category as well as the features that are irrelevant to category membership. The participants in the Similar First condition could then apply this abstract representation acquired early in learning to the more challenging dissimilar pairs. Participants in the Dissimilar First condition, however, may have formed an incomplete or inaccurate conceptual representation of a fault. Because the less alignable pairs did not highlight the category-relevant features, the Dissimilar First participants could have encoded irrelevant information into their representation. This could explain why they performed relatively poorly on the similar pairs that they received later in the task. In the absence of feedback, the participants would not know whether their representation was accurate or flawed.

The advantage of the progressive alignment condition (the Similar First condition) was greater for participants with geoscience experience than for inexperienced participants.[1] It is possible the novices were too lacking in the relevant geoscience knowledge to profit from even the easy matches. The lack of details in the instructional text compared to Experiment 1 could have also made it more difficult for the inexperienced participants to identify the geologically relevant features in the images.

In the next experiment, we examined concept learning using images that should support the analysis of geologically relevant features—namely, block diagrams. Block diagrams are 3D renderings that are commonly used in geoscience textbooks to display the spatial structure of faults and other formations (Jee et al. 2010). (Indeed, we included a single block diagram of a fault in the instructional text read by all participants in Experiment 1, but this factor was not manipulated). Our hypothesis is that, just as with real-world images, learning from a block diagram will be facilitated when participants can contrast a diagram of a fault with a similar no-fault diagram. This hypothesis was tested in Experiment 3.

In addition to using a simpler comparison manipulation, Experiment 3 differed from the prior studies in that we asked participants to indicate *where* the fault was located. This allows us an additional check on the accuracy of their understanding.

Experiment 3

In Experiment 3, we manipulated analogical comparison in the learning materials prior to the classification test, much as in textbook-based instruction. Participants read an instructional text that contained either a contrasting pair of 3D block diagrams, one with a fault and one without, or a single block diagram of a fault. Learning was tested in a classification task in which images were presented separately. The participant had to decide whether the image contained a fault and, if so, where the fault was located. If comparison of alignable contrasting block diagrams helps novice participants form an accurate conceptual representation of a fault, then they should perform better on the classification test than those who receive only a single fault diagram.

Method

Participants

The participants were 24 undergraduates from Northwestern University (12 females and 12 males) with no prior course experience in geoscience at the high school or university level. Participants were given course credit to participate in this experiment as part of an introductory psychology class.

Materials

The text from Experiment 2 was modified to create two versions of the instructions. All participants received some background information about structural geology and a conceptual definition of a geological fault. In addition, the instructions contained either a single block diagram of a

---

[1] The findings from Experiment 2 highlight the fact that the effects of comparison depend not only on the materials but also on the learner's prior knowledge. In the course of this research, we carried out several pilot studies that showed no effect—the distinction was either too subtle or too obvious given the participants' level of prior knowledge.

fault (Single Diagram condition) or a block diagram of a fault and a visually similar block diagram of a fracture with no fault (Paired Diagrams condition). The images from the instructions are shown in Fig. 5. We created the block diagram of the fracture by altering the fault diagram using Photoshop.

The materials for the fault identification test were 28 photographs of outcrops. Half of the test images contained a fault and half did not. All of the no-fault images contained multiple rock bedding layers, fractures, or both. The images were collected from geoscientist collaborators and from geoscience-related websites. None of the images had been used in the previous experiments.

Procedure

Participants were run individually. The instructional text and fault identification test were presented on a computer running E-Prime. Participants were randomly assigned to either the Single Diagram or Paired Diagrams condition. (There were an equal number of males and females in each condition). Participants in each condition completed the same 28-item fault identification test.

Participants were asked to read the instructional text carefully. In the Single Diagram condition, they were told that they could learn about faults by examining the block diagram that was provided. In the Paired Diagram condition, participants were told that they could learn what faults are like by comparing the block diagram that displays a fault with the block diagram that displays a fracture with no faulting. Participants were allowed to ask questions before beginning the fault identification test.

In the fault identification test, participants received the 28 test images (14 fault, 14 no fault) one at a time. For each image, the participant had to decide whether it contained a fault by pressing a button on the keyboard, either “1” for *fault* or “3” for *not a fault*. These category labels were presented at the bottom of the computer screen, one next to the other, along with a reminder about the designated keys. When the participant responded that the image contained a fault, they were then asked to place two Xs on different parts of the same fault. The Xs appeared on the computer screen and could be dragged and dropped into position using the mouse. After placing both Xs, the next image was presented. When the participant responded that the image did not contain a fault, the next image was presented immediately. Participants received all 28 images in random order. There was no feedback following the participant’s decisions.

