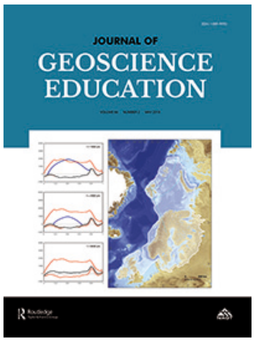

Journal of Geoscience Education

ISSN: 1089-9995 (Print) 2158-1428 (Online) Journal homepage: http://www.tandfonline.com/loi/ujge20

Reasoning about time from space: Visual continuity may disrupt reasoning about the passage of time within accreted materials

Rachel A. Myer, Thomas F. Shipley & Alexandra K. Davatzes

JOURNAL OF GEOSCIENCE EDUCATION
2018, VOL. 66, NO. 2, 147–165
https://doi.org/10.1080/10899995.2018.1451183

Reasoning about time from space: Visual continuity may disrupt reasoning about the passage of time within accreted materials

Rachel A. Myer[a], Thomas F. Shipley[a], and Alexandra K. Davatzes[b]

[a]Department of Psychology, Temple University, Philadelphia, Pennsylvania 19122-6085; [b]Department of Earth and Environmental Science, Temple University, Philadelphia, Pennsylvania 19122-6085

ABSTRACT
The ability to accurately reason using three-dimensional visualizations is vital to success in STEM disciplines, particularly the geosciences. One impediment to learning from visualizations is spatially-based misconceptions. Such errors can arise from a range of sources (e.g., prior beliefs, inaccurate application of analogy, and visual illusions). Of these sources, the potential for perceptual illusions to cause difficulty when reasoning with visualizations has been relatively unexplored. The experiments reported here consider misconceptions evident in a common type of geoscience diagram, 3D block diagrams of depositional environments (facies diagrams). Our results demonstrate a pattern of errors in temporal reasoning that we interpret to have a perceptual origin. Two experiments explored novice conceptions about the relative age of different locations in the rocks depicted in a facies diagram. Novices had difficulty reasoning about the temporal evolution of these regions. Errors appeared to be systematic, not the product of random guessing. The most common error was incorrectly treating all visually similar and connected areas as a single unit that had a common temporal origin. This error can be interpreted as the product of visual unit formation. Each experiment also explored the effectiveness of a diagrammatic intervention intended to highlight the relationship between space and time. Despite salient spatial indicators of the relationship between time and space, visual continuity and discontinuity strongly controlled temporal estimates of novice viewers.

ARTICLE HISTORY
Received 09 January 2018
Revised 22 February 2018
Accepted 06 March 2018
Published online 04 April 2018

KEYWORDS
Depositional environment; facies diagram; space for time; perceptual illusion

Introduction

Visualizations are omnipresent across the science, technology, engineering, and mathematics (STEM) disciplines. They convey information about objects and processes that are too small or large, fast or slow to be seen directly. Visualizations are also important for education and problem solving because they simplify complex situations to focus attention on a subset of the potential spatial and temporal aspects of an object or event (Hegarty, 2004). However, many students have difficulty understanding and using visualizations (Mayer, 2005), which can lead to misconceptions about important scientific concepts (Black, 2005). Understanding the nature, causes, and malleability of these misconceptions is an important part of design-based research to improve visualizations and pedagogy to better support student learning.

One source of difficulty comprehending visualizations is that many diagrams require reasoning about complex spatial relations, meaning comprehension depends on strong spatial reasoning skills. Studies have found that students with strong spatial skills have an advantage over students with weaker skills when learning with visualizations in a variety of STEM domains (for a recent meta-analysis see Höffler, 2010). One broad category of visualization challenges is reasoning about three-dimensional (3D) relationships. Learning advantages for students with strong spatial skills have been found for a variety of tasks involving 3D diagrams, including cross-sectioning a complex object that simulates a biological organ with ducts, mentally rotating chemical molecules, and cross-sectioning rock volumes (Kali & Orion, 1996; Keehner, Hegarty, Cohen, Khooshabeh, & Montello, 2008; Tuckey, Selvaratnam, & Bradley, 1991). Many STEM concepts can only be understood by grasping 3D spatial relations; thus, successful learning requires accurate understanding of 3D diagrams.

Misconceptions about 3D diagrams may arise from the same sources that contribute to misconceptions broadly, including inaccurate prior beliefs (Lawson, 1983; Libarkin & Kurdziel, 2001), incorrect inferences from analogies (Blake, 2004), and perceptual illusions

CONTACT Rachel A. Myer rachel.myer@temple.edu Department of Psychology, Temple University, Weiss Hall, 1701 N 13th Street, Philadelphia, PA 19122-6085, USA.

(Gagnier & Shipley, 2016). Overcoming misconceptions based on false beliefs or spurious analogies may be accomplished by direct instruction to replace or revise mental models (Chi, 2013). However, most visual illusions (a perceptual illusion in vision) are cognitively encapsulated (Fodor, 1983), thus making it difficult to alter misconceptions with direct conscious instruction. An example of a misconception that has been interpreted as the result of a visual illusion is the erroneous belief in block diagrams that an interior cross-section will always match the parallel face (Kali & Orion, 1996). Gagnier and Shipley (2016) found that this error was not restricted to block diagrams, it was pervasive, occurring in all situations in which people are asked about how a feature visible on a single surface projects into the volume under the surface. A visual mechanism is implicated by three observations: The appearance of patterns projecting 90 degrees into a volume (e.g., the pattern of wood grain seen on the face of a board is seen to project straight down into the board) is uniform across individuals; novices report high levels of certainty in this judgment (for unfamiliar objects the orientation is, in fact, unknowable from a single cross section); and the experience does not change for most observers even after the observers' visual estimate has been shown to be incorrect—the illusion is relatively impervious to knowledge. Thus, the pervasive difficulty in understanding block diagrams may be due, in part, to an illusion of seeing surfaces as projecting straight in from each face of the diagram. Knowing the source of the misconception allows targeted strategies for anticipating and addressing student difficulty in reasoning with cross-sections (Gagnier, Atit, Ormand, & Shipley, 2017). For example, knowing students' potential difficulty, one beneficial task can be to have students complete problems in which they are prone to errors and then show them the correct answer. Such tasks may allow students to develop spatial reasoning skills that allow them to integrate information across multiple faces, and not to stop and try to solve the problem visually, from a single face. Such explicit strategies to avoid errors resulting from normal perceptual processes may develop as part of traditional disciplinary practice (e.g., at an outcrop learning to always search for the expression of a feature, such as a fault, on surfaces with different orientations).

The potential for perceptual illusions to interfere with learning would argue for developing a program of research on STEM misconceptions that are grounded in normal perceptual processes. An initial step in constructing such a program should include identifying examples of such misconceptions. Documenting this category of misconception would allow instructors to anticipate them, and serve as a starting point for research to develop teaching strategies to mitigate their disruption of learning and practice. Given their perceptual basis, these misconceptions may be relatively impervious to simple correction, but might be best handled through informing students of their existence and providing conscious strategies to apply (just as carpenters and pilots know when making judgments about what is level, "Trust the instrument, not your eye"). In a small step toward such a broader enterprise, here we consider the potential role of systematic perceptual illusions in understanding diagrams that require reasoning about time from space—specifically, diagrams of depositional environments.

Prior work has found that students have difficulty reasoning about two-dimensional (2D) diagrams showing the temporal and spatial evolution of depositional environments (Dodick & Orion, 2003a, 2003b). Dodick and Orion found that students tend to believe that thicker strata must take more time to deposit than thinner strata. Because deposition rate depends on many factors, thinner strata could take a much longer time to deposit. Students showed improved understanding of the evolution of 2D depositional environments after viewing animations simulating deposition rates (Cheek, 2013). Here we consider students' understanding of the relationship between passage of time and sequences of sediment buildup in a near-shore ocean environment.

Three-dimensional diagrams of depositional environments provide the spatial distribution patterns that reveal how the depositional environment changed over space and time (e.g., during a transgression, shallow water depositional environments become deeper water environments, resulting in a sedimentary facies change). Fundamental to recognizing and reasoning about such processes is the spatial logic embodied by Steno's *law of horizontality*—that sediment must have originally been deposited horizontally—and *law of superposition*—that older layers underlie younger layers. Walther's Law integrates both space and time by showing that facies that occur in conformable vertical successions of strata also occur in laterally adjacent environments. Understanding that rock distribution within a horizontal plane reflects variation in sediment deposition in different sedimentary environments at the same time allows interpreting the vertical changes in rock distribution as the trace of a changing sea level over time.

Informal reports from instructors indicate that novices presented with diagrams of depositional environments, such as Figure 1, often experience difficulty understanding the complex spatial and temporal progressions. For example, students perform poorly on exam questions involving such diagrams even after completing a full lab on deposition in which they were taught Walther's Law, asked to interpret sea level changes based

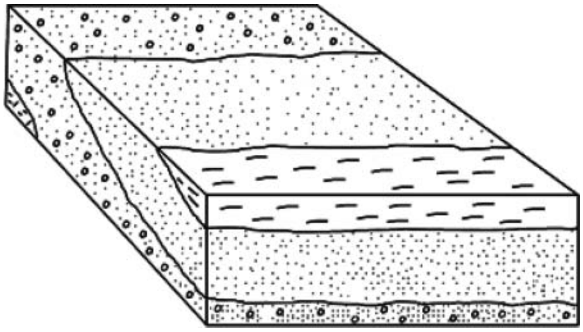

Figure 1. Basic facies diagram. The trace provided by the border between the conglomerate (the lithology with the dot and circles pattern) and the sandstone (the lithology with just the dotted pattern) gives experts all the information they need to comprehend this diagram. Experts know, and expect student to learn that, everything on the same horizontal level is the same age; everything vertically lower in the diagram is older than everything vertically higher in the diagram; and over time, the ocean rose such that the shoreline progressed from the front of the diagram toward the back of the diagram.

on stratigraphic sections, and asked to correlate stratigraphic sections using litho-, chrono-, and sequence stratigraphy. Experts, on the other hand, readily pick out the patterns in these diagrams indicating, for example, transgression or regression.

