Journal of Experimental Psychology: Applied

# Spatial Alignment Supports Comparison of Life Science Images

Nina K. Simms, Bryan J. Matlen, Benjamin D. Jee, and Dedre Gentner

Online First Publication, April 6, 2023. https://dx.doi.org/10.1037/xap0000471

CITATION

Simms, N. K., Matlen, B. J., Jee, B. D., & Gentner, D. (2023, April 6). Spatial Alignment Supports Comparison of Life Science Images. *Journal of Experimental Psychology: Applied*. Advance online publication. https://dx.doi.org/10.1037/xap0000471

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: Applied

ISSN: 1076-898X

https://doi.org/10.1037/xap0000471

# Spatial Alignment Supports Comparison of Life Science Images

Nina K. Simms[1], Bryan J. Matlen[2], Benjamin D. Jee[3], and Dedre Gentner[4]
[1] Spatial Intelligence and Learning Center, Northwestern University
[2] WestEd, San Francisco, California, United States
[3] Department of Psychology, Worcester State University
[4] Department of Psychology, Northwestern University

Visual comparisons are pervasive in science, technology, engineering, and mathematics (STEM) instruction and practice. In previous work, adults' visual comparisons of simple stimuli were faster and more accurate when the layout of a display facilitated alignment of corresponding elements—the *spatial alignment principle* (Matlen et al., 2020). Here, we asked whether the spatial alignment principle extends to rich, educationally relevant stimuli, and how prior experience and spatial skill relate to spatial alignment effects. Participants were asked to find an incorrect bone within a skeleton, presented individually or paired with a correct skeleton in a layout that did (direct placement) or did not (impeded placement) support alignment (Kurtz & Gentner, 2013). Consistent with the spatial alignment principle, undergraduates (Study 1) showed an advantage of direct over impeded placement. Middle schoolers (Study 2) showed a direct advantage on items presented in atypical orientations. That atypical items showed the strongest effects suggests that direct placement may help most when materials are less familiar. However, neither individual differences in undergraduates' STEM course history, nor undergraduates' or middle schoolers' spatial skills moderated spatial alignment effects. Thus, applying the spatial alignment principle in science, technology, engineering, and mathematics has potential to improve visual comparisons, especially those that are challenging, for students of all spatial skill levels.

**_Public Significance Statement_**

Visual comparisons are integral to learning in science. In this study, both undergraduate and middle-school students were better at finding differences between atypical life science images if the spatial layout supported aligning the images. This was true regardless of participants' spatial skill, suggesting that considering the spatial layout of figures in science education has potential to improve visual comparison for students of all spatial skill levels, across grade levels.

*Keywords:* visual comparison, spatial alignment, visualizations, science, technology, engineering, and mathematics, structure mapping

Visual comparisons are pervasive in human experience. We make them when deciding which suitcase will fit our stack of travel necessities, putting together a jigsaw puzzle, or even finding the freshest produce at the market. Comparing two visually present examples is especially important in the fields of science, technology, engineering, and mathematics (STEM), where visuospatial representations like models, graphs, and diagrams are critical for learning and reasoning (e.g., Ainsworth, 2008; Arcavi, 2003;

---

Nina K. Simms https://orcid.org/0000-0003-0830-5194
Bryan J. Matlen https://orcid.org/0000-0001-5989-6997
Benjamin D. Jee https://orcid.org/0000-0003-1007-8883
Dedre Gentner https://orcid.org/0000-0002-5120-6688

The research reported here was supported by the Spatial Intelligence and Learning Center at Northwestern University, created through National Science Foundation Science of Learning Grants SBE-0541957 and SBE-1041707 to Dedre Gentner, and the Institute of Education Sciences, U.S. Department of Education, through Grant R305A170165 to WestEd. The opinions expressed are those of the authors and do not necessarily represent views of the Institute or the U.S. Department of Education.

Nina K. Simms played lead role in data curation, formal analysis, investigation, project administration, visualization and writing–original draft, supporting role in conceptualization and methodology and equal role in supervision and writing–review and editing. Bryan J. Matlen played supporting role in formal analysis, project administration and visualization and equal role in conceptualization, funding acquisition, methodology and writing–review and editing. Benjamin D. Jee played supporting role in formal analysis, project administration and writing–original draft and equal role in conceptualization, funding acquisition, methodology and writing–review and editing. Dedre Gentner played lead role in resources, supporting role in formal analysis, investigation, project administration and writing–original draft and equal role in conceptualization, funding acquisition, methodology, supervision and writing–review and editing.

The study design and hypotheses were not preregistered; the analytic plan was also not preregistered. All data and analytic code needed to reproduce analyses are available at https://osf.io/5hyn6/?view_only=85d1a6bf95d84d24a17764636fb9133b (Simms, 2023). Materials are available by request to Nina K. Simms.

Correspondence concerning this article should be addressed to Nina K. Simms, Spatial Intelligence and Learning Center, Northwestern University, 2120 Campus Drive, Evanston, IL 60208, United States. Email: ninasimms@northwestern.edu

Eilam & Gilbert, 2014; Evagorou et al., 2015; Moore et al., 2013; Parsons & Sedig, 2014; Schönborn & Anderson, 2006). Comparisons between a model and its real-world counterpart, graphs of different data sets, and diagrams of structures or processes are just a few routine examples common across STEM disciplines. Thus, understanding the processes that underpin visual comparison will not only enrich our models of cognition but also improve our ability to support STEM instruction and learning.

Here, we focus on whether and how spatial arrangement affects the processing of STEM-relevant visual comparisons. In previous work, adults' visual comparisons of simple stimuli were faster and more accurate when stimuli were arranged in accord with the *spatial alignment principle*, derived from structure-mapping theory (SMT; Matlen et al., 2020). In this work, we investigate whether this principle extends to rich, educationally relevant images, for undergraduate and middle-school students. If so, this would have important implications for how visual comparisons are used and supported in the instruction and practice of STEM.

## Learning From Visual Comparisons

Visual representations are integral to STEM. When presented along with text, students can integrate the information from these complementary sources to develop rich, conceptual understanding (Ainsworth, 2008; Mayer, 2021; Schnotz & Wagner, 2018). Decades of research have established a set of multimedia principles that support learning from visual representations (Mayer, 2017). These principles operate by reducing extraneous processing and facilitating essential and generative processing, and they help students interpret visual representations and synthesize insights with text. The spatial arrangement of diagrams can influence students' conceptual understanding by inviting mental models that align with the spatial structure. Learning and performance are enhanced when a diagram is structurally compatible with the conceptual material but are stifled when the diagram is a poor structural match (Schnotz & Bannert, 2003).

Often, deriving insights learning from educational materials also requires comparisons within and between visual representations (Jee et al., 2022; Rau, 2017). Visual comparisons are valuable tools for learning. For example, in mathematics, comparing two worked examples fosters conceptual and procedural knowledge (Begolli & Richland, 2016; Rittle-Johnson & Star, 2007). In geoscience, comparing diagrams of geological structures facilitates the identification of real examples of those structures (Jee et al., 2013), and comparing timelines helps students grasp geological timescales (Resnick et al., 2017). In engineering, comparing stable and unstable structures can promote discovery of critical structural principles (Gentner et al., 2016; Hoyos & Gentner, 2017). In medicine, comparing radiographs of healthy and diseased patients improves medical students' diagnostic accuracy (Kok et al., 2013).

However, students can fail to benefit when visual comparisons are too difficult or unclear. For example, children were more likely to learn a key engineering principle when they compared stable and unstable structures that were visually *similar* as opposed to dissimilar (Gentner et al., 2016). When given two highly similar items, children readily compared them and discovered the key difference (a diagonal brace) that distinguished the stable from the unstable structure; but when given two dissimilar items, this key difference was far less likely to be noticed. Supporting visual comparison may be especially important for those with limited prior experience in a domain (Chen & Klahr, 1999; Guo & Pang, 2011; Mix, 2008; Rittle-Johnson et al., 2009; Rittle-Johnson & Star, 2009). As students gain knowledge in a domain, they become able to efficiently encode visual representations, focusing on key domain-relevant information (Braithwaite & Goldstone, 2015; Cook, 2006; Rau, 2017); but early in learning, learners need considerable support.

## Visual Comparison as Structural Alignment

Visual comparison can be understood as a process of structural alignment, as described by SMT (Gentner, 1983, 2010; Gentner et al., 2016; Gentner & Markman, 1997; Kurtz & Gentner, 2013; Matlen et al., 2020; Sagi et al., 2012). According to SMT, in a structural alignment, elements of one stimulus are put into one-to-one correspondence with the elements of another based on matching relational structure. For example, Figure 1 shows a diagram of two food chains, each depicting hierarchical predation relationships among species in two different environments. Such diagrams demonstrate that organisms can be both predator and prey and highlight that organisms higher in the hierarchy depend on those lower in the hierarchy—even those they do not directly prey upon. Importantly, aligning the full structures of the diagrams further reveals notable roles in the hierarchies—for instance, apex predators—that share important commonalities across environments.

During alignment, priority is given to structurally connected matches over idiosyncratic matches (Gentner, 1983, 2010). In Figure 1, the apex predator (brown shark) in the saltwater food chain (right) looks quite similar to the penultimate predator (brown pike) in the freshwater food chain (left). Indeed, both are fish. However, the shark and the pike do not ultimately correspond because the overarching hierarchical structure dictates that the apex predators should correspond instead. This applies to relations as well—although shrimp (freshwater) and crabs (saltwater) are both eaten by other organisms, these two *eaten by* relations do not correspond in the final alignment because the more coherent hierarchical structure incorporating relations among all predators/prey takes priority. That is, according to SMT, correspondences between systematic,[1] higher order structure (e.g., *hierarchical predation*) govern correspondences of lower order matches (e.g., *eaten by* relations) so that they are consistent with the overall alignment.

