Drawing on Experience: How Domain Knowledge Is Reflected in Sketches of Scientific Structures and Processes

**Benjamin D. Jee · Dedre Gentner · David H. Uttal ·
Bradley Sageman · Kenneth Forbus ·
Cathryn A. Manduca · Carol J. Ormand ·
Thomas F. Shipley · Basil Tikoff**

Abstract Capturing the nature of students' mental representations and how they change with learning is a primary goal in science education research. This can be challenging in spatially intense domains, such as geoscience, architecture, and engineering. In this research, we test whether sketching can be used to gauge level of expertise in geoscience, using new technology designed to facilitate this process. We asked participants with differing levels of geoscience experience to copy two kinds of geoscience images—photographs of rock formations and causal diagrams. To permit studying the process of sketching as well as the structure and content of the sketches, we used the CogSketch system (Forbus et al. 2011, Topics in Cognitive Science 3:648–666) to record the time course of sketching and analyze the sketches themselves. Relative to novices, geoscience students included more geological structures and relational symbols in their sketches of geoscience materials and were more likely to construct their sketches in a sequence consistent with the order of causal events. These differences appear to stem from differences in domain knowledge, because they did not show up in participants' sketches of materials from other fields. The findings and methods of this research

D. Gentner · D. H. Uttal · B. Sageman · K. Forbus
Northwestern University, Evanston, IL 60208, USA

C. A. Manduca · C. J. Ormand
Carleton College, Northfield, MN 55057, USA

T. F. Shipley
Temple University, Philadelphia, PA 19122, USA

B. Tikoff
University of Wisconsin-Madison, Madison, WI 53706, USA

B. D. Jee (✉)
Department of Psychology, Rhode Island College, 600 Mount Pleasant Avenue, Providence, RI 02908, USA
e-mail: bjee@ric.edu

suggest new ways to promote and assess science learning, which are well suited to the visual–spatial demands of many domains.

A crucial concern for researchers and educators alike is how to study the development of people's mental representations as they acquire domain knowledge. This challenge can be particularly great for domains that are intensely spatial, such as geoscience, engineering, and other STEM disciplines. Much of the content in these domains is expressed through diagrams, maps, and other visual representations. Traditional written assessments, including multiple-choice tests, are of limited value for highly spatial topics because of the difficulty of conveying complex spatial relations through speech or writing alone (e.g., Larkin and Simon 1987). Sketches, by contrast, readily capture spatial information (Goel 1995; Heiser and Tversky 2002; Suwa and Tversky 1997; Vosniadou and Brewer 1992). Sketches can also reveal students' conceptual representations, a point that is amplified below. All this suggests that sketching is crucial to capturing and assessing students' knowledge in STEM disciplines, such as geoscience.

Indeed, in a recent *Science* article, Ainsworth et al. (2011) go further, stating that sketching is vital to science learning. They maintain that sketching helps students learn to represent and reason about scientific phenomena, and that "Unlike other constructive strategies, such as writing summaries or providing oral self-explanations, visual representations have distinct attributes that match the visual-spatial demands of much of science learning" (p. 1097). Based on a survey of the literature on drawing during learning, Van Meter and Garner (2005) argued that learner-generated drawing shows particularly strong effects in higher-order assessments, such as transfer. Schwanbörn et al. (2010), for example, provide evidence that drawing during learning leads 9th-grade students to improve on tests of transfer in chemistry. Recent work by Gagnier et al. (2014, accepted for publication) suggests that sketching improves a learner's comprehension of diagrams, including inferences about the internal structure of a 3D object. These findings tell us that sketching can be important in learning and assessment, but they do not focus on what properties of sketches might be measured to aid in assessment. This paper provides evidence relevant to that question.