We verified that the participants had no prior geoscience course experience using the same survey of geology/Earth sciences experience as in the other experiments.

Results

For trials in which the participant correctly responded that the image contained a fault, we had to determine whether the two Xs were placed correctly on the fault line. To do so, the coordinate points for each pair of Xs were plotted on a scatterplot and overlaid on top of the original image. If the points fell on the same fault, the trial was scored as a hit. Cases in which the participant responded that a fault was present but did not correctly place the Xs on a fault were scored as misses. We also created a scatterplot for trials in which the participant *incorrectly* responded that the image contained a fault. In such cases, the participant was still prompted to drag and drop the two Xs; however, there was not actually a fault in the image. These false alarm responses shed light on the information that the participant used to identify faults in general. Before discussing false alarm responses, we consider the overall fault identification performance of the two groups.

On average, participants who received the paired diagrams in the instructions ($M = 0.92$, SD = 1.01) performed better than those who saw the single diagram ($M = 0.08$, SD = 0.72), $t(22) = 2.33$, $p < .05$, $d = 0.11–1.80$ (95 % CI). We also explored the basis of the paired diagram advantage by analyzing each component of d-prime—hits and false alarms—separately. The Paired Diagrams condition had about the same average number of hits ($M = 6.00$,

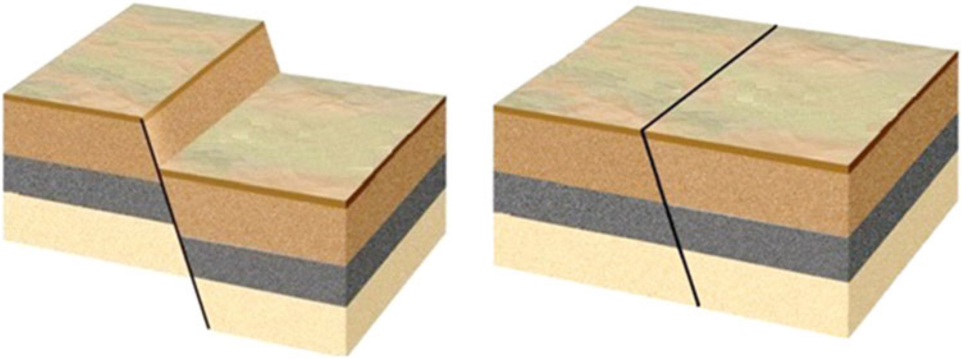


Fig. 5 Block diagrams showing a geological fault (*left*) and a fracture with no fault (*right*) from Experiment 3

SD = 3.84) as the Single Diagram condition (*M* = 5.67, SD = 2.23), *t*(22) = 0.26, *ns*, *d* = −0.70 to 0.90 (95 % CI); however, the Paired Diagrams condition had significantly fewer false alarms (*M* = 2.17, SD = 1.99) than the Single Diagram condition (*M* = 5.25, SD = 2.53), *t*(22) = 3.32, $p < .05$, *d* = 0.47–2.24 (95 % CI). Comparing the fault and no-fault pair of block diagrams appears to have increased the participants' ability to correctly reject images that contained multiple rock bedding layers and fractures but no fault. There was no significant difference in response time between the Single Diagram (*M* = 5.95 s, SD = 3.18 s) and Paired Diagram conditions (*M* = 5.03 s, SD = 1.22 s), *t*(22) < 1, $p > .35$, *d* = −0.43 to 1.19 (95 % CI).

The performance of the Single Diagram condition suggests that participants who lacked the comparison advantage failed to make a clear distinction between fractures and faults (i.e., fractures along which there has been displacement). The scatterplots for the false alarm responses support this conclusion. Figure 6 displays participants' (incorrect) responses for two of the no-fault images from the fault identification test (the red circles represent where participants dropped the Xs; there were 2 Xs per participant). All of the Xs appear on locations where the rock had fractured.

Discussion

The results of Experiment 3 show that participants who could compare a block diagram of a fault to a highly alignable diagram without a fault were better at identifying faults in real-world cases. Comparing the fault and no-fault block diagrams appears to have helped participants gain an understanding of the distinction between faults and fractures. When participants were presented with only the fault block diagram, as is often the case in textbooks and other forms of instruction (Jee et al. 2010), they were about as likely to identify a fault correctly as they were to misidentify a prominent fracture as a fault. This pattern held despite the fact that participants had been provided with a conceptual definition that emphasized the displacement of the rock along the fracture and a block diagram that clearly illustrated this displacement.