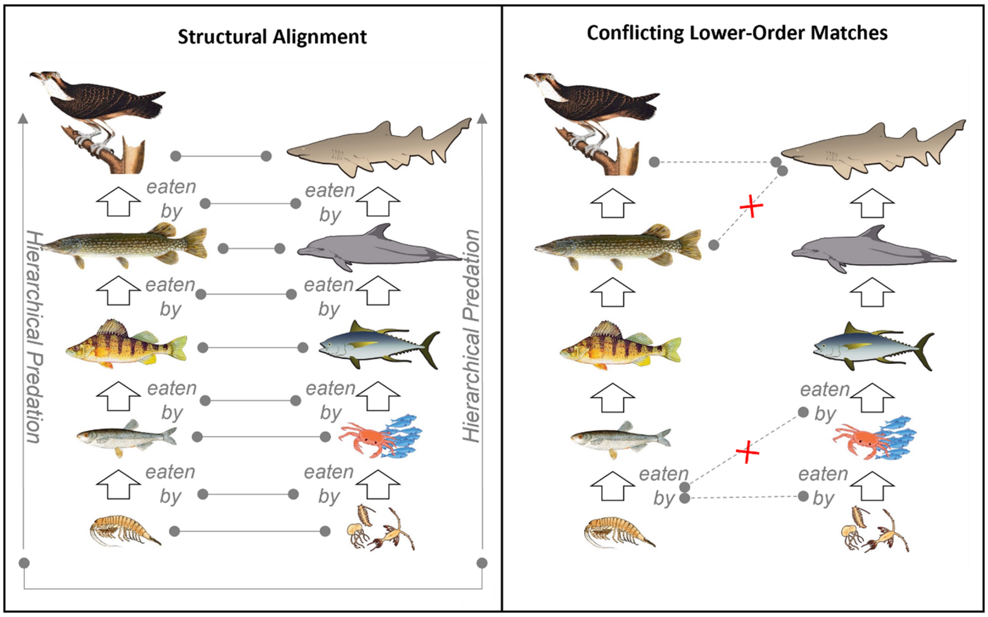
Achieving a structural alignment highlights the common relational structure, and this supports learning and abstracting of the highlighted structures (e.g., Christie & Gentner, 2010; Gadgil et al., 2012; Gick & Holyoak, 1983; Kurtz et al., 2001). Structural alignment also highlights alignable differences—differences between the instances that play the same role in the common structure (e.g., Gentner et al., 2007; Gentner & Gunn, 2001; Hoyos & Gentner, 2017; Jee et al., 2013; Markman & Gentner, 1993, 1996; Sagi et al., 2012). Once aligned, candidate inferences can also be projected from one instance to another, providing opportunities for further insight and refinement (Lassaline, 1996; Markman, 1997). These structure-mapping outcomes contribute to the substantial benefits of learning and reasoning with visual comparisons.

---

[1] More precisely, in structure mapping, the *systematicity principle* refers to an implicit preference for systems of relations governed by higher order constraining relations, such as causality, or (as in Figure 1) an overarching relation of transitivity (Falkenhainer et al., 1989; Gentner & Forbus, 2011).

**Figure 1**

*Schematic of a Structural Alignment and Conflicting Lower Order Matches for a Diagram of Food Chains*



*Note*. Annotations of theorized processes are in gray. The left figure shows a complete and systematic mapping, with correspondences (solid bars) governed by the overarching structure of hierarchical predation. The right figure shows a subset of potential object (top) and relational (bottom) matches (dotted bars). Matches that conform to the overarching structural match should be accepted in the final alignment; those that conflict with the overarching structure should be rejected. Such conflicting local matches can cause interference during alignment processing. See the online article for the color version of this figure.

However, when lower order matches (here, perceptually similar organisms) conflict with the systematic higher order alignment (here, the transitive *eaten-by* sequence comprising hierarchical predation), they can interfere with processing. Comparisons like this one, in which matching objects play different roles in two examples, are called cross-mappings and can prevent children (and sometimes adults) from achieving a structural alignment (e.g., Gentner & Toupin, 1986; Loewenstein & Gentner, 2005; Markman & Gentner, 1993; Paik & Mix, 2006; Richland & McDonough, 2010; Ross, 1987). Inconsistent lower order matches can disrupt alignment processes, for example, by making them less efficient or more prone to erroneous correspondences. Thus, according to SMT (and as implemented computationally in the structure-mapping engine; Falkenhainer et al., 1989; Forbus et al., 2017), structural alignments are easier to achieve when lower order matches are consistent with the systematic alignment and conflicting matches are minimized (Gentner & Colhoun, 2010; Gentner & Toupin, 1986; Sagi et al., 2012).

## Spatial Supports for Alignment

Structural alignment is not only affected by object similarity—in visual comparisons, the spatial arrangement of a display can also help or hinder alignment. All visual arrays have spatial structure. In some cases, that spatial arrangement is meaningful. For example, it may iconically represent spatial relationships (e.g., a diagram of a skeleton preserves integral spatial relations of real skeletons) or be used to structure nonspatial information (e.g., lifecycle diagrams organize events in a circle to illustrate cyclical reproduction). In other cases, spatial arrangement is incidental (e.g., an array of animals showcasing variability in a trait, arranged in no particular order). Regardless, the spatial organization of the items affects the alignment process (Hribar et al., 2012; Kurtz & Gentner, 2013; Matlen et al., 2020; Paik & Mix, 2008; Sagi et al., 2012).

Displays that make structural alignment harder disrupt visual comparison (Kurtz & Gentner, 2013; Matlen et al., 2020; Sagi et al., 2012). For example, Kurtz and Gentner (2013) gave participants an error-detection task in which they were asked to find anomalous bones within diagrams of skeletons. Participants were told that the skeletons were reconstructed by student archaeologists who had made a mistake in each. On some trials, the student reconstructions were shown alone (solo); on others, they were presented next to a correct, expert-constructed skeleton (in pairs). Not surprisingly, participants were far more accurate when skeletons were presented in pairs than solo. The opportunity to visually compare highlighted the key alignable differences between the skeletons, allowing participants to find the anomalous bone in the student skeleton more easily. Importantly, for our purposes, there was a second key factor in this study: The paired skeletons differed in how readily they could be aligned. Some pairs were presented facing the same way (high alignable), while others were presented facing opposite directions (low alignable; Figure 2), making them harder to align. Participants were more accurate (Experiment 1) and more efficient (Experiment 2) on high-alignable than low-alignable pairs. The benefits of comparison extended beyond online processing to improve error detection on new solo items compared to studying items individually.

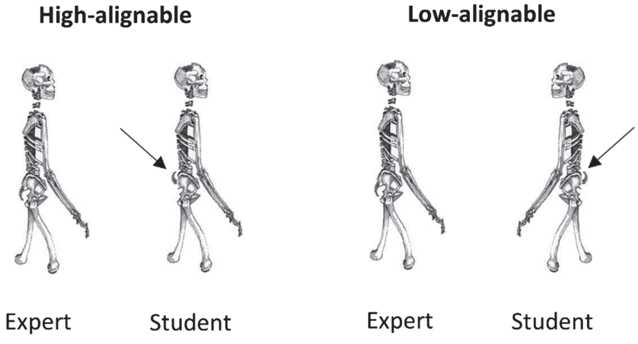
## Spatial Alignment Principle

Building on evidence that spatial structure can influence alignment, Matlen et al. (2020) proposed a general *spatial alignment principle* to describe the optimal placement of paired figures during visual comparison. Structured visual representations often have a primary axis of orientation. The skeletons in Figure 2, for example, have vertical axes, as do the food chain diagrams in Figure 1. When a display includes two vertical structures that are to be compared, alignment should be easiest when they are placed in such a way as to limit interference from spurious matches, that is, a side-by-side or horizontal layout. In this layout, correspondences can be drawn directly from one structure to the other, unobstructed—a *direct* placement. In contrast, when vertical structures are laid out vertically (one below the other), this creates an *impeded* placement. In this layout, finding the correspondences requires visually passing through other visual elements, drawing attention to noncorresponding matches that can disrupt efficient alignment processing (Gentner & Colhoun, 2010). For horizontal structures, the optimal placement is reversed—vertical placement is direct, and horizontal is impeded (see Figure 3, for examples). Thus, the spatial alignment principle states that visual comparisons should be most successful and efficient when spatial placement is *direct*.

Across several experiments, Matlen et al.'s (2020) findings were consistent with the spatial alignment principle, providing further evidence that structure-mapping processes underpin visual comparison. They gave participants pairs of shape triplets (like those in Figure 3) and asked them to judge whether they were the same or different. In some cases, the triplets in a pair were composed of the same objects, so that both object and relational similarity were relevant for same/different judgments; in other cases, the triplets were composed of different objects, so only the relational structure could match. Either way, when pairs were presented in direct placement, participants were significantly faster and more accurate than when they were presented in impeded placement (see also Zheng et al., 2022, for a parallel pattern in young children). An important point to clarify is that performance was influenced specifically by interference from potential, but incorrect, correspondences and not just visual noise. Placing a black rectangle between a pair of triplets in direct placement did not disrupt performance, but placing an irrelevant triplet between the pair led to more errors and slower reaction times (Matlen et al., 2020, Experiment 2). The irrelevant triplet contained matches that competed with the correct correspondences of the target pair.

**Figure 2**

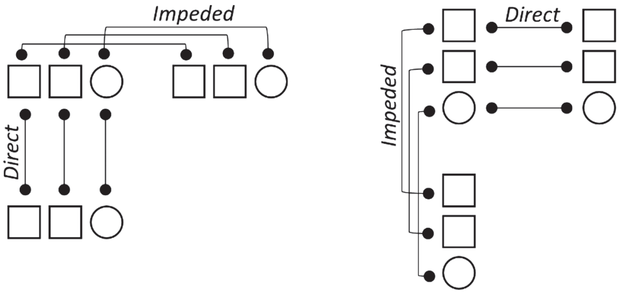
*High- and Low-Alignable Pairs From Kurtz and Gentner (2013)*



*Note.* Arrows point to the errors in the student skeletons.