For experts, the key to understanding the temporal evolution of the area depicted in Figure 1 is simple: Younger rocks overlie older ones, unless there has been later deformation. The apparent simplicity of this concept does not align well with the difficulty students experience understanding facies diagrams (Dodick & Orion, 2003a, 2003b). Could the cause of students' difficulty be perceptual in nature? Informal observations of novices in the classroom suggested they interpreted the layers in Figure 1 as temporally organized (e.g., the conglomerate was older than the sandstone, which in turn was older than the siltstone), even when told that the diagram represented the geological result of a slowly rising ocean. Such reports suggest a mental model in which all of the same rock type was the same age. Why did novices appear to be grouping similar material together and treating it as if it formed at the same time, in direct conflict with the instructor's account of the process?

We suggest that this error may reflect basic object perception processes. The human visual system must solve the complex problem of inferring the 3D structure of objects in the world from a projection to the eye, in which nearer objects occlude more distant objects (Shipley & Kellman, 2001). The importance of grouping visible fragments was first recognized by Gestalt psychologists (Koffka, 1935), who proposed that the visual system worked according to a small set of basic principles (e.g., the principle of similarity, which states that visually similar elements will tend to be grouped together; Wertheimer, 1923, 1938). In the case of facies diagrams, areas with the same pattern would be grouped together such that each different pattern represents a separate visual unit. We note "unit" is a term of art in psychology referring to a psychological group, as in, the visible fragments of a partially occluded object will be seen as a unit, a complete object. To distinguish between the geological and psychological meanings here, we refer to a psychologically discrete entity as a "visual unit"; and when using the geoscience disciplinary term, we refer to a "lithologic unit."

Although visual grouping using similarity may be beneficial in everyday life, because things that look similar are often parts of the same object (Brunswik & Kamiya, 1953), applying this principle to a geoscience diagram would be detrimental to understanding the spatiotemporal evolution of the region. Students may grasp that lower regions of the diagram must be older, analogous to a stack of objects in which lower objects were likely in position before higher objects were placed on top. However, observers treating all similar material as one visual unit—a single thing created at the same time through a single event—would evince significant misconceptions about the age of the different regions of rock. We refer to this hypothesized source of misconceptions in facies diagrams as "visual unitization." If temporal judgments are based on a strategy of using the visual appearance of materials in the diagram, we would expect misconceptions based on unitization processes. For contrast, we consider the alternative that novices are confused and do not understand the spatial or temporal processes well enough to generate a coherent mental model. They resort to guessing when asked any question that requires an inference from the diagram.

To determine which of these strategies novices use when reasoning with this diagram, or if they possibly use some other strategy that would be revealed by systematic judgments about the time course of rock formation in the diagram, we took two approaches. We directly asked participants to report how they were making decisions about which part of the diagram was older, and we developed a set of inferential questions about the relative ages of parts of the diagram. These questions were designed to discern which strategy was likely being employed. As we anticipated misconceptions, we also created an augmented diagram in which the visual units were broken by separating lithologic units, with a new lithologic unit that was linked to a brief period of time, to make the spatial pattern of specific point(s) in time salient. Participants answered the same inferential questions for both the basic and augmented diagrams to determine the effectiveness of this intervention. Thus, the two main research questions were, “How do novices conceive of time represented in this facies diagram?” and, “Does incorporating an intervention meant to make the horizontality of time more salient improve understanding?”

The first experiment was exploratory in nature. We aimed to develop a measure of errors in temporal reasoning and characterize the pattern of any errors. In the second experiment, we refined the measure using the pattern observed in Experiment 1. Here, we aimed to find out if the pattern of errors replicated, and to learn how resistant the misconception was to multiple illustrations of the horizontal nature of time.

Experiment 1

The purpose of Experiment 1 was to explore novice understanding of the temporal evolution represented in the facies diagram (shown in Figure 1). Specifically, what strategies do novices use when asked about the ages of specific parts of the diagram? Do they (a) employ the normative strategy, judging temporal origin based on relative height regardless of lithology; (b) employ a strategy based on visual unitization, treating each connected part of a lithologic unit as having a single common temporal origin; (c) employ no systematic strategy and guess randomly; or (d) employ some other systematic strategy? Participants also made the same judgments for an augmented diagram in which we sought to make salient that all horizontal points in the rock volume were the same age through the addition of a layer of volcanic ash (Figure 2).

Method

Materials

The facies diagram represented an area where the ocean rose over time. The basic diagram shows deposition of conglomerate, sandstone, and siltstone over some bedrock. Because these diagrams were presented to geology novices, each rock type was labeled by sediment type to provide more familiar terms. To emphasize the horizontal nature of time in such environments we created a disciplinarily appropriate diagram in which a horizontal feature, an ash layer mid-block, represented a geologically instantaneous volcanic eruption. The horizontal nature of time was made salient by showing deposition over a brief time period; however, this also broke up the visual continuity of the lithologic units. To develop our understanding of how novices reason about time in this diagram, we employed a temporal sequence task in which participants determined the relative age of pairs of points in the diagram.

We selected 16 point-pairs, used in both diagrams, to explore four qualitatively different spatial relations (described in the second and third columns of Table 1, with examples in Figure 3A). We anticipated that different strategies would yield differing patterns of responses across the four categories. A random guessing strategy would lead to an equivalent pattern of choices for all types of spatial categories. The normative strategy would yield consistent judgments of “the lower point is older” and “points on the same horizontal plane are the same age,” regardless of which rocks the points were in. By contrast, if participants employed a strategy based on visual unitization, we anticipated a different pattern for each of the four categories, which we consider in turn:

- The first category, labeled DH, included all cases in which the two points were in **different** lithologic units and on the same **horizon** (Figure 3A—diamond). Employing the strategy based on visual unitization would result in concluding that the point in the “lower lithologic unit” was older, as each lithologic unit would be seen to have been deposited sequentially. Thus, points in broken up rocks would be judged to be older than points in both sand and fine silt, and points in sand would be judged older than in fine silt.
- The DVU category included all cases in which the two points were in **different** lithologic units and **vertically** offset with the vertically lower point, the

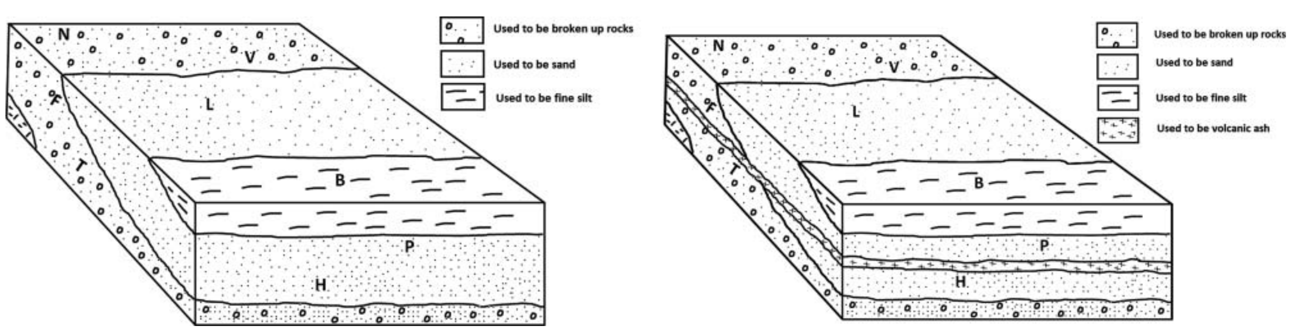

Figure 2. Basic facies diagram (left) and ash layer diagram (right) for Experiment 1 in one of the three point configurations.

Table 1. Description of spatial categories for Experiment 1.

Name	Lithologic unit(s)	Vertical point placement	Ash layer	Number of questions for the ash layer diagram
DHN	Different	Same **horizon**	Not crossed	1
DVUN	Different	**Vertically** offset, lower point (older) in **upper** lithologic unit compared to higher up (younger) point.	Not crossed	1
DVLN	Different	**Vertically** offset, lower point (older) in **lower** lithologic unit compared to higher up (younger) point.	Not crossed	1
DVLC	Different	**Vertically** offset, lower point (older) in **lower** lithologic unit compared to higher up (younger) point.	Crossed	2
SVN	Same	**Vertically** offset	Not Crossed	4
SVC	Same	**Vertically** offset	Crossed	5

Note: As there was no ash layer in the basic diagram, combining the number of questions with the same first two letters gives the number of questions in the not-crossed category for the basic diagram. Also note, two of the 16 original questions were removed from analysis due to spatial ambiguity and are not included in the "number of questions" section above.

older point, in an **upper** lithologic unit compared to the higher point, the younger point (Figure 3A—oval). The strategy based on visual unitization would lead to the vertically higher point, the one in the visually "lower lithologic unit," being judged to be older.

- The DVL category included all cases in which the two points were in **different** lithologic units and **vertically** offset with the vertically lower, older point in a **lower** lithologic unit compared to the higher up, younger point (Figure 3A—rectangle). The strategy based on visual unitization would lead to the point in the "lower lithologic unit" being judged to be older. So, both this and the normative strategy would lead to the vertically lower point being considered older, but for different reasons.
- Finally, the SV category included all cases in which the two points were in the **same** lithology and **vertically** offset (Figure 3A—oval). The strategy based on visual unitization would result in the two points being judged to be the same age since all points in the same lithologic unit would be seen to have been deposited at the same time regardless of vertical placement.