General discussion

The results across three experiments suggest that learning to classify faults is facilitated through analogical comparison. Experiment 1 showed that paired classification performance was superior to classification of single images, especially when the contrasting pair was superficially similar. Experiment 2 extended this result by presenting similar contrasting examples early versus late in learning. This study found that receiving similar pairs early in learning—the progressive alignment sequence—led to superior classification performance overall. Progressive alignment appeared to benefit participants with background geoscience experience the most, perhaps because these participants had sufficient knowledge to be able to analyze geological images into their relevant features.

In Experiment 3, we tested whether comparison of block diagrams would aid in learning about faults. Block diagrams are often used to depict geologically relevant structures for novice students, but they are generally shown singly. We found that participants who compared alignable contrasting block diagrams—one with a fault and one with only a fracture—were significantly better at subsequently classifying faults versus fractures than were those who saw only a single block diagram of a fault. Specifically, the participants who viewed the single diagram were far more likely to misidentify fractures as faults than the participants who compared two diagrams.

The results of the experiments support a perhaps counterintuitive point: rather than overloading a novice learner,

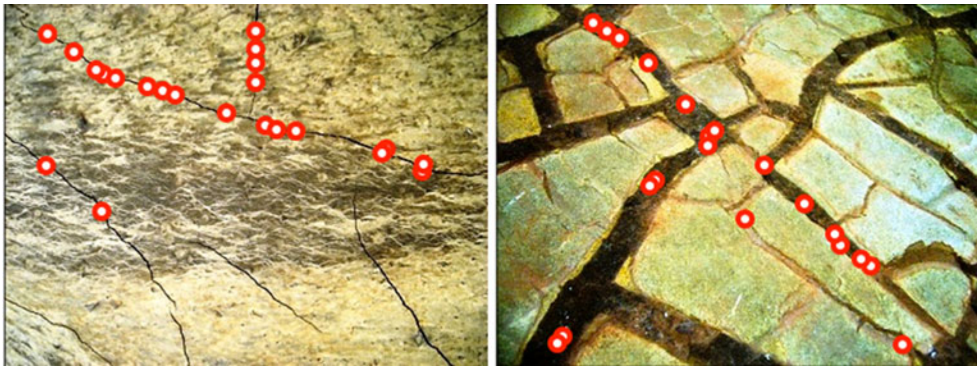


Fig. 6 False alarm responses on the fault identification test. The *red circles* represent the locations where the participant thought a fault was located

simultaneously processing a pair of novel examples can lead to deeper and more extensive understanding than processing each example separately. One explanation of this result is that processing a pair of items was both more effortful and more rewarding—a "desirable difficulty" (cf. Bjork 1994). Yet, there is evidence that comparison of highly similar images can make feature detection *easier*. High-similarity comparison makes certain features (alignable, contrasting features) "pop out" of a visual display (Sagi et al. 2012). Thus, comparison of similar, contrasting cases could direct attention to the relevant features and require fewer attentional resources from the learner.

Our findings are consistent with much developmental work showing that comparison aids children's learning (Christie and Gentner 2010; for a review, see Gentner 2010). Developmental research has also shown that learning from contrasting cases is facilitated when the cases are easy to align (Gentner et al. 2007; Kotovsky and Gentner 1996; Kurtz et al. 2001; Thompson and Opfer 2010). The present research shows that techniques that are successful with young learners—contrast and progressive alignment—can be applied to older learners (see also Loewenstein et al. 1999) and complex spatial concepts. Comparing similar examples facilitates analogical mapping, making structural commonalities and alignable differences more salient (Gentner and Markman 1994; Markman and Gentner 1993, 1996). Thus, a student is more likely to notice the central features of a concept—features high in both category and cue validity—by comparing two examples than by processing each example separately.

When contrasting cases are compared early in learning, as in Experiments 2 and 3, the learner may be more likely to generate an accurate conceptual representation. However, if comparison is not possible or examples are difficult to align, the student may form an incorrect representation. This could explain why the participants in the Dissimilar First condition in Experiment 2 performed worse than the Similar First condition, even for the similar pairs that they received later in the task. The false alarm data from Experiment 3 suggest that novice learners are likely to confuse faults and fractures. Participants who saw a single block diagram of a fault were likely to identify fractures as faults, apparently insensitive to signs of displacement. Comparing the fault and no-fault diagrams highlighted the displacement of the blocks of rock along the fracture and allowed the participant to understand that fractures were not a valid cue for classification.