**Figure 3**

*Correspondences in Direct and Impeded Placement According to the Spatial Alignment Principle*



*Note.* Figure adapted from Matlen et al. (2020). Bars represent correspondences between figures. In a direct placement, correspondences can be drawn directly between figures, unobstructed. In impeded placement, correspondences are drawn through other elements that introduce additional potential, but incorrect, correspondences, as signified by the crossing bars.

In sum, visual comparisons are frequent, beneficial for learning, and indispensable in STEM. As reviewed above, there is considerable evidence that visual comparisons—including those involving scientifically relevant materials—are carried out through a process of structural alignment. There is also strong evidence for the spatial alignment principle—that the spatial arrangement of figures is important in supporting alignment (Matlen et al., 2020)—but only for very simple materials. This could have important ramifications for STEM instruction, as students are often expected to perform visual comparisons when learning new material (Jee et al., 2022). Thus, there is great potential to improve instruction and learning. But first, we must ask whether the spatial alignment principle applies to complex, STEM-relevant materials.

## The Present Studies

The aim of these studies was to assess whether the spatial alignment principle extends to rich, educationally relevant stimuli. In two studies, we presented undergraduate (Study 1) and middle-school (Study 2) students with realistic diagrams of human and nonhuman animal skeletons and asked them to find an anomalous bone (as in Kurtz & Gentner, 2013). This task requires participants to notice small differences between visually presented items, something that is foundational to many aspects of discovery and learning in science. For example, comparing skeletons from earlier and later evolutionary time points can highlight structural changes that provide insight into the process of natural selection.

Participants saw skeletons either individually (solo) or in pairs (one correct skeleton and one containing an error). Our prediction, based on the spatial alignment principle, was that direct placement would facilitate alignment for pairs, leading to greater accuracy and/or faster reaction times,[2] compared to impeded placement (Matlen et al., 2020). Based on prior research, we also expected higher accuracy for pairs than solo items (Kurtz & Gentner, 2013).

In addition, we explored two factors that might moderate placement effects: participants' prior experience and their spatial skill. Prior experience with certain types of figures or with the content being represented can foster more efficient processing and expert-like encoding of visual representations (Canham & Hegarty, 2010; Miller et al., 2016; Shah & Freedman, 2011). This in turn may facilitate alignment during visual comparison (Jee et al., 2013; Rau, 2017; Rittle-Johnson et al., 2009). Students may be less reliant on support from direct spatial placement when they are familiar with the representations (Gentner et al., 2011; Jee & Anggoro, 2019; Rittle-Johnson & Star, 2009). In the present studies, we operationalized prior experience in two ways: First, between subjects in terms of their experience with skeletal structures (through STEM coursework); and second, within subjects in terms of the conventionality of the items (whether a skeleton was shown from canonical or noncanonical viewpoint; cf. Palmer et al., 1981). In both cases, we predict that spatial alignment effects should be greater when students are less familiar with the items. That is, we expect a larger advantage of direct over impeded placement for participants who lack relevant coursework, and for items with noncanonical presentations.

Another factor that could moderate the effects of spatial alignment is spatial skill. Higher spatial skill is associated with better learning from visual representations (Höffler, 2010). Visual comparison involves encoding and mentally manipulating visuospatial representations, so it seems likely that spatial skill will contribute to successful alignment. Moreover, impeded placement may place greater demands on these skills by introducing interference from intervening elements. If so, students with better spatial skills should experience a reduced cost of impedance and be better able to compensate for the visuospatial demand imposed by impeded placements (Höffler & Leutner, 2011). In these studies, we assessed spatial skill using standard measures and correlated performance on these tasks with placement effects. If spatial skills moderate placement effects, then we should see larger spatial alignment effects in low-skilled than in high-skilled students. In general, we expect the more challenging an alignment between visuospatial representations is to achieve, the more that spatial placement will matter.

# Study 1

## Method

### *Transparency and Openness*

Research activities were approved by institutional review board (IRB) at the first author's university and adhered to all ethical and local legal guidelines. This study was not preregistered. We have reported how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. All data and analytic code needed to reproduce the analyses are available at https://osf.io/5hyn6/?view_only=85d1a6bf95d84d24a17764636fb9133b. Materials are available by request.

### *Participants*

Forty undergraduate students were recruited from the introductory psychology subject pool at the first author's university. Students were given course credit for their participation. Five participants were excluded for poor performance on catch trials (see Design and Procedure section), leaving 35 participants in the final sample ($M_{\text{age}} = 19$ years). Thirty of these participants completed a demographics survey. This sample included 13 males, 15 females, and two other genders; was predominantly White (19 White, four Asian/Asian American, six multiracial, one unreported) and non-Hispanic (26 not Hispanic, two Hispanic, two unreported); and reported high parental educational attainment (18 graduate degrees, nine bachelor's degrees, two General Educational Development Test/high school, one did not complete high school).

**Sample Size.** Sample size was determined by a power analysis in G*Power 3.1.9.4 (Faul et al., 2007) for the key $t$ test between the direct and impeded conditions, calculated based on the average of the smallest effect size for this comparison reported in Matlen et al. (2020), $d = .55$, and the effect size for the parallel comparison reported in Kurtz and Gentner (2013), $d = 0.43$. At 80% power, $\alpha = 0.05$, and $d = 0.49$, a sample size of 35 was estimated.

### *Design and Procedure*

Participants completed two tasks: The skeleton error-detection task, followed by a spatial skills assessment.

**Error-Detection Task.** The error-detection task was modeled after Kurtz and Gentner (2013). In the cover story, participants were told that they would be viewing skeletons that had been assembled by student archaeologists, who had made a single mistake in each. The student skeletons could appear by themselves (solo), or with a second, correct skeleton that had been assembled by an expert archaeologist (paired). Participants' task was to find the erroneous bone in each student skeleton and click on it using the computer mouse as quickly as possible.

Test stimuli consisted of 24 human and animal skeleton pairs from Kurtz and Gentner (2013; Figure 4). "Mistakes" in the student skeletons were created by changing the size, orientation, or shape of a bone in the original, correct skeleton (the "expert" skeleton).

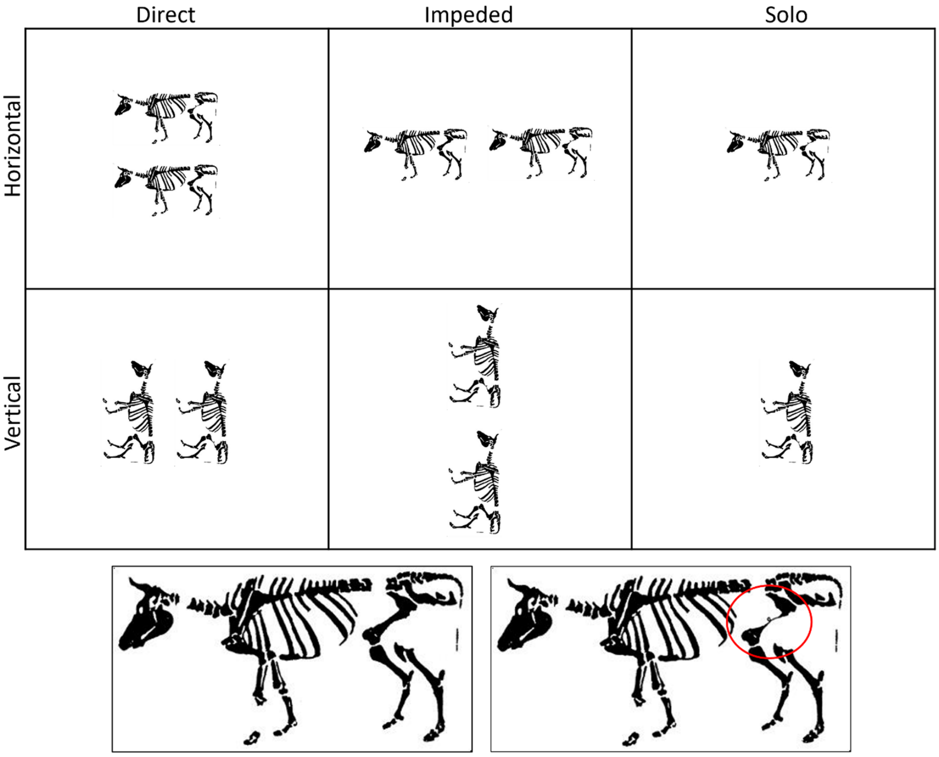
Skeletons were presented with the main axis oriented horizontally or vertically (Axis). Student skeletons were presented solo or paired with the expert example in direct or impeded placement (Placement; Figure 4). This resulted in a within-subjects Axis (2) × Placement (3) design. Horizontal skeletons laid out one above the other were in direct placement, and those laid out side by side were in impeded placement. Likewise, vertical skeletons laid out side by side were in direct placement, and those laid out one above the other were in impeded placement. Participants completed 24 test trials and saw each student skeleton only once. Axis and Placement were counterbalanced and varied within participants. Six fixed-order stimuli lists were created that balanced, as well as possible, the distribution of error type (size, orientation, shape), animal type (e.g., human, dinosaur, ungulate), and canonical orientation (see below) across Axis and Placement conditions.

---

[2] Because of the speed–accuracy trade-off (e.g., Wickelgren et al., 1980), we cannot predict whether the difference will show up more strongly in response time versus in accuracy.

**Figure 4**

*Sample Stimulus in Each Condition of the Error-Detection Task*



*Note.* Stimuli adapted from Kurtz and Gentner (2013). The bottom pair enlarges the horizontal, impeded example from the table to highlight the anomaly (circled). See the online article for the color version of this figure.