Ainsworth et al. (2011) also point out that sketching is currently underutilized in science instruction. It is not hard to see why—grading sketches from a large class can be extremely time-consuming. It is also a challenge to score drawings reliably, because they are often quite variable (e.g., White and Gunstone 1992). This not only places a burden on the instructor but also results in delayed (and therefore less useful) feedback to the student (Dihoff et al. 2004; Kulik and Kulik 1988). Recent innovations in computer-based sketching have made the task of collecting and assessing sketches substantially easier (de Silva et al. 2007; Do 2005; Forbus et al. 2011; Pargas et al. 2007; Valentine et al. 2012). These developments could have great impact on science education research and practice. In this paper, we describe findings from research using *CogSketch*, a computer-based sketching platform (Forbus et al. 2008, 2011), to investigate the effects of spatial scientific knowledge on sketching. Our focus here is on geoscience—an intensely spatial domain in which visualization, modeling, and sketching are fundamental to instruction (Jee et al. 2010; Kastens et al. 2009; Libarkin and Brick 2002; Liben and Titus 2012; Piburn et al. 2011). However, we believe our findings are relevant to spatially intensive fields in general.

This research has two linked goals. First, we aim to discover whether and how geoscience understanding is reflected in sketches and in the process through which sketches are

constructed. This is especially relevant to theories that aim to capture and explain changes in knowledge that result from science learning. To this end, we elicited students' sketches of two distinct kinds of geoscience materials—photographs of geological formations, and causal diagrams. We collected sketches from both novices (psychology students) and geoscience students. Our second goal is methodological: to explore the utility of CogSketch software (Forbus et al. 2008, 2011) for collecting and analyzing people's sketches and the time course of their construction. These two goals are interconnected. At present, evaluating sketches is very laborious and often inaccurate, and thus their use in classroom evaluation as well as in research is limited. If we find that certain properties of a sketch reveal a person's level of domain knowledge, then having a less effortful, more reliable, and more sensitive way to analyze sketches will be of great value both to instructors and to researchers in science learning. In this paper, our focus is on possible differences between novice and advanced geoscience students that may be revealed through analyzing their sketches with the CogSketch system. We will discuss CogSketch's methodological utility more broadly in the "General Discussion."

Spatial and Causal Knowledge in Geoscience

Geoscience students learn about the spatial structures that characterize the Earth, such as faults, folds, and fractures (Marshak 2005). They also learn about the causal processes that create these structures—both local deformation processes that create faults and folds, and large-scale global processes such as tectonic movements. All these topics are intensely spatial (Jee et al. 2010; Jee et al. 2013; Kastens et al. 2009; Libarkin and Brick 2002). Consider, for example, the geologic concept of a *fault*—a fracture in rock along which movement has occurred (Fig. 1). A fault is defined by specific spatial relationships between sections of rock, and these structures provide evidence of the effects of causal forces. For the fault in Fig. 1, a geoscientist would discern that the block of rock above the fault plane has moved downward relative to the block of rock below the plane. This pattern indicates that extensional forces acted on the rock until it fractured and was displaced along the plane. Thus, spatial and causal knowledge are tightly connected in this field.

Fig. 1 Photo of a fault (the fault plane runs diagonally from *top left to bottom right*)

How might the content and active construction of a sketch change as students gain knowledge of geologic structures and the causal forces underlying them? We consider this question with regard to two common types of images that geoscience students encounter: photographs of geological structures and causal diagrams of geologic processes.

Sketching from Photographs An important aspect of geological expertise is the ability to accurately characterize geological formations, such as outcrops of rock, and to be able to sketch such formations (either from photographs or from real-world scenes). Therefore, in experiment 1, we asked our participants to sketch rock outcrops. To ensure that participants had adequate opportunity to view the materials, the photographs remained present until they finished sketching.