Although the present study focused on between-categories contrasts, comparing examples *within* a category can also enhance learning. Yamauchi and Markman (2000) found that when the features of exemplars had several possible instantiations, comparing exemplars from the same category greatly improved classification learning. These within-category comparisons highlighted the abstract commonalities between the feature instantiations, facilitating the learning of rules that capture the commonalities. Higgins and Ross (2011) claimed that within-category comparisons are generally more effective than between-categories contrasts when the surface features of category members vary widely, as in the Yamauchi and Markman (2000) materials. Yet, such generalizations require qualification. It is important to keep in mind the goal of learning—what will the conceptual representation be used for? In the present study, participants often had to distinguish between faults and fractures, which are structurally similar and also share many surface features. Comparing two examples of faults could highlight the spatial structure of the category, but would not highlight the distinction between faults and fractures (i.e., displacement) that was crucial to classification performance. The type of comparison that is more effective will depend on the knowledge that is required to complete the task at hand.

The results of the present study have straightforward applications to instruction. For example, the findings of Experiment 3 suggest that providing a visually similar contrasting case in a textbook, instructional webpage, or lecture could highlight the relevant spatial structure of the target concept and enhance student learning. Indeed, alignment and contrast could support learning of a wide variety of structural concepts in geology, including different types of faults (Jee et al. 2011), as well as concepts in other domains (e.g., Kok et al. 2013). Of course, the effectiveness of an analogy is also influenced by the instructional supports that surround it (Harrison and Treagust 1993; Iding 1997; Jee et al. 2010; Sibley 2009; Treagust 1993). As Jee et al. (2010) discuss, analogical instruction can be enhanced when instructors explicitly map the correspondences (e.g., Richland et al. 2007) or when the student does so (e.g., Kurtz et al. 2001). This explicit mapping can help the student focus on the relevant relational structure of the cases and enhance their ability to notice alignable differences. Students may also benefit by having the instructional examples available during learning. In the case of learning the fault concept, the availability of contrasting block diagrams would enable the student to compare new cases with a known example of a fault and a non-fault (a within-category and a between-categories comparison). To get the most out of analogical comparison, instructors should think about the representation that they want their students to acquire and use analogies and other methods that best support the acquisition of this representation.

In conclusion, analogical comparison is a potentially powerful process for spatial learning in geoscience and other STEM disciplines. Future research is needed to explore how different types of comparison affect learning, how comparison can be applied to different concepts, and how analogical processing could be supported in real-world learning environments.

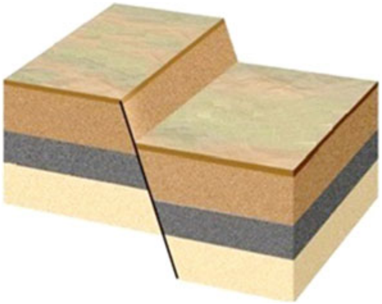
Acknowledgments The authors thank Maggie Carlin, Greg Erickson, and Lynn Koehler for assistance with data collection. This research was supported by NSF grant SBE-0541957, the Spatial Intelligence and Learning Center (SILC), and the Humboldt Foundation and the Hanse-Wissenschaftskolleg, which provided support to the third author (Dedre Gentner) during preparation of this paper.

Appendix: Instructional text from Experiment 1

To make a mountain, Earth forces lift cubic kilometers of rock skyward against the pull of gravity. The process of mountain forming not only uplifts the surface of the crust, but also causes rocks to undergo deformation, a process by which rocks squash, stretch, bend, or break in response to squeezing, stretching, or shearing. Geologists refer to the changes in shape caused by deformation as strain. Sometimes the rock changes only temporarily and then changes back when the force that caused the strain is removed—an elastic strain. Rocks can also develop permanent strain, in two fundamentally different ways. In ductile deformation, a material changes shape without breaking, like a ball of dough squeezed beneath a book. However, during brittle deformation, a material breaks into two or more pieces, like a plate shattering on the floor.

A fault is an example of a brittle deformation. A fault is a fracture in Earth's crust along which there has been slipping (or displacement) of the rocks. The amount of displacement can vary from a fraction of an inch to many thousands of feet. Some faults, like the San Andreas, intersect the ground surface and thus displace the ground when they move. Others involve the sliding of rock at depth within the crust and remain invisible at the surface unless later exposed by erosion. Movement along a fault generally takes place suddenly, commonly involving distances up to 20–40 feet, and rarely more.