Because familiar objects like humans and other animals often have a canonical axis of orientation (Blanz et al., 1999; Palmer et al., 1981)—for example, cows are usually seen oriented horizontally—our manipulation of Axis provided the means for producing figures that were matched in complexity, but (by virtue of their noncanonical orientations) less familiar—for example, vertical cows. Out of 24, 20 stimuli items were identified as having a canonically horizontal (13) or vertical (seven) orientation. Four others did not have a canonical orientation (human hands and frogs).

Prior to the test trials, participants completed warm-up and practice trials. The warm-up was designed to acclimate participants to the trial procedure and to ensure that they could click on a precise location within a figure quickly and accurately. The 10 warm-up trials consisted of shapes filled with geometric, lattice-like patterns with an "x" embedded in the image. Participants were asked to click on the "x" as quickly and precisely as possible.

Following the warm-up, participants completed four practice trials to familiarize them with the stimuli and ensure that they understood the instructions. Practice trials consisted of one nonskeleton figure (a chair with a misplaced leg) and three solo skeletons with obvious mistakes (one each illustrating a size, shape, and orientation error).

After the practice section, participants proceeded to the test trials. The trial procedure for the warm-up, practice, and test trials was the same. To begin a trial, participants clicked on a center fixation cross, ensuring that the cursor was at the center of the screen at the start of the trial. After 250 ms, the stimuli appeared and remained on screen until the participant clicked the mouse to respond. After their response, accuracy feedback ("Correct" or "Incorrect") was displayed for 1,500 ms before proceeding to the next trial or section of the task. No further information was given to participants about the correct answer. For each stimulus, the smallest rectangle that could encompass the correct object (e.g., the anomalous bone) was defined as the target area; clicks that fell within the target area were considered correct and clicks anywhere else on the screen were considered incorrect. The time to respond to the stimulus was also recorded (RT).

Following the 24 test trials, participants received three catch trials consisting of solo skeletons with obvious mistakes (e.g., limb attached to a skull). For participants, catch trials were indistinguishable from the test trials and ensured that participants were still attending and complying with the task. Participants who missed more than one catch trial were excluded from analysis.

**Spatial Skills Assessment.** The Mental Rotations Test (MRT; Peters, 1995) was used as an indicator of participants' spatial skills. The MRT is one of the most widely used tests of spatial visualization ability and has been found to correlate with a variety of STEM outcomes, including anatomy (Guillot et al., 2007). Thus, although mental rotation is not assumed in our process model, the MRT is an established assessment of visuospatial ability. The MRT requires participants to mentally rotate block figures to find those that match a standard. On each problem, a standard is presented on the left with four figures on the right. Two of the figures depict the standard object from a different angle (i.e., rotated), and two depict different objects that cannot be rotated to match the standard. Participants' task was to find the two objects that could be rotated to match the standard. The task was administered on paper in two timed sections. Participants were given 3 min to solve 12 problems in each section, with several minutes to rest between sections. Participants were told in advance that they may not complete all problems in each section within the time limit. Problems with both and only the correct figures identified were given 1 point; problems that were not completed, only partially completed, or that had one or more incorrect figures selected were given 0 points. Points were summed for a total score out of a possible 24.

## Results

### Placement

The effects of placement were assessed using participants' accuracy and reaction time for correct responses (RT) on the error-detection task. Based on prior work, we expected that (a) participants would be more accurate, but slower, when skeletons were presented in pairs compared to solo (Kurtz & Gentner, 2013) and that (b) participants would be more accurate and/or faster when skeletons were presented in direct than impeded placement (Matlen et al., 2020).

**Accuracy.** Participants' accuracy was entered into a 3 (Placement: Direct, Impeded, Solo) × 2 (Axis: Horizontal, Vertical) repeated measures analysis of variance (ANOVA; see Figure 5a). The main effect of Placement was significant, $F(1.50, 51.04)$[3] = 101.05, $p < .001$, $\eta_p^2 = 0.75$. Planned comparisons of the estimated marginal means revealed that, as predicted, participants were significantly more accurate on direct ($M = 0.90$, $SE = 0.02$) than impeded trials ($M = 0.84$, $SE = .03$), $p = .01$, $d = 0.49$. Participants were also significantly more accurate on both direct, $p < .001$, $d = 2.13$, and impeded, $p < .001$, $d = 1.56$, trials than solo trials ($M = 0.47$, $SE = .03$). No other effects or interactions were significant.

**RT.** Following Kurtz and Gentner (2013), RTs that exceeded 60 s were excluded from the analysis. Six participants collectively had a total of 11 trials (out of 840) that met this criterion.

Participants' RTs were entered into a 3 (Placement: Direct, Impeded, Solo) × 2 (Axis: Horizontal, Vertical) repeated measures ANOVA (see Figure 5b). Five participants made no correct, nonexcluded responses on any horizontal and/or vertical solo items, and were excluded from the repeated measures ANOVA. This left 30 participants in this analysis. A main effect of Placement was significant, $F(1.73, 50.29) = 19.64$, $p < .001$, $\eta_p^2 = 0.40$. Planned comparisons of the estimated marginal means revealed that participants were equally fast on direct ($M = 11{,}645$, $SE = 797$) and impeded trials ($M = 12{,}310$, $SE = 806$), $p = .56$, $d = 0.11$; however, both direct, $p < .001$, $d = 1.24$, and impeded trials, $p < .001$, $d = 0.99$, were significantly slower than solo trials ($M = 6{,}718$, $SE = 578$). No other effects or interactions were significant.

Because including the solo condition in the analyses meant excluding some participants, we also ran a repeated measures ANOVA on RTs excluding that condition (but including all 35 participants), resulting in a 2 (Placement: Direct, Impeded) × 2 (Axis: Horizontal, Vertical) design. As in the previous analysis, RTs to direct ($M = 11{,}968$, $SE = 765$) and impeded trials ($M = 12{,}383$, $SE = 746$) did not differ significantly, $F(1, 34) = 0.18$, $p = .68$, $\eta_p^2 = 0.01$. No other effects or interactions were significant.

### Prior Experience

We examined the relationship between prior experience and spatial placement effects in two ways. First, we expected larger Placement effects for students without relevant STEM course experience than for those with relevant experience. Second, we assumed that direct placement would be most important when students are comparing unfamiliar items. Thus, we predicted that the effect of placement would be larger for items in noncanonical orientation than for those in canonical (familiar) orientation.

**STEM Course History.** Thirty-four participants completed a survey of their high school and college STEM course history, either in a separate, in-class group testing session at the start of term (30) or at the end of the study session (4). Participants were coded as having relevant course experience if they took any college-level courses that would have involved skeletal structures (e.g., biology, biological anthropology, physiology) or if they listed a major that would require coursework involving skeletal structures (e.g., biology, premed). Out of 34, 14 participants met these criteria.

The relationship of course experience (with, without) to accuracy was analyzed in a mixed-measures ANOVA with Placement (direct, impeded, solo) and Axis (horizontal, vertical). The effect of Placement remained significant, $F(1.59, 50.74) = 92.57$, $p < .001$, $\eta_p^2 = 0.74$, but there were no significant effects or interactions with course experience. Overall, participants with relevant course experience ($M = 0.70$, $SE = 0.03$) fared no better than those without ($M = 0.75$, $SE = 0.02$), $F(1, 32) = 1.89$, $p = .18$, $\eta_p^2 = 0.06$, and the effect of Placement did not differ across groups, $F(1.59, 50.74) = 0.14$, $p = .82$, $\eta_p^2 = 0.00$.

**Canonical Orientation.** Out of 24, 20 stimuli items were identified as having a canonically horizontal (13) or vertical (seven) orientation. These items were coded as presented in a canonical or noncanonical orientation for each participant.[4]

Participants' mean accuracy on the 20 items was entered into a 3 (Placement: Direct, Impeded, Solo) × 2 (Orientation: Canonical, Noncanonical) repeated measures ANOVA (see Figure 5c). A main effect of Placement was significant, $F(1.60, 54.22) = 115.21$, $p < .001$, $\eta_p^2 = 0.77$, as was a main effect of Orientation, $F(1, 34) = 16.46$, $p < .001$, $\eta_p^2 = 0.33$, with higher accuracy on canonical ($M = 0.79$, $SE = 0.02$) than noncanonical items ($M = 0.67$, $SE = 0.03$).

Planned comparisons of the estimated marginal means between direct and impeded pairs were carried out separately for canonical and noncanonical items. For canonical items, there was no difference between direct ($M = 0.92$, $SE = 0.03$) and impeded pairs ($M = 0.91$, $SE = 0.03$), $p = .86$, $d = 0.01$. In contrast, for noncanonical items, participants were significantly more accurate on direct items ($M = 0.89$, $SE = 0.03$) than impeded items ($M = 0.79$, $SE = 0.04$), $p = .03$, $d = 0.37$. However, the interaction between Placement and Orientation was not significant, $F(2, 68) = 2.08$, $p = .13$, $\eta_p^2 = 0.06$.