What differences might we expect between novices and geoscience students? A number of studies suggest that expertise involves conceptualizing domain-related objects in terms of their deeper causal and goal-relevant structure, deemphasizing superficial details (e.g., Chi et al. 1981; Gentner and Rattermann 1991; Hay et al. 2013; Jee and Wiley 2007; Lowe 1993). As students acquire domain knowledge, their drawings emphasize domain-relevant causal structure (Gobert 2000; Gobert and Clement 1999). Kindfield (1992), for example, found that more expert biology students tended to create more schematized representations of the entities involved in mitosis. Sketches by less expert participants were more detailed, often including irrelevant physical dimensions and omitting relevant structural characteristics. This result is consistent with recent research by Hay et al. (2013), which explored expertise-related differences in drawings of brain cells. Experts in neuroscience created schematized sketches, biased toward the conventions and features relevant to their subfield. Undergraduate neuroscience students made drawings that were similar to textbook images of neurons. Similar patterns have been found in studies that ask participants to sort visual depictions. For example, in a well-known study, Chi et al. (1981) found that physics experts sorted diagrams of physics problems on the basis of common underlying principles, whereas novices relied on superficial commonalities, such as the presence of common objects such as ramps and pulleys. Likewise, when asked to group together elements of a diagram, experts tend to group together objects that are causally related, even if they are visually dissimilar and spatially distant (Lowe 1993). Based on findings from expertise studies, we could expect people with a background in geoscience to emphasize features related to domain-relevant spatial structures, such as faults and folds, in their sketches of geological photos. By contrast, novices are likely to include geologically irrelevant but perceptually salient details, such as vegetation.

Sketching Diagrams of Causal Processes The second kind of image we asked our participants to sketch was diagrams of large-scale causal processes. Causal diagrams are commonly used in geoscience to convey complex processes such as *subduction*, the process through which one tectonic plate moves under another. In geoscience instruction, these processes are often conveyed through causal diagrams like the one shown in Fig. 2. Understanding the process of subduction involves identifying both the spatial structure—that one plate is moving underneath another—as well as the causal structure. The student needs to understand that subduction is causally related to the process of seafloor spreading, which creates new ocean crust at divergent spreading zones, and is influenced by differences in crustal density between the converging plates, which determine which plate moves below the other. Thus, knowledge of both the kinematics (the displacements and deformation—both heavily spatial) and the dynamics (the causal forces at work) is needed to understand subduction.

We hypothesized that as students gain knowledge of spatially complex scientific material, such as plate tectonics, their diagrammatic sketches will become more focused on domain-relevant structures and more similar to standard geoscience depictions (Gobert and Clement 1999). The inclusion of relational and causal symbols in a sketch (e.g., directed arrows and
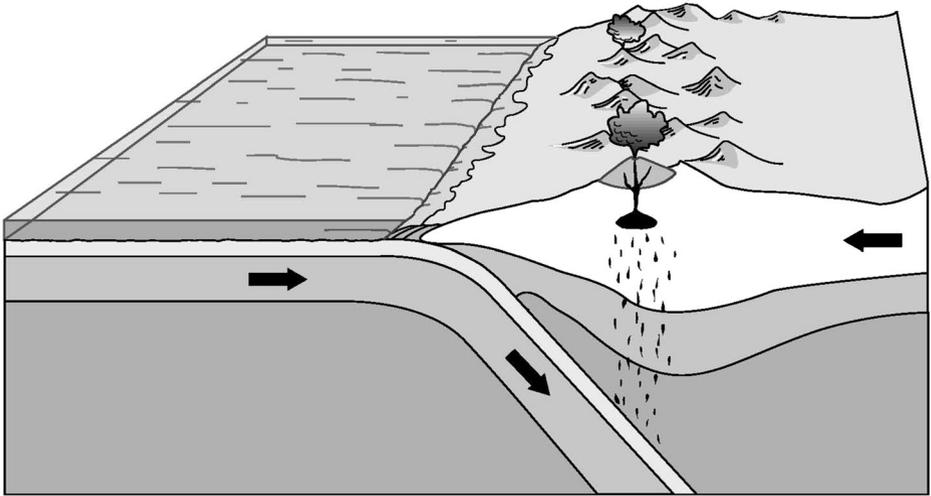
Fig. 2 Causal diagram showing the subduction process. Image source: U.S. Geological Survey (http://www.usgs.gov/)

other connectors) also reflects a person's causal understanding of the system they are sketching. When people are provided with a functional description of a system, they tend to include more of these symbols in their drawings (Heiser and Tversky 2006). As functional and causal knowledge increases with expertise (e.g., Chi et al. 1988; Hmelo-Silver and Pfeffer 2004; Proffitt et al. 2000), we predict that knowledgeable participants will include more relational symbols in their sketches than novices will.