Introducing the ash layer resulted in new qualitative spatial relations for some of the point-pairs (described in the fourth column of Table 1, examples in Figure 3B). The ash layer segmented the space and resulted in some vertically offset points being separated by the ash layer. This affected points in the DVL and SV categories. Thus, DVLC (two point-pairs) and SVC (five point-pairs), refer to cases in which the space between the point-pairs was **crossed** by the ash layer. All cases in which the point-pairs were **not** crossed by the ash layer are labeled as *N*, including all cases in the basic diagram where there was no ash layer. This resulted in the following six spatial categories: DHN, DVUN, DVLN, SVN, SVLC, and SVC.

Follow up interviews revealed two questions that required comparing points on different sides of the diagram were spatially ambiguous (one DHN and one SVN). Participants found it difficult to determine if the points were on the same horizontal plane or not. These questions were not included in further analysis.

Beneath each diagram was a short text that explained how each sedimentary rock originally formed. For the basic diagram, the text was as follows:

> The above diagram shows the geologic history of an area where the ocean rose over time. Areas that used to be broken up rocks were sediment resulting from mountains eroding. Areas that used to be sand were originally close to the shoreline. As the sea level rose and the shoreline moved inland, areas where the shoreline used to be were then under water and fine silt settled out in the deeper water. Please answer the following questions about the age of the material in the diagram above by circling your answer.

For the ash layer diagram, the only difference was that the sentence, "While the ocean was rising, a volcano erupted and deposited a layer of volcanic ash," was added before the final sentence.

In order to better understand how spatial cognitive processes might be involved in reasoning about time in this diagram, we also administered two individual difference measures of spatial skills. The first was the Redrawn Set A of Vandenberg & Kuse Mental Rotation Test (MRT-A; Peters et al., 1995). This rigid mental transformation task measured mental rotation skill by asking participants to determine which two of four possibilities were rotated versions of a target object. This task was included because prior research has found a learning advantage for students with strong mental rotation skills when completing tasks involving 3D STEM visualizations (Hegarty, Keehner, Khooshabeh, & Montello, 2009; Hoyek et al., 2009; Stieff, 2007). The second was a mental folding task, the Differential Aptitude Test: Space Relations (DAT:SR; Bennett, Seashore, & Wesman, 1973). This nonrigid mental transformation task measured skill to mentally construct a 3D box

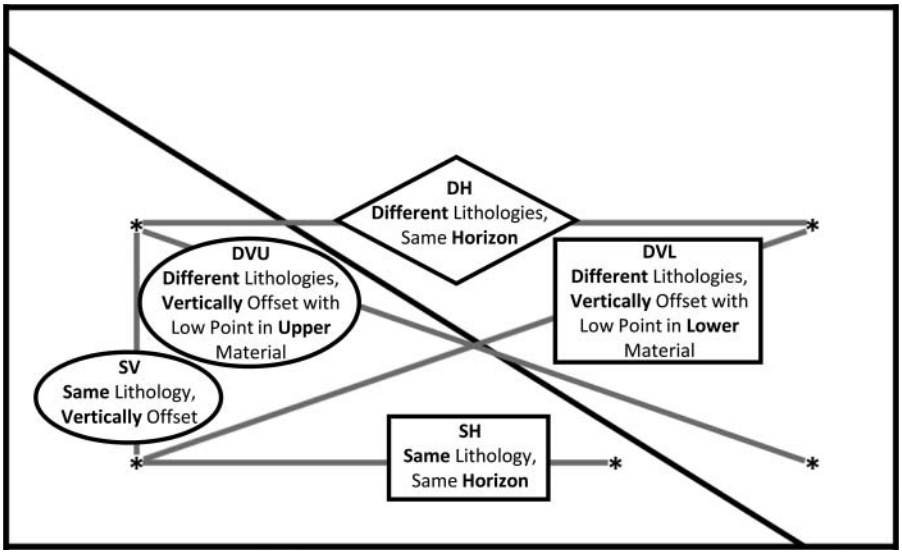

Figure 3A. A generalized schematic cross-section illustrating examples of the first two components of each spatial category (i.e., information on lithologic unit and vertical point placement are included, but whether or not ash was crossed is not included) for both experiments. Note that the SH category only appears in Experiment 2. All other categories are the same for both experiments.

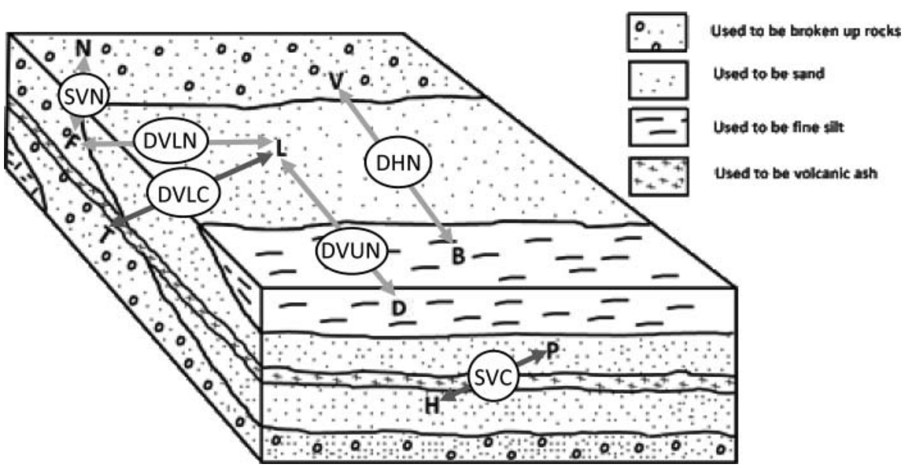

Figure 3B. Examples of all three components of each spatial category for Experiment 1 on the facies diagram. *Note:* These are just examples of what each question type would look like; pairs were not necessarily those used in the experiment; and point D was not originally part of this letter configuration.

from a 2D outline by asking participants to determine which of four images was a 3D version of a 2D target object. This task requires reasoning about which sequence of events results in a final form, which might be helpful reasoning about depositional sequences. The computerized versions were the same as traditional paper versions, except that participants were presented questions one at a time, could not revisit questions once answered, and were allowed to answer the final question they were working on when time ran out for each task.

Procedure

Participants completed 16 questions for both the basic and ash layer diagrams in counterbalanced order. Participants chose among four options for the relative age for each designated point-pair. For example, when presented with Figure 2, participants responded to "which is older P or H" and selected among the choices, "P is older," "H is older," "They are the same age," or "Cannot tell with given information." All 16 questions had a similar structure. After completing relative age questions for both diagrams, participants verbally explained their solution strategies for one set of five or six pair-wise questions. Next, they completed the MRT-A, then the DAT:SR, and finally a demographics sheet that asked about gender, age, and geology courses completed.

Participants

Sixty-three Temple University undergraduates from the psychology participant pool participated for course credit. We chose psychology students because it was unlikely they had received extensive formal education on facies diagrams and thus would serve as a good proxy for a disciplinary student encountering this diagram for the first time. We note that prior research has found that psychology students and introductory geology students have similar spatial skills (Ormand et al., 2014). All participants were native English speakers with normal or corrected-to-normal vision. The following participants were removed from all analyses: four participants for missing diagram-related data, one for choosing the third answer option for all questions, and one for having taken a geology course beyond an introductory course. One participant with missing MRT-A data was excluded only from analyses for which that data was missing. Data from these participants did not substantively change the overall pattern of findings; cases in which specific statistical results differ are noted in the text. We report data from the remaining 57 participants. Participants ranged in age from 18 to 30 years old ($M_{age} = 20.09$, $SD_{age} = 2.17$), and six were male (10.53% male), reflecting the demographics of the university's participant pool. Of these 57 participants, 28 answered questions for the basic diagram first and 29 answered questions for the ash layer diagram first.

Results

To avoid complicating interpretations with potential order-effects, in both experiments initial analyses focus on the first diagram participants completed.

Basic diagram

Overall accuracy on this task (selecting the vertically lower point as older and selecting "same age" when points were on the same horizontal plane) was low (52%), consistent with instructors' informal reports, with considerable variance ($SD = 0.26$). Performance varied widely across the types of question. Figure 4 groups the 14 questions into the four spatial categories (Table 1) and splits the SVN category between cases in which the point-pairs were on either side of the horizontal middle of the diagram and cases in which the points were on the same side of the middle, so their data can be compared to corresponding data from the ash layer diagram, where an ash layer was in the middle. Overall performance on questions in which points were in separate lithologic units and the horizontally lower point was in the visually lower lithologic unit (DVL in Figure 3A) was notably greater than the other conditions. Furthermore, performance on questions in which points were in separate lithologic units and on the same horizontal plane (DH in Figure 3A) was notably lower than the other conditions; here, participants often chose the point in the lower lithologic unit as older (39.3%) than the point in the higher unit (28.6%). When points were in the same lithology (SVN), participants were likely to judge them to be the same age (33.7%), which they rarely did when they were in different lithologic units (all other categories were less than 14.3%).