### Spatial Skill

Visual comparison involves encoding and manipulating visuospatial representations, so we expected that performance on the

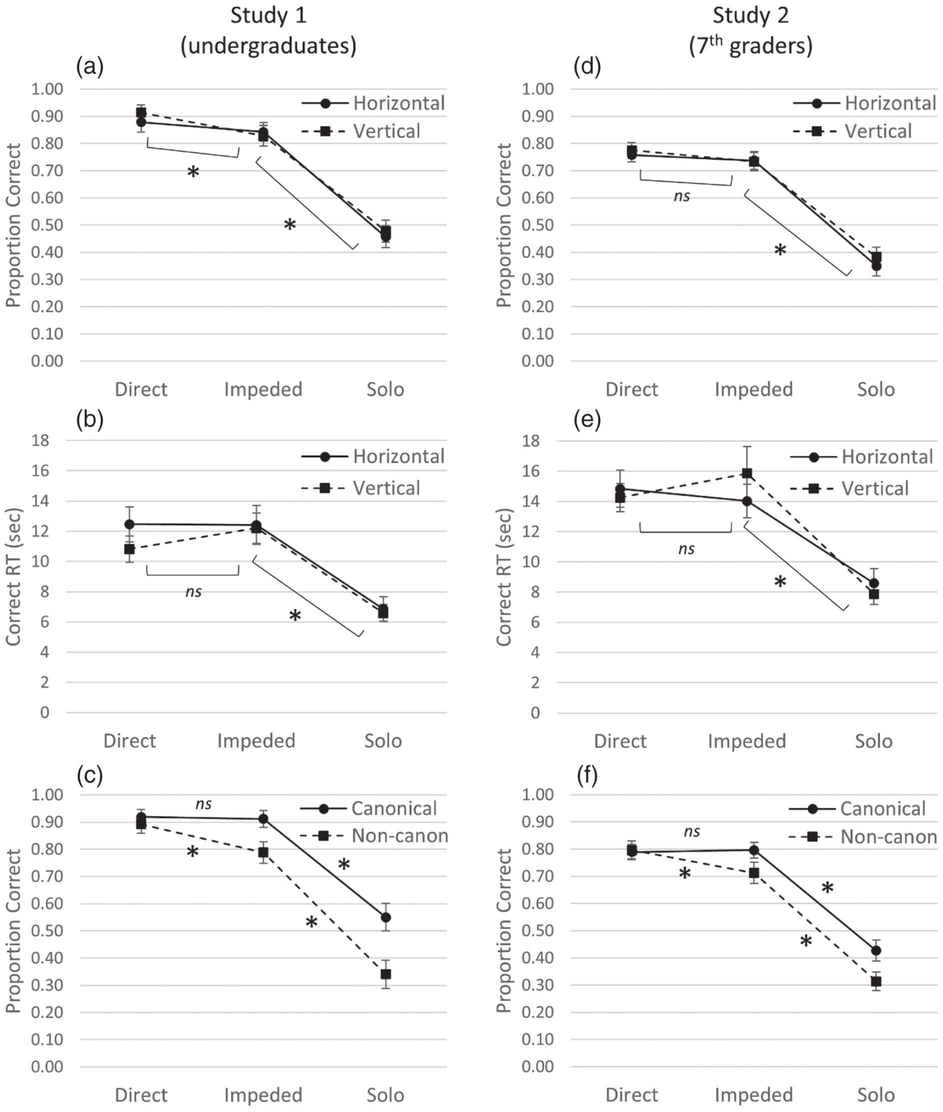
---

[3] In all cases, where the assumption of sphericity was not met, a Huynh–Feldt correction was applied.

[4] Because canonical orientation was not evenly balanced within the main design, different participants saw different numbers of items in canonical and noncanonical orientations, overall and across Placement conditions.

**Figure 5**

*Error-Detection Task Accuracy and Correct RTs for Studies 1 and 2*



*Note*. Square brackets indicate effects collapsing across conditions. Error bars depict within-subject standard errors. Graphs (a) and (d) include all items and all participants. Graphs (b) and (e) include all items but exclude participants who did not have any correct responses to one or more cells of the design. Graphs (c) and (f) include all participants but only a subset of 20 items for which there are clear expectations as to the normal (canonical) orientation. RT = reaction time.

* $p < .05$.

MRT would correlate with participants' overall performance on the error-detection task. We also asked whether spatial skills would moderate Placement effects, such that greater skill was associated with diminishing advantages of direct placement. To do so, we created individual "direct advantage scores" by subtracting mean accuracy on the impeded trials from accuracy on the direct trials and correlated this with participants' MRT performance.

Participants' MRT scores ($M = 10.66$, $SD = 4.93$, range = 3–23, total possible: 24) were not significantly related to their overall accuracy on the error detection test, $r = 0.23$, $p = .19$. However, on the subset of items with canonically horizontal or vertical orientations, MRT scores were marginally correlated with performance on trials where items were presented in a noncanonical orientation, $r = 0.29$, $p = .09$. MRT performance also was not significantly correlated with direct advantage scores, $r = -0.12$, $p = .47$.

## Discussion

The results of Study 1 suggest that the spatial alignment principle extends beyond simple, artificial stimuli to more complex, educationally relevant representations. Undergraduate students were better able to detect errors in skeletal structures when a correct reference was provided in direct placement than in impeded placement. This finding is noteworthy considering that the skeletons in each pair were nearly identical, with ample perceptual support for alignment. When other support is lacking, the effects of spatial alignment could potentially be even larger.

Participants were much less accurate when they were given a single skeleton than a pair, even when the pair was in impeded placement, highlighting the power of comparison to support processing visual representations. Although correct responses were also faster for single skeletons than for pairs, this is unlikely to reflect a simple speed–accuracy trade-off—on average, participants were slower on incorrect ($M = 9{,}791$ ms) than correct ($M = 7{,}119$ ms) trials. Rather, this might stem from having one, rather than two, images to process on solo trials.

Importantly, we did not find a pattern of RTs consistent with a speed–accuracy trade-off for our key comparison between direct and impeded placement, since accuracy, but not RT, differed between these conditions. Although Matlen et al. (2020) found RT differences, responses in the present study were much slower in general, and the stimuli and task were both more complex than in Matlen et al. (2020). Kurtz and Gentner (2013), who used the same stimuli and task to explore a different facet of alignability, also did not find RT differences.

The results also provide some insight into the role of prior experience in spatial alignment. The advantage of direct placement over impeded placement was especially evident for stimuli in noncanonical orientations, suggesting that direct placement may be an effective way to support visual comparison when the figures being compared are unfamiliar or hard to make sense of. (Consistent with the idea that noncanonical figures were more demanding, participants' MRT scores were marginally correlated with accuracy on noncanonical, but not canonical, trials—potentially because participants may have been compelled to engage mental rotation processes in order to interpret the noncanonical figures; Tarr & Pinker, 1989.) In fact, direct placement fully compensated for the decrements in accuracy on noncanonical trials seen in the impeded and solo conditions (Figure 5c). Although prior experience within participants at the item level affected accuracy, we did not find that between-participant experience with skeletal structures—inferred from STEM course history—related to accuracy or placement effects. It is possible that this measure did not adequately capture individual differences in experience with skeletal structures and diagrams, and future work should assess prior experience more directly.

Finally, we neither found evidence that spatial skills, as measured by the MRT, mediated the effects of placement nor did we find that spatial skills related to overall accuracy on the error-detection task. One possibility for these null results is that while the MRT was meant to provide a measure of participants' overall spatial skill levels, mental rotation is not a skill specifically needed for the error-detection task. The compared figures were always displayed with axes in the same orientation and so aligning them did not require rotation (though as we noted above, participants may still have engaged in mental rotation when faced with noncanonically oriented stimuli). Spatial skills are multifaceted and support task performance variably based on the spatial demands of the task (Atit et al., 2020; Newcombe & Shipley, 2015). In the next study, we included additional measures to explore whether other potentially more relevant facets of spatial skill were related to performance on the error-detection task.

# Study 2

Study 2 had two main aims. First, to determine whether the placement effects seen in undergraduates would extend to middle-school students. This has important practical and theoretical implications. Practically, it is crucial for understanding who might benefit from applying the spatial alignment principle in educational contexts. This may also help us to understand the developmental course of comparison processing. There is evidence from studies involving simple stimuli that children as young as 6 years old show spatial alignment effects (Zheng et al., 2022). Exploring the developmental trajectory of these effects for rich stimuli provides insight into the underlying processes and how they develop. Our second main aim was to more carefully explore the relationship between spatial skill and effects of placement by including additional spatial measures.

## Method

### *Transparency and Openness*

Research activities were approved by IRB at the first author's university and adhered to all ethical and local legal guidelines. This study was not preregistered. We have reported how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. All data and analytic code needed to reproduce analyses are available at https://osf.io/5hyn6/?view_only=85d1a6bf95d84d24a17764636fb9133b. Materials are available by request.

### *Participants*

Seventy-five seventh-grade students were recruited from an existing developmental participant database. Students received $20 for their participation. One participant was excluded for reporting diagnosed difficulties with visual perception. An additional 14 participants were excluded for poor performance on catch trials (see Study 1's Design and Procedure section), leaving 60 participants in the final sample ($M_{\text{age}} = 12$ years). Fifty-eight participants completed a demographics survey. This sample included 20 males and 38 females, was predominantly White (39 White, seven Asian, seven Hispanic or Latino, one not reported), and reported high parental educational attainment (36 graduate degrees, 18 bachelor's degrees, one General Educational Development Test/high school degree, one did not complete high school, four not reported). Eight students qualified for free or reduced lunch programs (41 did not, nine did not know).

**Sample Size.** Sample size was determined by a power analysis in G*Power 3.1.9.4 (Faul et al., 2007) for the comparison of interest with the smallest effect size in Study 1 (direct vs. impeded accuracy for noncanonical items), $d = 0.37$. To achieve 80% power at $\alpha = 0.05$, it was estimated that 60 participants would be needed.

### *Design and Procedure*

The design and methodology for the error-detection task were as in Study 1. The MRT was also administered as in Study 1. In addition, another spatial assessment, the Group Embedded Figures Test (EFT; Oltman et al., 1971), was added at the end of the procedure. The EFT requires participants to find simple shapes embedded within more complex images. It was selected because it mirrors the demands of the error-detection task: Holding in mind a visual representation of part of an image (e.g., a bone) to find and compare within another, complex image (e.g., the other skeleton). The task was administered with paper test booklets. On each problem, participants were instructed to look up one of eight simple shapes at the back of the test booklet before returning to the problem page to find and trace that shape within a larger, more complex geometric figure. The task included three timed sections with several minutes to rest between each. The first section served as practice and was not included in the scoring of the task. In each of the last two sections, participants were given 5 min to solve nine problems. Participants were told in advance that they might not finish all problems in each section within the time limit. Problems were awarded 1 point if the simple shape was correct and complete (all lines drawn in full, without deformation of shape), and no other lines or marks were present (no additional lines or lines extending past the shape boundaries). Problems that did not meet these criteria, including those that were not started or attempted, received 0 points. Points were summed for a total score out of a possible 18. Three participants did not complete the EFT.