An advantage of using a computer-based sketching program is that it can record the real-time process by which a sketch is drawn. Such timing information could reveal knowledge of key causal and spatial relations (Chase and Simon 1973; Gobet et al. 2001; Moss et al. 2006; Taylor and Tversky 1992; Tversky 2002). Taylor and Tversky (1992) found that people tend to sketch large features of a scene, such as mountains, before smaller-scale features, such as houses, suggesting that people organize their knowledge of the scene in terms of a hierarchical structure. A different kind of organization may be revealed when people sketch diagrams that depict a set of interrelated causal processes. Moss et al. (2006) found that senior engineering students recreated a diagram of a device starting with its input and moving to its output. Knowledge of how the device works influenced how they assembled it later. If people with geoscience knowledge actively interpret a causal diagram in terms of the sequence of causal events, they may construct their sketch in line with this sequence. They may draw an antecedent causal event before the resulting event, and a primary causal event before secondary causes.

Overview of the Experiments

In summary, our goals are, first, to discover whether and how patterns of sketching reveal differences in geoscience expertise, and, second, to test the efficacy of a computer-based sketching platform, CogSketch, for collecting and analyzing such sketches.

To ensure that the novices had access to the same visual information as the geoscience students, including causal symbols that could support diagram interpretation, each source image was available for the entire time that the participant sketched. The participants' task was simply to copy the image onto the computer screen. The question was whether differences in spatial and causal knowledge would affect the relational content and temporal construction of a sketch even in this copying task, where there was no inherent need to go beyond the information in the images. We tested this possibility across three experiments. Experiment 1 provides an initial test of the effects of geoscience knowledge on sketching. Experiments 2 and 3 further explore sketching of geological formations and causal diagrams, respectively, and explore whether differences between novices and geoscience students are specific to geoscience-related materials (suggesting an effect of domain-specific knowledge) or if they generalize to nongeoscience materials. In each experiment, we collected sketches on a tablet computer running CogSketch software, which tracks the spatial and temporal properties of the digital ink strokes that a person uses to create a sketch (using a stylus, mouse, or other input device). The software also records the conceptual labels and descriptions that the participant assigns to different parts within a sketch, as described below. By using Cogsketch as a research tool, we can examine not only the drawings and conceptual labels that comprise a sketch, but also the order in which different parts were created.

Experiment 1

Participants with differing levels of geoscience experience used CogSketch to sketch two photographs containing geological formations, and a diagram of a geoscience-related causal process. The participants copied a photograph or diagram onto the screen of a tablet computer using a stylus. For the geologic formations, the main question was whether geoscience students included more features related to geologically relevant structures, such as rock layers and folds. For the causal diagram, the main questions were whether geoscience students included more causal and relational information in their sketches, and whether the ordering of their sketches reflects the sequence of causal events depicted in the diagram.

Method

Participants

Twenty students from Northwestern University participated: ten geoscience students (six graduate students and four advanced undergraduates) and ten undergraduate psychology students who had no prior course experience in geoscience. The geoscience students received $12 for their participation, and the psychology students received course credit. All students participated individually.

Materials

Images We asked participants to sketch three geoscience-related images: two photographs of geological formations—one that showed a number of horizontal rock layers and one that showed folded rock—and one causal diagram illustrating subduction. The images are listed in the Appendix.

Tablet Computer An HP tablet computer with a 12-in. screen and a stylus was used for data collection.

CogSketch Software CogSketch is an open-domain, general-purpose sketch understanding system (see Forbus et al. 2011 for a more detailed description). CogSketch allows users to tell it what entities they are sketching via a specialized interface. Every sketch drawn in CogSketch is made up of *glyphs*. A glyph is a shape, an object, or a relationship that has been drawn by the user. The user presses a button to indicate that they are about to start drawing a glyph; they then draw as many ink strokes as they want, and end the glyph by pressing the button again[1]. As a separate step, users describe what they intend a glyph to represent by labeling it with either a concept from the knowledge base associated with CogSketch or with a word or phrase. Each glyph is represented internally as digital *ink*—essentially the lines drawn by the user when creating the glyph—and *content*—a conceptual entity representing what the glyph depicts. A video demonstration of CogSketch can be viewed at http://www.youtube.com/watch?v=o6f1udbIkrI.