Accuracy for the four spatial categories was compared using paired samples *t*-tests. Results showed performance was significantly higher for questions from the DVLN category, in which both the normative solution strategy and the

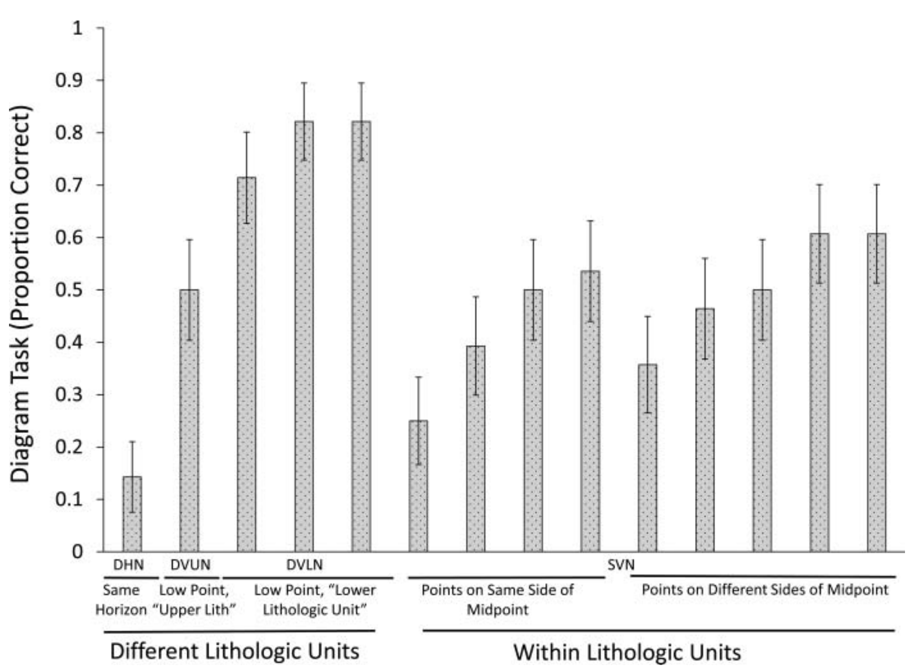

Figure 4. Diagram task scores for the basic diagram for the Experiment 1. Error bars represent +/– 1 standard error.

strategy based on visual unitization would have led participants to the same answer, than the DHN, SVN, and DVUN categories, $t(27) = 6.85$, $p < .001$, $t(27) = 3.64$, $p = .001$, and $t(27) = 2.21$, $p = .04$ (note the last of these only trends toward significance with data from all participants included). In marked contrast, performance on questions in the DHN category was lower than chance and significantly lower than the other categories, $t(27) = -3.88$, $p < .001$ and $t(27) = -3.04$, $p = .005$, compared to SVN and DVUN categories, respectively.

Ash layer diagram

Overall accuracy was slightly higher (59%) and variance lower ($SD = .19$) for the ash layer diagram. However, improved performance was mostly attributable to improvement on questions in the SVC category and one of the two questions in the DVLC category, all cases in which the ash layer divided the two points. Otherwise, the overall pattern was similar to the basic diagram, as can be seen in Figure 5, which groups the 14 questions into the six spatial categories listed in Table 1. This overall impression was confirmed by comparing the spatial categories across diagrams using Welch's t-tests. Across spatial categories, performance only differed significantly between ash and basic diagrams for the SVC category, $t(53.23) = 3.37$, $p = .001$, in which the ash layer separated the two points, effectively segmenting what had been a single visual unit. In the cases in which points are in the same lithology, there is a trend for participants looking at the ash layer diagram to show more errors than in the basic diagram (this trend reaches significance only when all participants are included in the analysis). These results suggest that the addition of the ash layer did not change how people interpret the vertical nature of time in this diagram, only how they visually segment the lithologic units in the diagram. However, it does rule out an interpretation in which participants believe each lithology (e.g., all sandstone) is the same age, because they report that the lithologic units above the ash layer are younger than those below the layer.

Order effects

We explored order effects to determine if experience with the ash layer diagram led to learning such that there was a decrease in misconceptions for the second diagram. Welch's t-tests were performed comparing the responses to the basic diagram between those who completed it first and those who completed it second, after the ash layer diagram. Performance for the basic diagram was not significantly different for the two conditions,

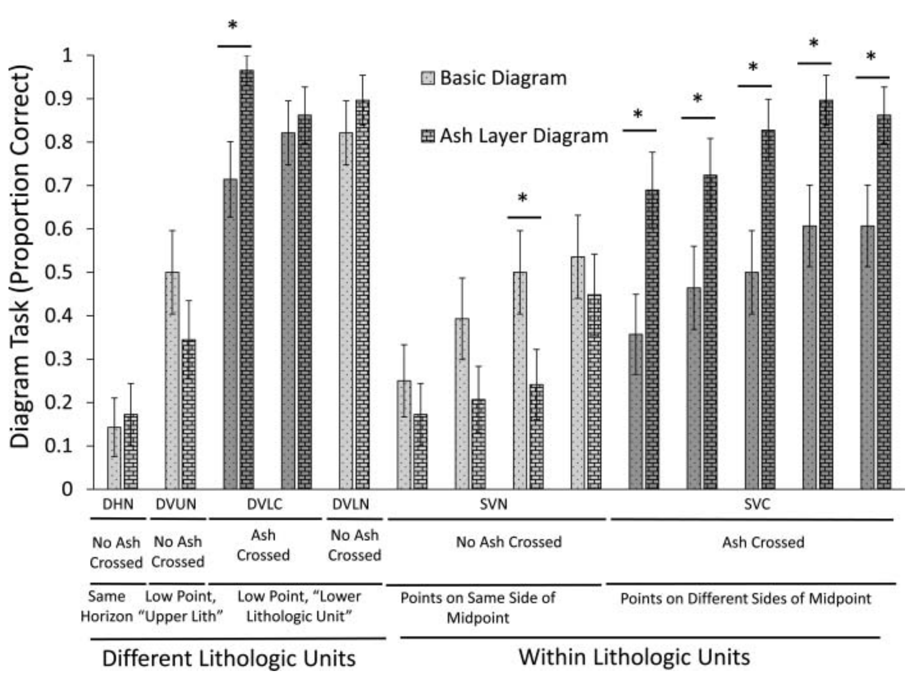

Figure 5. Diagram task scores for the basic diagram and the ash layer diagram for Experiment 1. Error bars represent +/– 1 standard error. *$p < .05$ for Welch's t-test comparing results from the ash layer diagram to results from the basic diagram. *Note:* With all participants included in the analyses, the second and last questions in the SVC category are no longer significant. Additionally, the second question in the SVN category is significant such that those who completed the basic diagram first outperformed those who completed the ash layer diagram first.

suggesting that completing the ash layer diagram first did not lead to a decrease in misconceptions.

Qualitative characterization of strategies

Information from the talk aloud suggests two predominant solution strategies: One was to reason using the normative strategy of judging relative height regardless of lithology, and the other was to judge based on the relative height of the visual units containing the points. Participants were categorized as normative or visual unitization based on their predominant strategy (the strategy used for the majority of participants' responses). Many participants alternated between these two strategies, and if a participant mentioned both strategies in response to different point-pairs, regardless of how often, they were coded as alternating between the two. The qualitative evidence for visual unitization is consistent with the quantitative findings that errors were likely when unitization indicated the incorrect answer. No other solution strategy was employed frequently. When no strategy was reported for a majority, or when the majority of responses suggested strategies different from those above, participants were coded as having an "individual strategy." When a participant reported guessing, he or she was coded as such. Finally, when two or more questions were unable to be coded for a single participant, the participant was coded as "uncodable" due to a lack of information. Table 2 shows proportions of participants for each category, criteria for each category, and examples of responses. All strategies were independently coded by two coders, following extensive discussion and review of the cases, with good reliability (percentage agreement = 78.57%; Cohen's kappa: $\kappa = .72$). Cases in which the two coders did not agree were discussed and a consensus was reached for final reporting.

Spatial tasks

The two spatial tasks, MRT-A and DAT:SR, were moderately correlated for participants who completed the basic diagram first, $r = .42$, $n = 27$, $p = .03$. Neither spatial task was correlated with performance on any of the spatial categories for the basic diagram, suggesting that higher spatial reasoning skills do not lead to increased performance on this task and that low spatial skills are not likely to lead to the misconceptions with this diagram.

Discussion

Overall, results showed that although some participants employed the normative strategy, those who employed it consistently were a small minority. In contrast, a strategy based on visual unitization was widely employed in some or all instances and often resulted in misconception about relative age. From the perspective of detecting the application of this strategy, it is notable that the use of visual unitization with its accompanying misconceptions was only evident in performance for certain categories of questions in which the vertical arrangement of visual units and the vertical height of points suggested different answers. Such a strategy based on visual unitization may represent a heuristic, supported by basic visual mechanisms, for inferring events from objects' spatial arrangements that serves to yield the correct answer for a novice some of the time. If so, it is not surprising that the MRT-A and DAT:SR were unrelated to performance, as these tests measure complex spatial reasoning, not basic visual processes.

Experiment 2

The primary goal of Experiment 2 was to see if visual unitization would be reduced by more explicitly highlighting the horizontal nature of time. Experiment 2 also included additional location pairs to address the unequal sampling in Experiment 1 and uniformly probe the different spatial categories. We reasoned that although the ash layer in Experiment 1 was treated as a break in the temporal sequence of events, perhaps its horizontal structure was not sufficiently salient to participants. To increase the salience of the horizontality of time, we revised the augmented diagram to have two ash layers (explained as the result of a sequence of two eruptions, shown in Figure 6). Thus, not only was each ash layer horizontal, the interval of time between eruptions was also horizontal. The apparent ineffectiveness of the single clear indicator that all points on the same horizon were deposited at the same time led us to question the resistance to change of the misconception based on visual unitization. This was explored in Experiment 2 through a questionnaire on response confidence, which could reveal a willingness to believe in alternate conceptions. High confidence in responses would suggest an unwillingness to change, so that misconceptions would be hard to alter (van Loon, Dunlosky, van Gog, van Merriënboer, & de Bruin, 2015). Finally, we took advantage of this replication to learn more about how much participants understood about the spatial nature of the events depicted in the diagram. To do this, after the participants had made their temporal judgments, we asked them to indicate where the deeper water would have been in the diagram and how the shoreline would have moved over time.

Method

Materials

We modified the task employed in Experiment 1 in two ways. First, in order to get a clearer understanding of

Table 2. Example verbal descriptions of strategies from Talk-Alouds.