## Results

### Placement

As in Study 1, the effects of placement were assessed using participants' accuracy and reaction time for correct responses (RT) on the error-detection task.

**Accuracy.** Participants' mean accuracy was entered into a 3 (Placement: Direct, Impeded, Solo) × 2 (Axis: Horizontal, Vertical) repeated measures ANOVA (Figure 5d). A main effect of Placement was significant, $F(1.66, 97.91) = 100.76$, $p < .001$, $\eta_p^2 = 0.63$. Planned comparisons of the estimated marginal means revealed that as in Study 1, solo trials ($M = 0.37$, $SE = 0.30$) were significantly less accurate than direct ($M = 0.77$, $SE = 0.02$), $p < .001$, $d = 1.52$, and impeded trials ($M = 0.74$, $SE = 0.03$), $p < .001$, $d = 1.32$. However, unlike Study 1, accuracy did not differ significantly between direct and impeded trials, $p = .17$, $d = 0.17$. No other effects or interactions were significant.

**RT.** RTs that exceeded 60 s were excluded from the analysis. Fifteen participants collectively had a total of 31 trials (out of 897) that met this criterion.

Participants' RTs were entered into a 3 (Placement: Direct, Impeded, Solo) × 2 (Axis: Horizontal, Vertical) repeated measures ANOVA (Figure 5e). Twenty-seven participants made no correct, nonexcluded responses on any horizontal and/or any vertical items in one or more of the Placement conditions, and were excluded from the repeated measures ANOVA, leaving 33 participants in this analysis. A main effect of Placement was significant, $F(2, 64) = 19.37$, $p < .001$, $\eta_p^2 = 0.38$. Planned comparisons of the estimated marginal means revealed that while RTs to solo items ($M = 8{,}228$, $SE = 715$) were significantly faster than to direct ($M = 14{,}544$, $SE = 861$), $p < .001$, $d = 0.70$, or impeded items ($M = 14{,}943$, $SE = 1{,}162$), $p < .001$, $d = 0.83$, RTs to direct and impeded items did not differ, $p = .78$, $d = 0.04$.

Including the solo condition in the previous analysis meant that almost half of the participants were excluded. To include more participants, and to narrow in on the key comparison between direct and impeded placement, we conducted a second analysis including only the direct and impeded conditions. Three participants were excluded from the analysis for lacking any correct, nonexcluded responses on any horizontal and/or any vertical items in the direct or impeded conditions. This left 57 participants in the 2 (Placement: Direct, Impeded) × 2 (Axis: Horizontal, Vertical) repeated measures ANOVA. However, Placement was not significant, $F(1, 56) = 1.04$, $p = 31$, $\eta_p^2 = 0.02$. Direct RTs ($M = 14{,}556$, $SE = 689$) were no faster than impeded RTs ($M = 15{,}679$, $SE = 937$). No other effects or interactions were significant.

### Prior Experience

We examined the relationship between prior experience and spatial placement effects by comparing accuracy when figures were in canonical versus noncanonical orientations. We expected that, as in Study 1, the effect of Placement would be larger when the stimuli were unfamiliar (noncanonical) than when they were familiar (canonical).

**Canonical Orientation.** As in Study 1, participants' mean accuracy on the 20 items with canonical orientations was entered into a 3 (Placement: Direct, Impeded, Solo) × 2 (Orientation: Canonical, Noncanonical) repeated measures ANOVA (see Figure 5f). A main effect of Placement was significant, $F(1.72, 101.74)^5 = 96.23$, $p < .001$, $\eta_p^2 = 0.62$, as was a main effect of Orientation, $F(1, 59) = 8.75$, $p = .004$, $\eta_p^2 = 0.13$. The interaction between Placement and Orientation was marginally significant, $F(2, 118) = 2.72$, $p = .07$, $\eta_p^2 = 0.04$. Planned comparisons of the estimated marginal means between direct and impeded items were carried out separately for the canonical and noncanonical conditions. As found in Study 1, there was no difference between direct ($M = 0.79$, $SE = 0.03$) and impeded placement ($M = 0.80$, $SE = 0.03$) for canonical items, $p = .84$, $d = 0.03$. However, also as found in Study 1, accuracy was significantly higher for direct ($M = 0.80$, $SE = 0.03$) than impeded placement ($M = 0.71$, $SE = 0.04$) for noncanonical items, $p = .03$, $d = 0.29$.

### Spatial Skill

We explored the relationship between spatial skills and overall performance on the error-detection task, as well as individuals' direct advantage scores, using two spatial tasks: the MRT and EFT. Scores on the two spatial tasks—MRT ($M = 8.25$, $SD = 4.73$, range = 0–21, total possible: 24) and EFT ($M = 8.98$, $SD = 5.18$, range = 0–18, total possible: 18)—were significantly correlated, $r = 0.38$, $p = .003$.

**Mental Rotations Test.** The MRT did not correlate with participants' direct advantage scores, $r = -0.15$, $p = .27$. However, it did correlate significantly with overall performance on the error-detection task, $r = 0.40$, $p = .001$. Among the 20 items with canonically horizontal or vertical orientations, there was a stronger overall correlation for items presented in their noncanonical orientation, $r = 0.45$, $p < .001$, than in their canonical orientation, $r = 0.25$, $p = .05$.

**Embedded Figures Test.** EFT performance did not correlate significantly with either direct advantage scores, $r = -0.07$, $p = .59$, or overall performance, $r = 0.21$, $p = .12$. However, for the subset of items with canonical orientations, it correlated with accuracy on trials with noncanonical presentations, $r = 0.41$, $p = .03$.

---

[5] Because the assumption of sphericity was not met, a Huynh–Feldt correction was applied.

## Discussion

Overall, the middle-school students were less accurate (62% vs. 73%) and slower (16.2 vs. 11.2 s) than the undergraduates from Study 1. Nonetheless, both groups showed qualitatively similar error-detection patterns, in line with the spatial alignment principle. Unlike the adults in Study 1, seventh graders' overall accuracy for direct and impeded placement did not differ significantly in Study 2. However, seventh graders—like adults—were more accurate on the difficult noncanonical trials when the pair was in direct compared to impeded placement. That is, both middle-school and undergraduate students benefited from direct placement when the items did not follow their prior experience and were more difficult to identify visually.

Like Study 1, this study did not find any evidence that spatial placement effects were mediated by spatial skills, as neither MRT performance nor EFT performance was correlated with individual direct advantage scores.

However, in contrast to the adult results in Study 1, this study found that seventh graders' performance on the MRT was significantly correlated with overall error-detection accuracy, consistent with previous findings that spatial skills support the processing of visual representations (Hambrick et al., 2012; Höffler, 2010). For both adults and seventh graders, the results suggest that spatial skills were most related to performance on noncanonical items.[6] In Study 1, adults' MRT performance was marginally correlated with accuracy on noncanonical trials only, and in Study 2, seventh graders' performance on both the MRT and EFT was more strongly related to noncanonical than to canonical trial accuracy. This is consistent with the idea that spatial skills may have helped participants compensate for the additional difficulty posed by the noncanonical orientations (Höffler & Leutner, 2011). The association with EFT performance is notable because the EFT does not involve any changes in orientation, suggesting that students' ability to make sense of the figures may have been disrupted in multiple ways by noncanonical presentation.

It is interesting that seventh graders, but not adults, showed a significant correlation between MRT performance and performance on the error-detection task. One possible explanation is that the two groups may have been using different strategies, either on the error-detection task, the MRT, or both (Janssen & Geiser, 2010). Spatial skills are associated with successful deployment of spatial-visualization strategies (e.g., mentally imagining figures rotating), but the use of such strategies diminishes with growing experience and expertise with other (e.g., analytical) strategies (Hambrick et al., 2012; Schwartz & Black, 1996; Stieff, 2007; Stieff et al., 2012). Consistent with this, the younger students in Study 2 may have been more likely to use spatial-visualization strategies and thus may have shown a greater association between spatial skill and task performance.

# General Discussion

Our results provide evidence that the spatial alignment principle (Matlen et al., 2020) applies to rich, educationally relevant materials. Both undergraduate and middle-school students showed an advantage of direct over impeded placement when comparing skeletal structures depicted in unfamiliar, noncanonical orientations. No such benefit was seen when the figures were in familiar, canonical orientations—suggesting that prior familiarity with the materials moderated the effects of placement, with greater benefits of direct placement when figures were unfamiliar or hard to interpret. This interpretation is consistent with prior findings that students need more support for visual comparison when they lack experience with the materials (Braithwaite & Goldstone, 2015; Gentner et al., 2011; Guo & Pang, 2011; Jee & Anggoro, 2019; Mix, 2008; Rittle-Johnson et al., 2009; Rittle-Johnson & Star, 2009). We suggest that for familiar figures, learners were able to quickly form accurate, stable representations (Palmer et al., 1981), and that this facilitated comparisons, regardless of spatial layout. In contrast, for figures in noncanonical orientations, direct placement made it easier to align the two figures, despite their lack of familiarity. In fact, direct placement fully compensated for the costs imposed by noncanonical presentation.

Although more work is needed to understand these effects, there is a potentially analogous pattern found in Matlen et al. (2020), using simple sequences such as ABA. They found that impeded placement was much less disruptive for horizontal sequences than for vertical sequences. They speculated that their participants were likely to be highly fluent at processing horizontal sequences (possibly due to extensive reading or other experiences); thus, they were able to encode the sequences quickly, rendering them more resilient to disruption from impeded placement (see also Zheng et al., 2022).