CogSketch offers several features relevant to data gathering and interpretation. First, CogSketch records a time stamp along with every ink point (sampled at a rate of 4 points/s). These data can be exported to common formats, such as Microsoft Excel. Second, because the user has manually segmented the digital ink into glyphs, there is no ambiguity about which parts belong to which depicted entity. Third, there is no need to analyze video and audio transcripts to ascertain what labels should be used for the parts of a sketch, making CogSketch data simpler and cheaper to analyze.

Procedures

The tablet computer was folded flat and placed on a desktop in front of the participant. The experimenter first gave the participant a brief tutorial and practice using the software and hardware, followed by an overview of the task and procedure. Participants were asked to try to copy the experimental images to the best of their ability, to use the glyph function in CogSketch to divide their sketch into the parts that seemed natural to them, and to label the parts as they saw fit, using words or phrases. They were given up to 3 min/sketch, with a warning given 1 min before the time limit. For the duration of each trial, a color printout of the image was presented next to the computer screen. The participant sketched directly on the computer screen using the stylus. The images were sketched in the same order for all participants: subduction 1, rock layers, and then folded rock. Each session took under 30 min to complete.

Results

Two raters coded each sketch independently with high overall inter-rater reliability (Cohen's Kappa=0.89). Sketches were scored blind to any participant identification information. Coding disagreements were settled through discussion.

Geological Formation Sketches

A geoscience expert identified two or three features in each photograph that captured the geologic structure of the image. For example, the expert identified three key features in the photograph of folded rock (Fig. 3): the three large folds. For each fold, the participant had to

---

[1] This accurately describes the interface used in these experiments; since then, the interface has been streamlined so that pressing a "start glyph" button is unnecessary, and users can merge or split their ink as they desire.
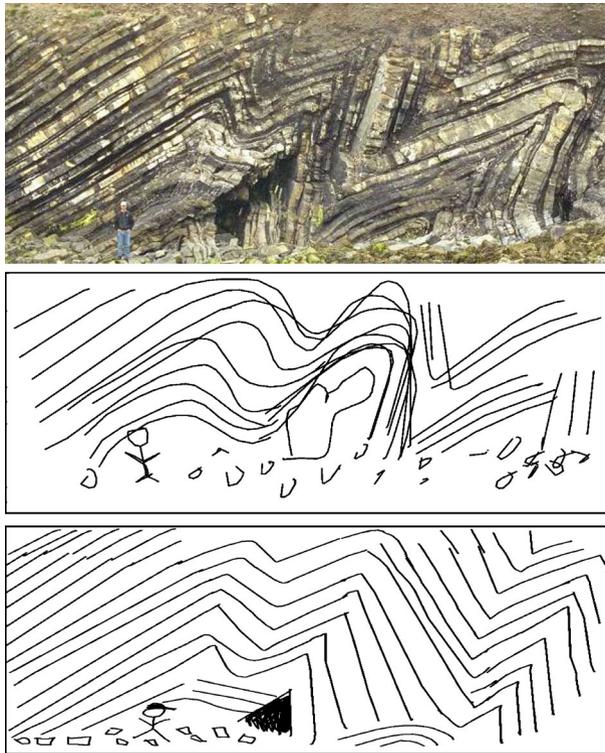
Fig. 3 Example sketches of a photograph of a geological formation, folded rock (original image in *top row*) by a novice (*middle row*) and geoscience student (*bottom row*)

include its approximate location and apex in their sketch, as well as the angle and length of the folded parts of rock. Figure 3 displays representative examples of sketches from novice and experienced participants for this photograph. The novice student's sketch did not distinguish the layers of folded rock—depicted as crisscrossing lines—and clearly failed to capture the rightmost fold. The geoscience student's sketch, however, captured all three folds.

Participants' sketches were coded for the presence of the key features from the two photos. We computed the mean proportion of key features in each sketch. Overall, geoscience students' sketches contained a higher proportion of the geological features ($M=0.78$, $SD=0.26$) than novices' sketches ($M=0.49$, $SD=0.19$), $t(18)=2.95$, $p<.05$, $d=1.27$.