You can see that a fault is a fracture in Earth's crust along which there has been slipping (or displacement) by examining the fault block diagram below. The diagram displays a fracture, and there has clearly been movement along it. In order to identify a fault, you must find evidence that the rocks on either side of an apparent fracture have been displaced. Note that this movement can be in any direction along the fracture, that the amount of movement can vary in different faults, and that the angle of the fault may also vary.



In this task, you will be presented with photographs of geological structures. Some of the photographs will contain a fault. Your job is to determine which photographs display a fault.

References

Baenninger M, Newcombe N (1989) The role of experience in spatial test performance: a meta-analysis. Sex Roles 20:327–344

Biederman I, Shiffrir M (1987) Sexing day-old chicks: a case study and expert systems analysis of a difficult perceptual learning task. J Exp Psychol Learn Mem Cogn 13:640–645

Bjork RA (1994) Memory and metamemory considerations in the training of human beings. In: Metcalfe J, Shimamura AP (eds) Metacognition: knowing about knowing. MIT Press, Cambridge, pp 185–205

Changing the Equation in STEM Education (2010) Retrieved 12 June 2012 from the whitehouse.gov website: http://www.whitehouse.gov/blog/2010/09/16/changing-equation-stem-education

Chase WG, Simon HA (1973) Perception in chess. Cogn Psychol 4(1):55–81

Christie S, Getner D (2010) Where hypotheses come from: Learning new relations by structural alignment. J Cogn Dev 11(3):356–373

Downs R, DeSouza A (eds) (2006) Learning to think spatially: GIS as a support system in the K-12 curriculum. National Academies Press, Washington D.C.

Gentner D (1983) Structure-mapping: a theoretical framework for analogy. Cogn Sci 7:155–170

Gentner D (1988) Metaphor as structure mapping: the relational shift. Child Dev 59:47–59

Gentner D (2005) The development of relational category knowledge. In: Gershkoff-Stowe L, Rakison DH (eds) Building object categories in developmental time. Erlbaum, Hillsdale, pp 245–275

Gentner D (2010) Bootstrapping the mind: analogical processes and symbol systems. Cogn Sci 34(5):752–775

Gentner D, Markman AB (1994) Structural alignment in comparison: no difference without similarity. Psychol Sci 5:152–158

Gentner D, Namy L (1999) Comparison in the development of categories. Cogn Dev 14:487–513

Gentner D, Loewenstein J, Thompson L (2003) Learning and transfer: a general role for analogical encoding. J Educ Psychol 95(2):393–408

Gentner D, Medina J (1998) Similarity and the development of rules. Cognition 65(2):263–297

Gentner D, Loewenstein J, Hung B (2007) Comparison facilitates children's learning of names for parts. J Cogn Dev 8:285–307

Gentner D, Anggoro FK, Klibanoff RS (2011) Structure-mapping and relational language support children's learning of relational categories. Child Dev 82(4):1173–1188

Gick ML, Holyoak KJ (1983) Schema induction and analogical transfer. Cogn Psychol 15:1–38

Gobert JD, Clement JJ (1999) Effects of student-generated diagrams versus student-generated summaries on conceptual understanding of causal and dynamic knowledge in plate tectonics. J Res Sci Teach 36:39–53

Goldstone RL, Medin DL, Gentner D (1991) Relational similarity and the non-independence of features in similarity judgment. Cogn Psychol 23:222–264

Harrison AG, Treagust AG (1993) Teaching with analogies: a case study in grade-10 optics. J Res Sci Teach 30:1291–1307

Hegarty M, Keehner M, Khooshabeh P, Montello DR (2009) How spatial ability enhances, and is enhanced by, dental education. Learn Individ Differ 19:61–70

Higgins E, Ross B (2011) Comparisons in category learning: how best to compare for what. In: Proceedings of the 33rd annual conference of the Cognitive Science Society. Cognitive Science Society, Austin, pp 1388–1393

Holyoak KJ, Koh K (1987) Surface and structural similarity in analogical transfer. Mem Cogn 15:323–340

Iding MK (1997) How analogies foster learning from science texts. Instr Sci 25:233–253

Jee BD, Uttal DH, Gentner D, Manduca C, Shipley T, Sageman B, Ormand CJ, Tikoff B (2010) Analogical thinking in geoscience education. J Geosci Educ 58:2–13