Our pattern of results held regardless of prior STEM experience and across different levels of spatial skill. In particular, spatial skills did not correlate with differences in participants' accuracy between direct and impeded placement. However, spatial skills, in particular for the seventh graders, were related to overall performance on the error-detection task, consistent with prior work showing links between spatial skill and successful reasoning with visual representations (Höffler, 2010). This relationship was stronger for noncanonical than canonical items, again suggesting that spatial skills may have helped to compensate for the difficulty posed by the unfamiliar presentation (Höffler & Leutner, 2011). Given the evidence for relationships between prior experience, spatial strategy use, and visual processing fluency (Höffler, 2010; Rau, 2017; Stieff et al., 2012), further studies should explore whether and how the relationship between spatial skill and placement effects changes with relevant experience.

## Educational Implications

These findings have implications for STEM education. Comparing, integrating, and translating between visual representations are integral to STEM learning (e.g., Ainsworth, 2008; Gleicher et al., 2011; Rau, 2017). For example, Jee et al. (2022) found that more than a third of the figures in the evolution and anatomy units of middle-school life science textbooks involved comparison. Further, of the comparisons in those figures, fewer than half were in direct placement. The findings in the present work suggest that the placement of paired visual representations can influence students' ability to identify important differences that are connected to the relational structure. Thus, there is significant potential to apply the

---

[6] We note that familywise error rate may be inflated within Studies 1 and 2. However, we found parallel patterns of key findings across both studies, namely, the accuracy advantage of direct over impeded placement for noncanonical figures and the stronger relationship between spatial skills and accuracy on noncanonical figures, which bolsters our confidence in these results.

spatial alignment principle in textbooks and other educational resources.

We note that the stimuli used in these studies provided a stringent test of the role of placement. The skeletons in each pair were nearly identical, apart from the target error, providing significant perceptual support for the alignment. Nonetheless, spatial placement improved comparison beyond these perceptual supports, suggesting it offered unique and additive benefits. In noncanonical orientations, performance on direct trials was improved by about 10% relative to impeded trials. For a student trying to make sense of textbook figures, this might mean the difference between readily seeing the point of a figure and drawing relevant insights versus deciding to move on without understanding. In comparisons lacking perceptual similarity, placement has the potential to produce even larger effects. This is encouraging because many visual comparisons students must make involve figures that are dissimilar (e.g., multiple representations; Ainsworth, 2008).

According to the spatial alignment principle, direct placement facilitates the alignment of systematic correspondences and minimizes distraction from competing, incorrect matches. Thus, like other factors proposed to support coordination of visual representations, direct placement supports necessary processes while reducing extraneous processing (Mayer, 2017; Rau, 2017; Sweller, 2011). For example, Mayer's spatial contiguity principle recommends placing text near corresponding parts of a graphic to ensure that students will make connections between them, without the added burden of scanning across a display (Mayer, 2017). Likewise, alignment can be supported in a number of ways, such as by avoiding inconsistent lower order matches that compete with higher order systematic correspondences (Gentner & Toupin, 1986; Loewenstein & Gentner, 2005; Markman & Gentner, 1993; Richland & McDonough, 2010), by highlighting common structure through language (Gentner & Toupin, 1986; Goldwater & Gentner, 2015; Loewenstein & Gentner, 2005), or by gesturing between corresponding objects (Gentner et al., 2011; Jee & Anggoro, 2019; Richland & Simms, 2015; Vendetti et al., 2015; Yuan et al., 2017). These all have in common reducing extraneous obstacles to fruitful comparison.

In conclusion, the spatial alignment principle applies to rich, educationally relevant figures. Applying this principle has the potential to improve visual comparisons in STEM education.

## References

Ainsworth, S. (2008). The educational value of multiple-representations when learning complex scientific concepts. In J. K. Gilbert, M. Reiner, & M. Nakhleh (Eds.), *Visualization: Theory and practice in science education* (pp. 191–208). Springer. https://doi.org/10.1007/978-1-4020-5267-5_9

Arcavi, A. (2003). The role of visual representations in the learning of mathematics. *Educational Studies in Mathematics*, *52*(3), 215–241. https://doi.org/10.1023/A:1024312321077

Atit, K., Uttal, D. H., & Stieff, M. (2020). Situating space: Using a discipline-focused lens to examine spatial thinking skills. *Cognitive Research: Principles and Implications*, *5*(1), Article 19. https://doi.org/10.1186/s41235-020-00210-z

Begolli, K. N., & Richland, L. E. (2016). Teaching mathematics by comparison: Analog visibility as a double-edged sword. *Journal of Educational Psychology*, *108*(2), 194–213. https://doi.org/10.1037/edu0000056

Blanz, V., Tarr, M. J., & Bülthoff, H. H. (1999). What object attributes determine canonical views? *Perception*, *28*(5), 575–599. https://doi.org/10.1068/p2897

Braithwaite, D. W., & Goldstone, R. L. (2015). Effects of variation and prior knowledge on abstract concept learning. *Cognition and Instruction*, *33*(3), 226–256. https://doi.org/10.1080/07370008.2015.1067215

Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction*, *20*(2), 155–166. https://doi.org/10.1016/j.learninstruc.2009.02.014

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, *70*(5), 1098–1120. https://doi.org/10.1111/1467-8624.00081

Christie, S., & Gentner, D. (2010). Where hypotheses come from: Learning new relations by structural alignment. *Journal of Cognition and Development*, *11*(3), 356–373. https://doi.org/10.1080/15248371003700015

Cook, M. P. (2006). Visual representations in science education: The influence of prior knowledge and cognitive load theory on instructional design principles. *Science Education*, *90*(6), 1073–1091. https://doi.org/10.1002/sce.20164

Eilam, B., & Gilbert, J. K. (2014). The significance of visual representations in the teaching of science. In B. Eilam & J. K. Gilbert (Eds.), *Science teachers' use of visual representations* (pp. 3–28). Springer. https://doi.org/10.1007/978-3-319-06526-7_1

Evagorou, M., Erduran, S., & Mäntylä, T. (2015). The role of visual representations in scientific practices: From conceptual understanding and knowledge generation to 'seeing' how science works. *International Journal of STEM Education*, *2*(1), Article 11. https://doi.org/10.1186/s40594-015-0024-x

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence*, *41*(1), 1–63. https://doi.org/10.1016/0004-3702(89)90077-5

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Forbus, K. D., Ferguson, R. W., Lovett, A., & Gentner, D. (2017). Extending SME to handle large-scale cognitive modeling. *Cognitive Science*, *41*(5), 1152–1201. https://doi.org/10.1111/cogs.12377

Gadgil, S., Nokes-Malach, T. J., & Chi, M. T. H. (2012). Effectiveness of holistic mental model confrontation in driving conceptual change. *Learning and Instruction*, *22*(1), 47–61. https://doi.org/10.1016/j.learninstruc.2011.06.002

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

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

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

Gentner, D., & Colhoun, J. (2010). Analogical processes in human thinking and learning. In B. Glatzeder, V. Goel, & A. Müller (Eds.), *Towards a theory of thinking* (pp. 35–48). Springer. https://doi.org/10.1007/978-3-642-03129-8_3

Gentner, D., & Forbus, K. D. (2011). Computational models of analogy. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(3), 266–276. https://doi.org/10.1002/wcs.105

Gentner, D., & Gunn, V. (2001). Structural alignment facilitates the noticing of differences. *Memory & Cognition*, *29*(4), 565–577. https://doi.org/10.3758/BF03200458

Gentner, D., Levine, S. C., Ping, R., Isaia, A., Dhillon, S., Bradley, C., & Honke, G. (2016). Rapid learning in a children's museum via analogical comparison. *Cognitive Science*, *40*(1), 224–240. https://doi.org/10.1111/cogs.12248

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

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, *52*(1), 45–56. https://doi.org/10.1037/0003-066X.52.1.45

Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science*, *10*(3), 277–300. https://doi.org/10.1207/s15516709cog1003_2

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*(1), 1–38. https://doi.org/10.1016/0010-0285(83)90002-6

Gleicher, M., Albers, D., Walker, R., Jusufi, I., Hansen, C. D., & Roberts, J. C. (2011). Visual comparison for information visualization. *Information Visualization*, *10*(4), 289–309. https://doi.org/10.1177/1473871611416549

Goldwater, M. B., & Gentner, D. (2015). On the acquisition of abstract knowledge: Structural alignment and explication in learning causal system categories. *Cognition*, *137*, 137–153. https://doi.org/10.1016/j.cognition.2014.12.001

Guillot, A., Champely, S., Batier, C., Thiriet, P., & Collet, C. (2007). Relationship between spatial abilities, mental rotation and functional anatomy learning. *Advances in Health Sciences Education*, *12*(4), 491–507. https://doi.org/10.1007/s10459-006-9021-7

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

Hambrick, D. Z., Libarkin, J. C., Petcovic, H. L., Baker, K. M., Elkins, J., Callahan, C. N., Turner, S. P., Rench, T. A., & Ladue, N. D. (2012). A test of the circumvention-of-limits hypothesis in scientific problem solving: The case of geological bedrock mapping. *Journal of Experimental Psychology: General*, *141*(3), 397–403. https://doi.org/10.1037/a0025927

Höffler, T. N. (2010). Spatial ability: Its influence on learning with visualizations—A meta-analytic review. *Educational Psychology Review*, *22*(3), 245–269. https://doi.org/10.1007/s10648-010-9126-7

Höffler, T. N., & Leutner, D. (2011). The role of spatial ability in learning from instructional animations—Evidence for an ability-as-compensator hypothesis. *Computers in Human Behavior*, *27*(1), 209–216. https://doi.org/10.1016/j.chb.2010.07.042

Hoyos, C., & Gentner, D. (2017). Generating explanations via analogical comparison. *Psychonomic Bulletin & Review*, *24*(5), 1364–1374. https://doi.org/10.3758/s13423-017-1289-5

Hribar, A., Haun, D. B. M., & Call, J. (2012). Children's reasoning about spatial relational similarity: The effect of alignment and relational complexity. *Journal of Experimental Child Psychology*, *111*(3), 490–500. https://doi.org/10.1016/j.jecp.2011.11.004