Causal Diagram Sketches

Relational Content Each part/glyph in each sketch was coded as either an *object*—a part that depicted some tangible entity—or a *relation*—a part that depicted the behavior or connectivity of either a tangible or intangible entity, e.g., a force. The information in both the sketch and label was used for this classification. A part was scored as a relation if (a) it contained a relational symbol, such as an arrow indicating movement, or (b) its label described the part in terms of its relation to other entities (e.g., "subducting slab"). Figure 4 displays parts that were coded as objects and relations.
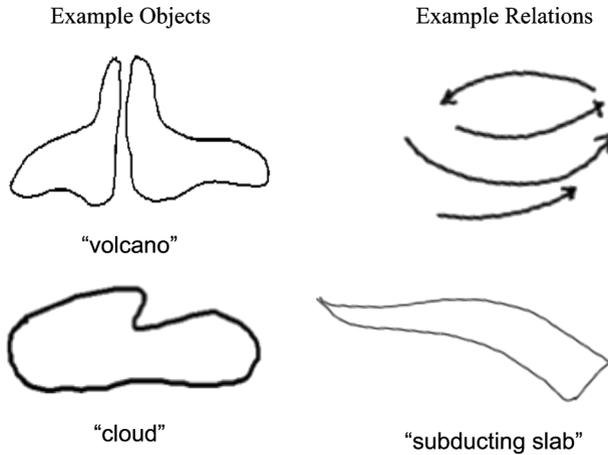
Example Objects   Example Relations

"volcano"

"cloud"   "subducting slab"

Fig. 4 Examples of parts coded as objects and relations. Note: The participant's label appears in quotes below the part. Our coding took these labels into account

Figure 5 displays sample sketches of the causal diagram from geoscience students and novices. The figure reveals that the novice sketch contains no relational symbols, whereas the geoscience student's sketch contains several (the directed arrows[2] that we have highlighted in red). The mean proportion of relations was compared between the two groups. Geoscience students' sketches contained a higher proportion of relations ($M=0.53$, $SD=0.30$) than novices' sketches ($M=0.11$, $SD=0.17$), $t(18)=3.90$, $p<.05$, $d=1.72$.

Sketch Ordering This analysis investigated whether the sketch displayed particular geologically significant events, and in what order these events were drawn. A geoscience expert identified a set of geologically significant events in the diagram— e.g., seafloor spreading and plate convergence—and specified their causal order. These events were conveyed by one or more directed arrows in the diagram, which were a subset of the total number of arrows in the diagram. The presence of these arrows in a sketch was taken as evidence that the participant represented the corresponding causal event. We scored the participants' sketches for the presence of these causal events, and used the timing data recorded by CogSketch to determine the order in which they were sketched.

On average, the geoscience students ($M=0.72$, $SD=0.23$) included a higher proportion of the causal events than the novices ($M=0.15$, $SD=0.32$), $t(18)=4.45$, $p<.05$, $d=2.05$. As the novices rarely included causal events, we were unable to analyze the ordering of events in their sketches. For the geoscience student sketches, we computed the mean ordering for each of the four events, including only the sketches that contained at least two causal events. This left us with useable data from seven of the geoscience student sketches. Mean ordering for the geoscience students was compared with the ordering of

---

[2] In our coding scheme, each part of the sketch that contained an arrow was coded as a relational part. Although there are five arrows in the geoscience student's sketch, the participant grouped the arrows into three spatially separated parts; thus, only three relational parts were counted. (Coding also took into account the labels given to the parts, so additional relational parts may be present in Fig. 5 in addition to those indicated in red.)
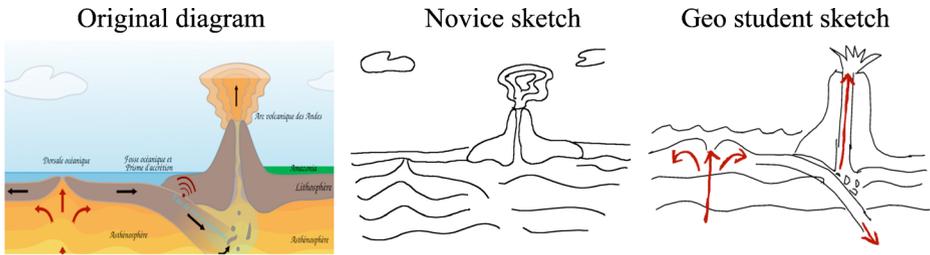
Fig. 5 Example sketches of a geoscience-related causal diagram (subduction 1) from experiment 1. Relational symbols are indicated in red