Strategy used	Criteria	Percentage coded	Example quotes
Random guessing	Participants using this strategy mentioned guessing or were otherwise uncertain of responses for the majority of question responses.	3.6%	Participant 37: Responding to multiple questions: "I'm not so sure why I choose this answer. … Ugh I said J is older because I don't know. … I really can't explain."
Normative strategy	Participants using this strategy mentioned that they chose each answer based on the relative height of points for the majority of question responses.	25.0%	Participant 31: Description of general strategy: "For each question I just looked at the position of the letters to see which was older. So, if one was below the other, I assumed it was older considering … that as the sea level rose the shoreline moved inland." Participant 33: Description of general strategy: "I mean, it makes sense that over time things layer on top of each other, so I just made the assumption that whatever was the lowest had been there the longest, so it would be oldest."
Strategy based on visual unitization	Participants using this strategy mentioned that they chose answers based on location of lithologic unit(s) for the majority of question responses.	14.3%	Participant 19: For a question in the DVLN category: "I said Y was older because the broken rocks are before the sand sediment." For a question in the SVN category: "Q and S are the same because they're both part of the sand layer." Participant 51: For a question DHN category: "I put V because I think the old rocks would be older than um the silt" For a question in the SVN "I put, umm, that they are the same because they both are in the same category (sand)"
Alternating between normative strategy and strategy based on visual initization	Participants using these strategies mentioned aspects from both the normative strategy and the strategy based on visual unitization for the majority of question responses.	28.6%	Participant 13: For a question in the SVN category (coded as using visual unitization): "I picked they were the same age just because they were the same substance." For a different question in the SVN category (coded as using a normative strategy): "I know they're the same substances. but I also picked O is older because it seemed closer to the bottom than R." Participant 49: For a question in the SVN category (coded as using normative strategy): "I said G was older because it looked lower." For a different question in the SVN category (coded as using visual unitization strategy): "I said that they were the same age uh … because they look like they have the same description that used to be sand."
Individual strategies	Participants who used individual strategies mentioned use of a strategy (or multiple strategies) that deviated from those mentioned above and did not answer the majority of questions in a way consistent with the above categories.	21.4%	1. Participant 10: Description of general strategy: "So I figure, like, it (the shoreline) moves forward so the deeper they are, the more new they are because if they're older they get moved forward because the shoreline moves inland." (higher and forward as older). 2. Participant 15: For a question in the SVN category: "I said H was older because it was further down a little bit" (normative strategy). For a different question in the SVN category: "I put F is older because it is a little bit more pushed back" (further back as older). 3. Participant 47: Description of general strategy: "Everything that is forward is supposed to be older than what's closer to the shoreline" (forward as older). 4. Participant 53: For a question in the SVN category: "I put that H is older because it had to be there as the layer before P" (normative strategy). For a different question in the SVN category: "I choose F again because F was further back than T" (further back as older).

(*Continued on next page*)

Table 2. (*Continued*)

Strategy used	Criteria	Percentage coded	Example quotes
			5. Participant 57: For a question in the DVLN category: "I think whatever is on the bottom is older and S used to be sand and Y is the rocks" (strategy based on visual unitization). For a question in the SVN category: "Q is older than S 'cause as the sea level rises, it kind of wash out the old things so this time the old part came up to the top and new part is at the bottom" (higher as older). 6. Participant 61: For a question in the SVN category: "I did G is older and that's because they said it was closer to the shoreline" (near shoreline as older). For a different question in the SVN category: "I said O is older because it was towards the bottom" (normative strategy).
Not able to be coded	Participants' responses to two or more questions did not provide enough information to determine what strategy (or strategies) was used.	7.1%	

Note: Descriptions and percentage data come from participants who completed the basic diagram first. "Majority" strategy was taken to mean at least three of five or four of six of the questions presented were coded as describing that strategy. Additionally, three of these seven individuals coded as "normative" mentioned lithology at least once in their reasoning, but indicated that it was of minor consideration in deciding relative age, suggesting that their choices may have depended somewhat on visual unitization.

how novices reason about time and points with the same horizontal locations, pairs were added in which the two points were on the same horizontal plane and within the same lithology (the SH category). Second, additional point-pairs were added so that all spatial categories had at least two questions. We also modified stimuli to improve clarity and to provide additional geological context. To ensure that participants compared precise locations, we added points next to each of the location letters. Finally, we changed the text supporting the materials to more closely resemble material and terms that would be employed in an introductory geology course by using proper geologic terms and explaining how each type of rock was formed.

Results from Experiment 1 indicated that a common incorrect solution strategy was to say that everything in the same lithologic unit was the same age, with "lower" lithologic units being older than "upper" lithologic units. To further explore both this strategy and the normative strategy, we explored five qualitatively different spatial relations (see the second and third columns of Table 3 and Figure 3A). We included all four of the spatial categories from Experiment 1 and one new category, SH, in which the two points were in the **same** lithology and on the same **horizon** (Figure 3A—rectangle). Application of the normative strategy would result in reasoning that both points were the same age, and the strategy based on visual unitization would also result in the two points being considered the same age, as they appeared in the same lithologic unit. In these cases, both the strategies led to the same judgments but for different reasons. Adding the additional questions resulted in a total of 22 questions. The answer options for each question were the same as in Experiment 1.

When the two ash layers were added to the diagram, the ash layer(s) segmented the space and resulted in two categories of points separated by the ash layers (Table 3 and Figure 7). First, the point-pairs could be separated by a single ash layer. These cases are indicated with a

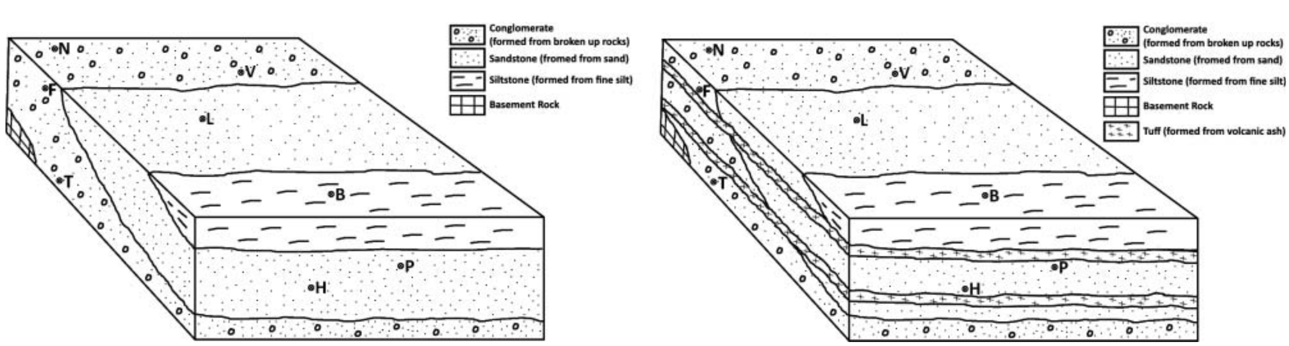

Figure 6. Basic diagram (left) and double ash layer diagram (right) for Experiment 2 in one of the four point configurations.

Table 3. Description of spatial categories for Experiment 2.

Name	Lithologic unit(s)	Vertical point placement	Ash layer(s)	Number of questions for the double ash layer diagram
DHN	Different	Same **horizon**	Not crossed	2
DVUN	Different	**Vertically** offset, lower point (older) in **upper** lithologic unit compared to higher up (younger) point.	Not crossed	2
DVUC	Different	**Vertically** offset, lower point (older) in **upper** lithologic unit compared to higher up (younger) point.	Crossed	2
DVUC2	Different	**Vertically** offset, lower point (older) in **upper** lithologic unit compared to higher up (younger) point.	Crossed 2	2
DVLN	Different	**Vertically** offset, lower point (older) in **lower** lithologic unit compared to higher up (younger) point.	Not crossed	2
DVLC	Different	**Vertically** offset, lower point (older) in **lower** material layer compared to higher up (younger) point.	Crossed	2
DVLC2	Different	**Vertically** offset, lower point (older) in **lower** lithologic unit compared to higher up (younger) point.	Crossed 2	2
SVN	Same	**Vertically** offset	Not crossed	2
SVC	Same	**Vertically** offset	Crossed	2
SVC2	Same	**Vertically** offset	Crossed 2	2
SHN	Same	Same **horizon**	Not crossed	2

Note: As there were no ash layers in the basic diagram, combining the number of questions with the same first two letters gives the number of questions in the not-crossed category for the basic diagram.

"C," where one ash layer **crossed** between the points (DVUC, DVLC, and SVC). Second, cases in which **two** ash layers **crossed** the space between the point-pairs were indicated as "C2" (DVUC2, DVLC2, and SVC2). Two questions in each of the DVUN, DVLN, and SVN categories in the basic diagram were categorized as DVUC, DVLC, and SVC in the double ash layer diagram, with an additional two questions categorized as DVUC2, DVLC2, and SVC2 in the double ash layer diagram. As in Experiment 1, all cases (including for the basic diagram, in which there were no ash layers) in which the point-pairs were **not** crossed by any ash layer are designated with "N." This resulted in the following 11 spatial categories: DHN, DVUN, DVLN, SVN, SHN, DVUC, DVLC, SVC, DVUC2, DVLC2, and SVC2.

Beneath each diagram, text similar to Experiment 1 explained how each sedimentary rock originally formed. The modified text for the basic diagram was as follows:

> The above diagram shows the geologic history of an area where the ocean rose over time. Conglomerate is formed out of broken up rocks eroded from mountains. Sandstone is formed from rocks that were broken up into sand near the shoreline. Siltstone is formed from particles that settled out in deeper water to form fine silt. Please answer the following questions about the age of material at locations indicated with ⊗s in the diagram above. Please circle your answers.