Jee BD, Desharnais C, Staley JV (2011) Category learning from alignable examples: an application to structural geology. In: Poster presented at the 33rd conference of the Cognitive Science Society. Boston, MA

Jones M, Love BC (2007) Beyond common features: the role of roles in determining similarity. Cogn Psychol 55:196–231

Kali Y, Orion N (1996) Spatial abilities of high school students in the perception of geologic structures. J Res Sci Teach 33:369–391

Kastens KA, Ishikawa T (2006) Spatial thinking in the geosciences and cognitive sciences: a cross-disciplinary look at the intersection of the two fields. In: Manduca CA, Mogk DW (eds) Earth and mind: How geologists think and learn about the Earth: Geological Society of America Special Paper 413, pp 53–76

Kastens KA, Agrawal S, Liben L (2009) How students and field geologists reason in integrating spatial observations from outcrops to visualize a 3-D geological structure. Int J Sci Educ 31:365–394

Keane MT (1996) On adaptation in analogy: tests of pragmatic importance and adaptability in analogical problem solving. Q J Exp Psychol 49(4):1062–1085

Kok EM, de Bruin AB, Robben SG, van Merriënboer JJ (2013) Learning radiological appearances of diseases: does comparison help? Learn Instr (23):90–97

Kotovsky L, Gentner D (1996) Comparison and categorization in the development of relational similarity. Child Dev 67:2797–2822

Krawczyk D, Holyoak K, Hummel J (2004) Structural constraints and object similarity in analogical mapping and inference. Think Reason 10:85–104

Kurtz KJ, Miao C, Gentner D (2001) Learning by analogical bootstrapping. J Learn Sci 10(4):417–446

Lesgold A, Glaser R, Rubinson H, Klopfer D, Feltovich P, Wang Y (1988) Expertise in a complex skill: diagnosing x-ray pictures. In: Chi MTH, Glaser R, Farr MJ (eds) The nature of expertise. Lawrence Erlbaum, Hillsdale, pp 311–342

Libarkin JC, Brick C (2002) Research methodologies in science education: visualization and the geosciences. J Geosci Educ 50:449–455

Liben LS, Kastens KA, Christensen AE (2011) Spatial foundations of science education: the illustrative case of instruction on introductory geological concepts. Cogn Instr 29(1):45–87

Loewenstein J, Thompson L, Gentner D (1999) Analogical encoding facilitates knowledge transfer in negotiation. Psychon Bull Rev 6(4):586–597

Markman AB, Gentner D (1993) Structural alignment during similarity comparisons. Cogn Psychol 25:431–467

Markman AB, Gentner D (1996) Commonalities and differences in similarity comparisons. Mem Cogn 24:235–249

Marshak S (2005) Earth: portrait of a planet. W.W. Norton, New York

Orton J, Anggoro F, Jee BD (2012) Mutual alignment comparison facilitates abstraction and transfer of a complex scientific concept. Educ Stud 38(4):473–477

Posner MI, Goldsmith R, Welton KE Jr (1967) Perceived distance and the classification of distorted patterns. J Exp Psychol 73:28–38

Richland LE, Zur O, Holyoak KJ (2007) Cognitive supports for analogies in the mathematics classroom. Science 316:1128–1129

Sagi E, Gentner D, Lovett A (2012) What difference reveals about similarity. Cogn Sci 36(6):1019–1050

Sibley D (2009) A cognitive framework for reasoning with scientific models. J Geosci Educ 57:255–263

Solomon KO, Medin DL, Lynch E (1999) Concepts do more than categorize. Trends Cogn Sci 3(3):99–105

Sorby SA (2009) Developing 3-D spatial visualization skills. Eng Des Graph J 63:21–32

Thompson CA, Opfer J (2010) How 15 hundred is like 15 cherries: effect of progressive alignment on representational changes in numerical cognition. Child Dev 81:1768–1786

Treagust AG (1993) The evolution of an approach for using analogies in teaching and learning science. Res Sci Educ 23:293–301

Uttal DH, Cohen CA (2012) Spatial thinking and STEM education: when, why and how. Psychol Learn Motiv 57:147–181

Uttal DH, Meadow NG, Tipton E, Hand LL, Alden AR, Warren C, Newcombe NS (2012, June 4) The malleability of spatial skills: a meta-analysis of training studies. Psychol Bull. Advance online publication. doi:10.1037/a0028446

Yamauchi T, Markman AB (2000) Inference using categories. J Exp Psychol Learn Mem Cogn 26:776–795