Janssen, A. B., & Geiser, C. (2010). On the relationship between solution strategies in two mental rotation tasks. *Learning and Individual Differences*, *20*(5), 473–478. https://doi.org/10.1016/j.lindif.2010.03.002

Jee, B. D., & Anggoro, F. K. (2019). Relational scaffolding enhances children's understanding of scientific models. *Psychological Science*, *30*(9), 1287–1302. https://doi.org/10.1177/0956797619864601

Jee, B. D., Matlen, B. J., Greenlaw, M., Simms, N., & Gentner, D. (2022). Spatial supports for comparison in educational science images. *Instructional Science*, *50*(6), 807–827. https://doi.org/10.1007/s11251-022-09599-0

Jee, B. D., Uttal, D. H., Gentner, D., Manduca, C., Shipley, T. F., & Sageman, B. (2013). Finding faults: Analogical comparison supports spatial concept learning in geoscience. *Cognitive Processing*, *14*(2), 175–187. https://doi.org/10.1007/s10339-013-0551-7

Kok, E. M., de Bruin, A. B. H., Robben, S. G. F., & van Merriënboer, J. J. G. (2013). Learning radiological appearances of diseases: Does comparison help? *Learning and Instruction*, *23*, 90–97. https://doi.org/10.1016/j.learninstruc.2012.07.004

Kurtz, K. J., & Gentner, D. (2013). Detecting anomalous features in complex stimuli: The role of structured comparison. *Journal of Experimental Psychology: Applied*, *19*(3), 219–232. https://doi.org/10.1037/a0034395

Kurtz, K. J., Miao, C.-H., & Gentner, D. (2001). Learning by analogical bootstrapping. *Journal of the Learning Sciences*, *10*(4), 417–446. https://doi.org/10.1207/S15327809JLS1004new_2

Lassaline, M. E. (1996). Structural alignment in induction and similarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(3), 754–770. https://doi.org/10.1037/0278-7393.22.3.754

Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology*, *50*(4), 315–353. https://doi.org/10.1016/j.cogpsych.2004.09.004

Markman, A. B. (1997). Constraints on analogical inference. *Cognitive Science*, *21*(4), 373–418. https://doi.org/10.1207/s15516709cog2104_1

Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology*, *25*(4), 431–467. https://doi.org/10.1006/cogp.1993.1011

Markman, A. B., & Gentner, D. (1996). Commonalities and differences in similarity comparisons. *Memory & Cognition*, *24*(2), 235–249. https://doi.org/10.3758/BF03200884

Matlen, B. J., Gentner, D., & Franconeri, S. L. (2020). Spatial alignment facilitates visual comparison. *Journal of Experimental Psychology: Human Perception and Performance*, *46*(5), 443–457. https://doi.org/10.1037/xhp0000726

Mayer, R. E. (2017). Using multimedia for e-learning. *Journal of Computer Assisted Learning*, *33*(5), 403–423. https://doi.org/10.1111/jcal.12197

Mayer, R. E. (2021). *Multimedia learning* (3rd ed.). Cambridge University Press.

Miller, B. W., Cromley, J. G., & Newcombe, N. S. (2016). Improving diagrammatic reasoning in middle school science using conventions of diagrams instruction. *Journal of Computer Assisted Learning*, *32*(4), 374–390. https://doi.org/10.1111/jcal.12143

Mix, K. S. (2008). Surface similarity and label knowledge impact early numerical comparisons. *British Journal of Developmental Psychology*, *26*(1), 13–32. https://doi.org/10.1348/026151007X189109

Moore, T. J., Miller, R. L., Lesh, R. A., Stohlmann, M. S., & Kim, Y. R. (2013). Modeling in engineering: The role of representational fluency in students' conceptual understanding. *Journal of Engineering Education*, *102*(1), 141–178. https://doi.org/10.1002/jee.20004

Newcombe, N. S., & Shipley, T. F. (2015). Thinking about spatial thinking: New typology. new assessments. In J. S. Gero (Ed.), *Studying visual and spatial reasoning for design creativity* (pp. 179–192). Springer. https://doi.org/10.1007/978-94-017-9297-4_10

Oltman, P. K., Raskin, E., & Witkin, H. A. (1971). *Group embedded figures test*. Consulting Psychologists Press.

Paik, J. H., & Mix, K. S. (2006). Preschoolers' use of surface similarity in object comparisons: Taking context into account. *Journal of Experimental Child Psychology*, *95*(3), 194–214. https://doi.org/10.1016/j.jecp.2006.06.002

Paik, J. H., & Mix, K. S. (2008). It's all relative: Different levels of relational similarity used in children's comparisons. *British Journal of Developmental Psychology*, *26*(4), 499–505. https://doi.org/10.1348/026151007X260163

Palmer, S., Rosch, E., & Chase, P. (1981). Canonical perspective and the perception of objects. In J. Long & A. Baddeley (Eds.), *Attention and performance IX* (pp. 135–151). Lawrence Erlbaum.

Parsons, P., & Sedig, K. (2014). Adjustable properties of visual representations: Improving the quality of human-information interaction. *Journal of the Association for Information Science and Technology*, *65*(3), 455–482. https://doi.org/10.1002/asi.23002

Peters, M. (1995). *Revised Vandenberg & Kuse Mental Rotations Tests: Forms MRT-A to MRT-D* [Technical report]. Department of Psychology, University of Guelph.

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

Resnick, I., Davatzes, A., Newcombe, N. S., & Shipley, T. F. (2017). Using analogy to learn about phenomena at scales outside human perception. *Cognitive Research: Principles and Implications*, *2*(1), Article 21. https://doi.org/10.1186/s41235-017-0054-7

Richland, L. E., & McDonough, I. M. (2010). Learning by analogy: Discriminating between potential analogs. *Contemporary Educational Psychology*, *35*(1), 28–43. https://doi.org/10.1016/j.cedpsych.2009.09.001

Richland, L. E., & Simms, N. (2015). Analogy, higher order thinking, and education. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(2), 177–192. https://doi.org/10.1002/wcs.1336

Rittle-Johnson, B., & Star, J. (2009). Compared with what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. *Journal of Educational Psychology*, *101*(3), 529–544. https://doi.org/10.1037/a0014224

Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology*, *99*(3), 561–574. https://doi.org/10.1037/0022-0663.99.3.561

Rittle-Johnson, B., Star, J. R., & Durkin, K. (2009). The importance of prior knowledge when comparing examples: Influences on conceptual and procedural knowledge of equation solving. *Journal of Educational Psychology*, *101*(4), 836–852. https://doi.org/10.1037/a0016026

Ross, B. H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(4), 629–639. https://doi.org/10.1037/0278-7393.13.4.629

Sagi, E., Gentner, D., & Lovett, A. (2012). What difference reveals about similarity. *Cognitive Science*, *36*(6), 1019–1050. https://doi.org/10.1111/j.1551-6709.2012.01250.x

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, *13*(2), 141–156. https://doi.org/10.1016/S0959-4752(02)00017-8

Schnotz, W., & Wagner, I. (2018). Construction and elaboration of mental models through strategic conjoint processing of text and pictures. *Journal of Educational Psychology*, *110*(6), 850–863. https://doi.org/10.1037/edu0000246

Schönborn, K. J., & Anderson, T. R. (2006). The importance of visual literacy in the education of biochemists*. *Biochemistry and Molecular Biology Education*, *34*(2), 94–102. https://doi.org/10.1002/bmb.2006.49403402094

Schwartz, D. L., & Black, J. B. (1996). Shuttling between depictive models and abstract rules: Induction and fallback. *Cognitive Science*, *20*(4), 457–497. https://doi.org/10.1207/s15516709cog2004_1

Shah, P., & Freedman, E. G. (2011). Bar and line graph comprehension: An interaction of top-down and bottom-up processes. *Topics in Cognitive Science*, *3*(3), 560–578. https://doi.org/10.1111/j.1756-8765.2009.01066.x

Simms, N. (2023, February 17). *Spatial alignment bones*. https://osf.io/5hyn6

Stieff, M. (2007). Mental rotation and diagrammatic reasoning in science. *Learning and Instruction*, *17*(2), 219–234. https://doi.org/10.1016/j.learninstruc.2007.01.012

Stieff, M., Ryu, M., Dixon, B., & Hegarty, M. (2012). The role of spatial ability and strategy preference for spatial problem solving in organic chemistry. *Journal of Chemical Education*, *89*(7), 854–859. https://doi.org/10.1021/ed200071d

Sweller, J. (2011). CHAPTER TWO—Cognitive load theory. In J. P. Mestre & B. H. Ross (Eds.), *Psychology of learning and motivation* (Vol. 55, pp. 37–76). Academic Press. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, *21*(2), 233–282. https://doi.org/10.1016/0010-0285(89)90009-1

Vendetti, M. S., Matlen, B. J., Richland, L. E., & Bunge, S. A. (2015). Analogical reasoning in the classroom: Insights from cognitive science. *Mind, Brain and Education*, *9*(2), 100–106. https://doi.org/10.1111/mbe.12080

Wickelgren, W. A., Corbett, A. T., & Dosher, B. A. (1980). Priming and retrieval from short-term memory: A speed accuracy trade-off analysis. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 387–404. https://doi.org/10.1016/S0022-5371(80)90276-5

Yuan, L., Uttal, D., & Gentner, D. (2017). Analogical processes in children's understanding of spatial representations. *Developmental Psychology*, *53*(6), 1098–1114. https://doi.org/10.1037/dev0000302

Zheng, Y., Matlen, B., & Gentner, D. (2022). Spatial alignment facilitates visual comparison in children. *Cognitive Science*, *46*(8), Article e13182. https://doi.org/10.1111/cogs.13182

Received May 2, 2022
Revision received November 14, 2022
Accepted February 5, 2023 ■