For the double ash layer diagram, two sentences were added: "Tuff is formed from volcanic ash that is deposited as a result of volcanic eruptions. The above diagram

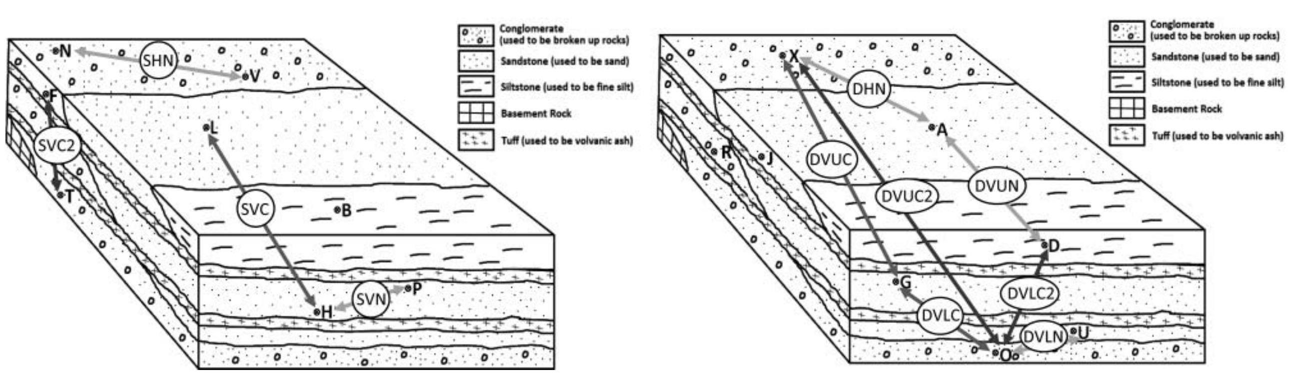

Figure 7. Examples of all three components of each spatial category for Experiment 2 on the facies diagram. The diagram on the left shows all categories within a single lithology, and the diagram on the right shows all categories between lithologic units. *Note*: These are just examples of what each question type would look like; pairs were not necessarily those used in the experiment.

represents an area where a volcanic eruption occurred and then later a second volcanic eruption occurred."

The same computerized versions of the MRT-A and DAT:SR employed in Experiment 1 were used (Bennett et al., 1973; Peters et al., 1995).

Confidence questionnaire. We added a questionnaire to measure the strength of the misconception by asking how confident participants were in their answers. This also allowed us to ask if there was a relationship between confidence and performance on the task. A positive correlation between performance and confidence, although not expected here, would suggest that participants understood the nature of the challenge and their strategies. However, a negative correlation would suggest misconceptions that may be resistant to change and likely difficulty adopting the normative solution strategy (van Loon et al., 2015).

Participants completed this measure directly following each diagram. Participants indicated how confident they were on a five-point Likert scale (from "not confident" to "confident"). They also indicated for how many of the 22 questions they were "confident," "not confident," and how many did they "guess." We included this to get a better understanding of how often random guessing was employed.

Deep water task. We added this task to investigate how well participants understood the spatiotemporal evolution of the diagram and how well they connected the story they were given to the diagram. This task used the basic diagram and narrative without the letters included. Participants shaded in where the deep water would have been on the diagram. Participants were considered correct if they shaded in either the siltstone or along the front of the diagram, not including the conglomerate. Shading the siltstone would imply a correct inference about events from the text and diagram, as the text explicitly stated that siltstone formed in deeper water. Shading the front of the diagram would also imply a correct understanding, as the area furthest to the front was always the area within the diagram that was under deepest water. Any response other than these was considered incorrect. This measure would allow us to analyze whether there was a relation between their mental model of ocean history in this diagram and strategies for judging the relative age of points.

Shoreline motion task. Like the deep water task, we designed this task to get a better understanding of how well participants understood the spatiotemporal evolution of the diagram. This task used the double ash layer diagram and story without the letters included. We identified the vertically lower ash layer "ash layer 1" and the higher layer "ash layer 2." Participants were asked how the shoreline changed from the time ash layer 1 was deposited to the time ash layer 2 was deposited. Did the shoreline move either "toward you" (toward the front of the diagram) or "away from you" (toward the back of the diagram)? The correct response was "away." Correctly inferring the answer required either accurate interpretation of the traces in the diagram or mentally animating the deposition process, both difficult tasks for novices. This measure would allow us to analyze whether there was a relationship between task accuracy and strategies for judging the relative age of points.

Procedure

Participants completed the 22 questions for both the basic diagram and the double ash layer diagram in a counterbalanced order, with the confidence questionnaire completed after each diagram. Next, they completed the MRT-A followed by the DAT:SR. Then they completed the deep water task, followed by the shoreline motion task. Finally, participants completed the same demographics sheet used in Experiment 1.

Participants

A new group of 88 Temple University undergraduates was drawn from the psychology participant pool. All participants were native English speakers with normal or corrected-to-normal vision and received course credit for participation. Four participants had missing or incomplete data from the spatial reasoning tasks and one person had missing Likert confidence data. These participants were only excluded from analyses for which they had missing data. Participants ranged in age from 18 to 27 years old (M_{age}=20.30, SD_{age}=1.84), and 28 were male (32% male). Of these, 43 completed the basic diagram first and 45 completed the double ash layer diagram first.

Results

Basic diagram

Overall accuracy (selecting the horizontally lower point as older and selecting same age when the points were on the same horizontal level) was low (59%), with considerable variance ($SD = 0.24$). Although low, the mean was higher than in Experiment 1, likely reflecting the new questions. As in Experiment 1, performance varied widely across the types of question. Figure 8 groups the 22 questions into the five spatial categories (Table 3) organized by location in same or different lithologic units, and how the points in the ash layer diagram were separated. The relative performance on the four categories that were present in both experiments were similar across experiments. Again, overall

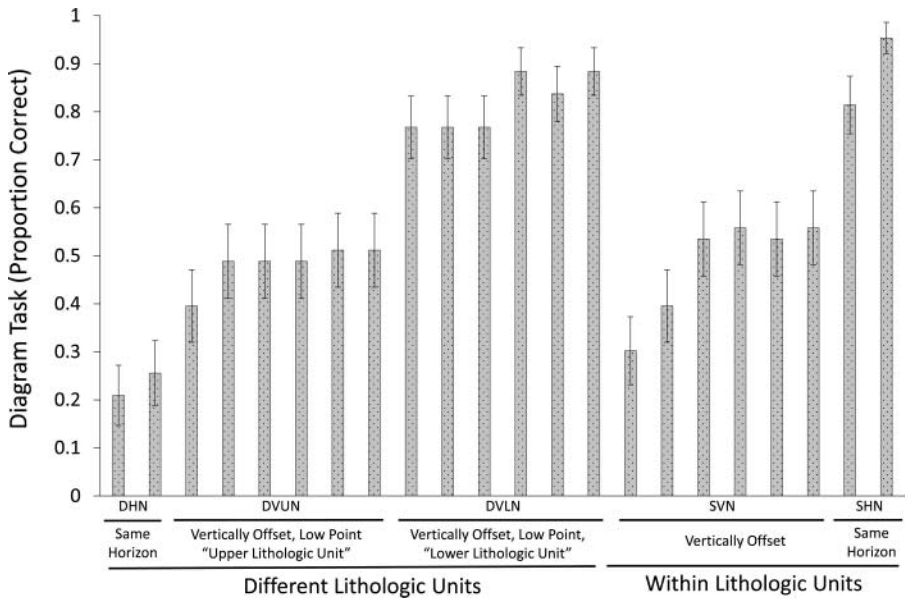

Figure 8. Diagram task scores for the basic diagram in Experiment 2. Error bars represent +/– 1 standard error.

performance on questions in which points were in separate lithologic units and the vertically lower point was also in the visually lower lithologic unit (DVL in Figure 3A, rectangle) was notably greater than the other conditions, and performance on questions in which the points were in separate lithologic units and on the same horizontal plane (DH in Figure 3A, diamond) was notably lower than other conditions. As in Experiment 1, participants more often chose the point in the lower lithologic unit as older (59.5%) than the point in the higher unit (13.0%). Additionally, participants were much more likely to say that points were the same age when they were in the same lithology (SVN: 40.1%, SHN: 90.5%) than when they were in different lithologic units (all other categories were less than 23.9%).

Accuracy for the five spatial categories was compared using paired samples t-tests. Results showed performance was significantly higher for questions in the DVLN category than the DHN, SVN, and DVUN categories, $t(42) = 7.71$, $p < .001$, $t(42) = 5.34$, $p < .001$, and $t(42) = 4.20$, $p < .001$. Participants performed significantly worse on the DHN category than the other categories, $t(42) = -3.58$, $p = .001$ and $t(42) = -3.99$, $p < .001$, compared to SVN and DVUN categories, respectively. Performance did not differ significantly for the SVN and DVUN categories. Performance on the new spatial category, in which both points were in the same lithology and on the same horizon (SH, Figure 3A, rectangle), was high. A paired samples t-tests showed performance did not differ significantly from the DVLN category, and participants performed significantly better on the SHN category than the DHN, SVN, and DVUN categories, $t(42) = 9.09$, $p < .001$, $t(42) = 5.37$, $p < .001$, and $t(42) = 5.83$, $p < .001$.

Double ash layer diagram

Overall accuracy was higher (75%) and variance lower ($SD = .16$) for the double ash layer diagram than the basic diagram. However, improved performance was mostly attributable to improvement on questions in which the points were separated by one or both of the ash layers; otherwise, the overall pattern was similar to the basic diagram (see Figure 9, which groups questions using the 11 spatial categories from Table 3). This was confirmed by comparing the spatial categories across diagrams using Welch's t-tests, comparing participants who completed the double ash layer diagram first to participants who completed the basic diagram first. Across spatial categories, performance differed significantly between double ash and basic diagrams for six of the 11 categories. These categories included five of the six categories that crossed ash: DVLC, SVC, SVC2, DVUC, and DVUC2; $t(64.68) = 2.31$, $p = .024$, $t(75.23) = 3.40$, $p = .001$, $t(73.65) = 3.47$, $p = .001$, $t(72.57) = 4.56$, $p < .001$, and $t(72.03) = 4.56$, $p < .001$. The final category that crossed ash, DVLC2, may not have differed between diagrams due to a ceiling effect. In comparison, performance only differed significantly between double ash and basic diagrams for one of the five categories that did not cross an ash layer: DHN, $t(85.09) = 2.73$, $p = .008$. These results suggest that the addition of the two ash layers generally did not change how participants interpret time within lithologies in this diagram.

It is clear from Figure 9 that there was significant variation across point-pairs within some of the spatial categories. Therefore, *post hoc* tests were run to contrast the double ash layer and the basic diagram for each point-pair. A Bonferroni correction was applied to Welch's t-test, and the results can be seen in Figure 9.

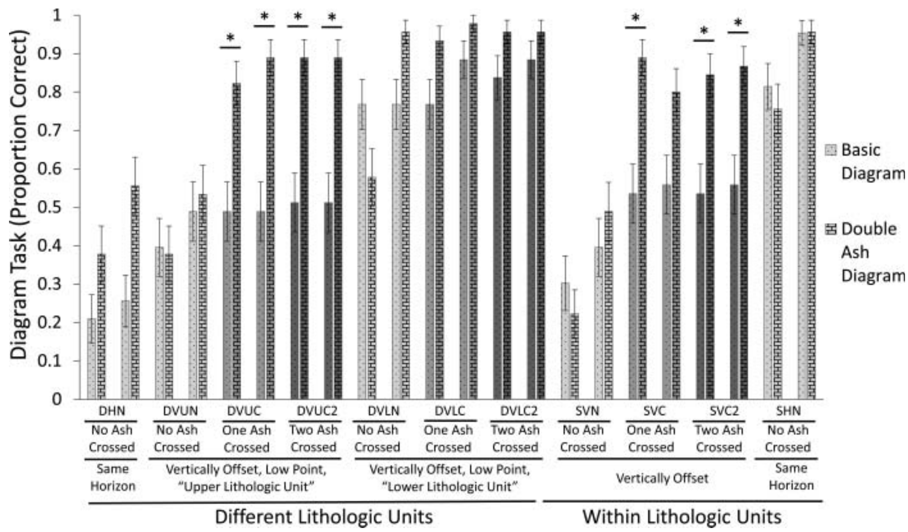

Figure 9. Diagram task scores for the basic diagram and the double ash layer diagram for Experiment 2. Error bars represent +/− 1 standard error. $^{*}p < .05$ for Welch's t-test comparing results from the double ash layer diagram to results from the basic diagram for each question with a Bonferroni correction.

Order effects

Order effects were again explored to determine if working with the double ash layer diagram led to learning and a decrease in misconceptions for the basic diagram. As in Experiment 1, Welch's t-tests indicated no significant benefit for the basic diagram after seeing the double ash layer diagram relative to completing the basic diagram first.

Confidence questionnaire

Participants who completed the basic diagram first reported low to neutral confidence on the Likert confidence test, $M = 2.81$ ($SD = .96$), with scores ranging from 1 ("not confident") to 5 ("very confident"). Participants reported being "confident" or "very confident" on 44% ($SD = .32$) of the questions. Participants reported being either "not confident" or "slightly confident" on 44% ($SD = .32$) of the questions, and guessing on 34% ($SD = .32$) of questions. No significant correlations were found between diagram task performance for the basic diagram and the Likert data, the number of guessed questions, or the number confident. The overall pattern with high confidence ratings for many questions suggests a lack of awareness of erroneous strategies consistent with a perceptual foundation. A moderate negative correlation was found between diagram task performance and proportion not confident data, $r = -.31$, $n = 43$, $p = .04$. This suggests that participants had modest metacognitive awareness of their skill on the task, as a high lack of confidence predicted low diagram performance; however, neither the number confident or number guessing ratings supported this pattern.

Deep water task

For participants who completed the basic diagram first, 76.7% ($SD = .43$) correctly completed this task. Welch's t-tests found no significant differences between those who correctly completed this task and those who did not for performance on any of the spatial categories.

Shoreline motion task

For participants who completed the basic diagram first, 60.5% ($SD = .49$) correctly completed this task. Welch's t-tests found no significant differences between those who correctly completed this task and those who did not for any of the spatial categories. These results are near chance (50%); the task may have been too difficult for novices.

Spatial tasks

As in Experiment 1, the two spatial tasks were correlated for participants who completed the basic diagram first, $r = .71$, $n = 42$, $p < .001$, and the DAT:SR was not significantly correlated with any spatial categories for the basic diagram. This time, there was a moderate correlation between the MRT-A and one spatial category, the SVN category, $r = .36$, $n = 42$, $p = .02$. We have no clear interpretation of this finding and suggest it should be followed up in future research.

Discussion

Generally, results from Experiment 2 were consistent with our initial inferences from Experiment 1 that the predominant strategy involves visual unit formation. This is evident in the relative accuracy patterns for the different spatial categories, in which accuracy was high for questions in which the visual units were consistent with the vertical height of the points, and accuracy was low when this was violated. Participants were generally only accurate judging points to be the same age when the points were in the same material. In both experiments, there was a lack of correlation between diagram understanding and complex spatial reasoning.

The intervention was not notably effective in either experiment. The benefits of a horizontal indicator of time were only evident for spatial categories where the two points were separated by the ash layer(s). The only exception to this was the DHN category. Adding the two ash layers may have improved thinking about points on the same horizontal level and in different lithologic units, but other aspects of the misconception persisted. In sum, developing an understanding that points in the same visual unit are not necessarily the same age and that vertically lower points are older regardless of lithologic unit may prove difficult.

Finally, results from the interpretation tasks suggest that participants did not fully understand the spatiotemporal evolution of this region. Correct completion of the deep water and shoreline tasks were unrelated to diagram task performance, meaning that even those who may have used a normative solution strategy may not have fully understood how this region evolved over time. That participants reported guessing the answer about a third of the time supports this interpretation. Although not very confident, participants did not recognize that they were using an error-prone strategy. The novices may have understood that they lacked necessary knowledge, which could lead to an openness to learning alternative strategies.

General Discussion

These two experiments confirmed what many introductory geology professors have already recognized: Novices have difficulty reasoning about time when presented with 3D block diagrams of depositional environments. Our findings offer a new explanation that accounts, in part, for this difficulty. The most common novice approach, which often led to misconceptions, was to judge relative time based on the relative locations of *visual units*. All points within a single connected lithology were considered the same age, and vertically lower lithologic units were considered older than upper lithologic units (e.g., all conglomerate was considered older than all sandstone). We hypothesize this inclination reflects a basic perceptual process to group visually similar elements. Supporting a lower-level cognitive mechanism was the finding that both rigid and nonrigid mental transformation skills were unrelated to participants' interpretation of this diagram, suggesting that novices do not use complex spatial reasoning when determining temporal relations in this diagram. Furthermore, this strategy (and its resulting misconceptions) was found to be resistant to diagrammatic interventions designed to make the spatial nature of time salient in the absence of instructor support. What do these findings mean for cognitive science, education researchers, and educators?

The human visual system developed through evolution and within individuals to accurately represent a cluttered environment, so that we can gracefully coordinate our actions with other objects, animate and inanimate (Gibson, 1979; Marr, 1982). The visual processes that serve to provide information about the world reflect fundamental aspects of the physics of our world, including the solid nature of objects and physical support (Kellman & Spelke, 1983). The routines work so well that we do not notice them, and our experience is akin to a naïve realist who believes the world is simply the way it appears. The nature of visual routines is often revealed by mistakes. Here we found that when the traces left by an event do not clearly mark the passage of time in the slow accretion of a solid, the visual system is left to use basic heuristics that serve reasoning about separable objects. Thus, regions that should not be grouped are grouped on the basis of similarity (Wertheimer, 1923, 1923/1938), and erroneous inferences follow.

Recognizing the perceptual basis of this misconception may be key to addressing the misconception and understanding why students make this error. This is but one instance of a misconception stemming from a perceptual illusion (e.g., Gagnier & Shipley, 2016). However, just as cognitive psychology has found many visual illusions (Luckiesh, 1922), it is possible that there are other geoscience misconceptions that stem from specific perceptual illusions. To develop a program of research on the general problem of misconceptions that have a perceptual basis, many such misconceptions need to be identified. Consider the clues that would suggest a particular misconception has a perceptual basis (is a visual illusion). Perceptual illusions tend to be widespread, consistent across people, and resistant to change, despite knowledge of the correct solution (Gagnier & Shipley, 2016). A key distinction between a widely held false belief and a misconception based on visual illusion is resistance to change. A false belief may be directly addressed with instruction. If the visual system is providing information that is inconsistent with direct instruction, the result may be confusion, or ignoring the instruction (Chi, 2013). Educators could flag for future study student errors that are widespread, consistent, and persistent.

Implications for educators

The difficulties novices experience when reasoning about the spatiotemporal relations in 3D block diagrams of depositional environments are likely similar to those students face in courses where this diagram is first introduced. Educators should be aware not just that students are likely to have trouble with these diagrams but also that misconceptions may have perceptual origins. As such, addressing this misconception may not be a simple matter of correcting an erroneous belief (Gagnier & Shipley, 2016). However, this does not mean that nothing can be done. It is possible to guide attention in such a way as to alter or reduce illusions (e.g., Wernery et al., 2015); however, further research is needed into how to do this before clear advice can be given. Educators should take care to recognize this perceptual basis, as students are likely to use this heuristic, which is correct some of the time, despite learning about Steno's law and explicit discussion of depositional processes.

It can be hard to know when a heuristic is being employed in place of a correct algorithm, especially when under some conditions the normative strategy and heuristic provide the same answer. An important research goal is identifying what strategies are employed when. This allows the educator to assess if students have achieved the desired understanding. By doing so the education researcher may develop materials in which the two strategies diverge and there is an opportunity to correct erroneous mental models by adopting new strategies. Explaining common misconceptions and their perceptual bases to students could be beneficial, just as it helps the weekend carpenter to know not to trust his or her eyes to accurately judge level. Another research approach would be to consider how experts reason with these diagrams. Although, to our knowledge, no research has directly explored how experts interact with facies diagrams, it is reasonable to assume they are not prone to the same error we have identified in novices, at least not for simple diagrams. The misconception would be avoidable with the disciplinarily appropriate strategy, but how experts avoid any perceptually prepotent bias is an open question. We speculate that protracted experience with disciplinary diagrams and multiple opportunities to correct mental models aid experts on the path to mastery. If true, successful interventions may require interacting with multiple facies diagrams and explicit opportunities to practice correcting mental models (for example, using the shoreline task as a bridging activity). Educators could use a compare and contrast strategy to structure such interactions (Rau, 2017).

As students struggle to develop accurate mental models of sedimentary environments, it may be helpful to provide animations simulating how deposition originally occurred, which has been shown to support learning about 2D depositional environments (Cheek, 2013). However, using animations to address misconceptions is not always successful (Tversky, Morrison, & Betrancourt, 2002). Additionally, if the error presented here is a visual illusion, then providing more temporal information may not help in the long run, as students might still see visual units in the diagram and continue to believe the separate lithologic units represented units of time (hence, our current advice that educators keep these errors in mind).

Limitations

An important limitation of these experiments is that only one facies diagram was used. It is possible that other facies diagrams, and more generally other diagrams intended to illustrate the passage of time using space, do not lead to the same issues presented here. Additionally, it might be possible to design a diagram that takes novices' perceptual processes into account to allow an interpretation closer to the one instructors desire. However, how to do that is not clear. Further research is necessary to find the best intervention to support reasoning about time in diagrams that do not have distinct temporal boundaries.

Generalization is also limited by our participant pool. Although participants likely represent a geology student's first approach to this facies diagram, we do not yet understand how meaning of such diagrams evolves with education. Thus, advice to instructors is necessarily limited and highlights the need for future work to understand the three-way interaction among diagram, student, and instruction.

Conclusions

Participants struggle to correctly understand the spatiotemporal evolution of depositional environments. The most common misconception was that areas that look visually similar must be the same age, ignoring lateral variability associated with a change in depositional environment controlling lithology. This error stems from everyday perceptual processes that serve natural vision, in which it is beneficial to group visually similar elements into a single object. Although there has been extensive work on STEM misconceptions (Francek, 2013; Gil-Perez & Carrascosa, 1990; Perkins & Simmons, 1988), relatively little research has focused on those with perceptual origins. Moving forward, education researchers, educators, and cognitive scientists should work together to identify and correct such misconceptions in service of students of science and the science of learning.

Acknowledgments

We thank two anonymous readers for their comments and careful read of an early draft of this article.

Funding

The work was made possible through the National Science Foundation, both its sponsored Spatial Intelligence Learning Center (SBE-0541957 and SBE-1041707), and its Science of Learning Collaborative Network grant (1640800).

References

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1973). *Differential aptitude tests, forms S and T*. New York, NY: Psychological Corporation.

Black, A. A. (2005). Spatial ability and Earth science conceptual understanding. *Journal of Geoscience Education*, *53*(4), 402–414. doi:10.5408/1089-9995-53.4.402

Blake, A. (2004). Helping young children to see what is relevant and why: Supporting cognitive change in Earth science using analogy. *International Journal of Science Education*, *26*(15), 1855–1873. doi:10.1080/0950069042000266173

Brunswik, E., & Kamiya, J. (1953). Ecological cue-validity of "proximity" and of other Gestalt factors. *American Journal of Psychology*, *66*, 20–32. doi:10.2307/1417965

Cheek, K. A. (2013). How geoscience novices reason about temporal duration: The role of spatial thinking and large numbers. *Journal of Geoscience Education*, *61*, 334–348.

Chi, M. T. H. (2013). Two kinds and four sub-types of misconceived knowledge, ways to change it, and the learning outcomes. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 49–70). New York, NY: Routledge.

Dodick, J., & Orion, N. (2003a). Cognitive factors affecting student understanding of geologic time. *Journal of Research in Science Teaching*, *40*(4), 415–442. doi:10.1002/tea.10083

Dodick, J., & Orion, N. (2003b). Measuring student understanding of geological time. *Science Education*, *87*, 708–731. doi:10.1002/sce.1057

Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Francek, M. (2013). A compilation and review of over 500 geoscience misconceptions. *International Journal of Science Education*, *35*(1), 31–64. doi:10.1080/09500693.2012.736644

Gagnier, K. M., Atit, K., Ormand, C. J., & Shipley, T. F. (2017). Comprehending 3D diagrams: Sketching to support spatial reasoning. *Topics in Cognitive Science*, *9*(4), 883–901. doi:10.1111/tops.12233

Gagnier, K. M., & Shipley, T. F. (2016). Visual completion from 2D cross-sections: Implications for visual theory and STEM education practice. *Cognitive Research: Principles and Implications*, *1*(1), 1–18.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton, Mifflin.

Gil-Perez, D., & Carrascosa, J. (1990). What to do about science "misconceptions". *Science Education*, *74*(5), 531–540. doi:10.1002/sce.3730740504

Hegarty, M. (2004). Dynamic visualizations and learning: Getting to the difficult questions. *Learning and Instruction*, *14*, 343–351. doi:10.1016/j.learninstruc.2004.06.007

Hegarty, M., Keehner, M., Khooshabeh, P., & Montello, D. R. (2009). How spatial abilities enhance, and are enhanced by, dental education. *Learning and Individual Differences*, *19*, 61–70. doi:10.1016/j.lindif.2008.04.006

Höffler, T. N. (2010). Spatial ability: Its influence on learning with visualizations—a meta-analytic review. *Educational Psychology Review*, *22*, 245–269. doi:10.1007/s10648-010-9126-7

Hoyek, N., Collet, C., Rastello, O., Fargier, P., Thiriet, P., & Guillot, A. (2009). Enhancement of mental rotation abilities and its effect on anatomy learning. *Teaching and Learning in Medicine*, *21*(3), 201–206. doi:10.1080/10401330903014178

Kali, Y., & Orion, N. (1996). Spatial abilities of high-school students in the perception of geologic structures. *Journal of Research in Science Teaching*, *33*(4), 369–391. doi:10.1002/(SICI)1098-2736(199604)33:4%3c369::AID-TEA2%3e3.0.CO;2-Q

Keehner, M., Hegarty, M., Cohen, C., Khooshabeh, P., & Montello, D. R. (2008). Spatial reasoning with external visualizations: What matters is what you see, not whether you interact. *Cognitive Science*, *32*, 1099–1132. doi:10.1080/03640210801898177

Kellman, P. J., & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology*, *15*, 483–524. doi:10.1016/0010-0285(83)90017-8.

Koffka, K. (1935). *Principles of Gestalt psychology: International library of psychology, philosophy, and scientific method*. New York, NY: Harcourt, Brace.

Lawson, A. E. (1983). Predicting science achievement: The role of developmental level, disembedding ability, mental capacity, prior knowledge, and beliefs. *Journal of Research in Science Teaching*, *20*(2), 117–129. doi:10.1002/tea.3660200204

Libarkin, J. C., & Kurdziel, J. P. (2001). Research methodologies in science education assessing students' alternative conceptions. *Journal of Geoscience Education*, *49*(4), 378–383.

Lukiesh, M. (1922). *Visual illusions: Their causes, characteristics, and applications*. New York, NY: Van Nostrand.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, CA: W. H. Freeman.

Mayer, R. E. (Ed.). (2005). *The Cambridge handbook of multimedia learning*. New York, NY: Cambridge University Press.

Ormand, C. J., Manduca, C., Shipley, T. F., Tikoff, B., Harwood, C. L., Atit, K., & Boone, A. P. (2014). Evaluating geoscience students' spatial thinking skills in a multi-institutional classroom study. *Journal of Geoscience Education*, *62*, 146–154. doi:10.5408/13-027.1

Perkins, D. N., & Simmons, R. (1988). Patterns of misunderstanding: An integrative model for science, math, and programming. *Review of Educational Research*, *58*(3), 303–326. doi:10.3102/00346543058003303

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg & Kuse mental rotations test: Different versions and factors that affect performance. *Brain and Cognition*, *28*, 39–58. doi:10.1006/brcg.1995.1032

Rau, M. A. (2017). Conditions for the effectiveness of multiple visual representations in enhancing STEM learning. *Educational Psychology Review*, *29*(4), 717–761. doi:10.1007/s10648-016-9365-3

Shipley, T. F., & Kellman, P. J. (Eds.). (2001). *From fragments to objects: Segmentation and grouping in vision*. Amsterdam, The Netherlands: Elsevier Science.

Stieff, M. (2007). Mental rotation and diagrammatic reasoning in science. *Learning and Instruction*, *17*, 219–234. doi:10.1016/j.learninstruc.2007.01.012

Tuckey, H., Selvaratnam, M., & Bradley, J. (1991). Identification and rectification of student difficulties concerning three-dimensional structures, rotation, and reflection. *Journal of Chemical Education*, *68*(6), 460–464. doi:10.1021/ed068p460

Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human-Computer Studies*, *57*, 247–262. doi:10.1006/ijhc.2002.1017

van Loon, M., Dunlosky, J., van Gog, T., van Merriënboer, J. J. G., & de Bruin, A. B. H. (2015). Refutations in science text lead to hypercorrection of misconceptions held with high confidence. *Contemporary Educational Psychology*, *42*, 39–48. doi:10.1016/j.cedpsych.2015.04.003

Wernery, J., Atmanspacher, H., Jürgen, K., Candia, V., Folkers, G., & Wittmann, M. (2015). Temporal processing in bistable perception of the Necker cube. *Perception*, *4*(2), 157–168. doi:10.1068/p7780

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt, II. [Investigations in Gestalt Theory: II. Laws of organization in perceptual forms]. *Psychologische Forschung*, *4*, 301–350. doi:10.1007/BF00410640

Wertheimer, M. (1938). Laws of organization in perceptual form. In W. D., Ellis (Ed.), *A source book of Gestalt psychology* (pp. 71–94). London, UK: Routledge & Kegan Paul. (Original work published 1923)