the expert by computing Kendall's coefficient of concordance (Kendall's W). This statistic provides an index of the similarity between distinct orderings—ranging from 0 (minimum similarity) to 1 (maximum similarity)—and can be tested for statistical significance using a chi-square distribution (see Taylor and Tverky (1992) for a similar analysis). The value of Kendall's W was .70, suggesting a relation between the mean[3] geoscience student ordering and the expert ordering; however, this statistic did not approach significance, $\chi^2(3)=4.20$, $p=.24$.

Discussion

The results of experiment 1 support the hypothesis that differences in geoscience knowledge lead to systematically different patterns of sketching. The results also are encouraging for our goal of capturing differences in scientific knowledge using CogSketch. The results revealed that even in a simple copying task, geoscience students and novices differed both in terms of what information they included and the order in which they drew the elements in the sketch. When sketching photographs of geological formations, geoscience students were more likely than novices to include spatial features that reflected key geologic structures. When sketching the causal diagram, geoscience students emphasized causal relations and processes, and their sketches often reflected the causal ordering of geologic events. Novices sketched fewer relations overall and often omitted the events that a geoscience expert considered important.

The finding that geoscience students differ systematically from novices in sketching geoscience materials supports the hypothesis that sketching reveals domain understanding. The results are consistent with the previous findings that knowledgeable individuals emphasize domain-relevant structures in their sketches (Gobert 2000; Hay et al. 2013; Kindfield 1992). Yet, the generality of this result is unclear. We do not know whether geoscience students would show similar effects when drawing images from different domains, because only geoscience-related materials were included in experiment 1. If our findings are attributable to differences in domain knowledge—and not some other factor, such as drawing ability or spatial skill—then the between-group differences we observed in experiment 1 should not carry over to nongeoscience materials, even those that overlap in their features and causal complexity. This is a critical issue. Prior work has found that experts often rely on domain-general knowledge for thinking and reasoning (e.g., Schunn and Anderson 1999).

---

[3] The coefficient of agreement for mean ordering was lower than the median and modal values because one student sketched the events in perfect reverse order (Kendall's $W=0$). The median value of Kendall's W was .75 and the mode was 1.0 for geoscience students who drew at least two causal events.

To determine whether the differences between geoscience students and novices are specific to geoscience, we asked participants in experiments 2 and 3 to sketch control images in addition to geoscience images. We chose spatially intricate pictures from domains unrelated to geoscience, such as a cross-section of a fruit and an artificial heart valve. The goal was to find images that were about equally familiar to both groups. For the control photographs, we sought images that contained similar features (folds, layers, etc.) to the geoscience images; for the control causal diagrams, we sought images similar in causal complexity. If geoscience students construct sketches differently to novices regardless of the domain, then we should find similar patterns for both the geoscience and nongeoscience materials. However, if geoscience students are utilizing domain-specific knowledge during sketching, then we should observe differences for the geoscience-related images but not for the control images.

A further goal of experiments 2 and 3 was to generalize our findings to a broader range of materials. Experiment 2 tested additional geoscience formations and experiment 3 tested additional causal diagrams. Both of these experiments included a greater number of images than did experiment 1 to gain experimental power. Finally, to explore how sketching effects relate to prior knowledge of the *particular* topics shown in the images, at the end of the study we asked people to rate their prior knowledge of the content of each image.

Experiment 2

This experiment focused on sketches of geological formations. As in experiment 1, participants were given the simple task of copying photographs: three photographs of geological formations (not used in experiment 1) and three control (non-geological) photographs. The geological and control photos contained similar surface features, such as curves, layers, and complex lines, and were comparably intricate. If geoscience students are simply better at detecting and depicting the spatial features of objects in a sketch, then they should sketch both the geoscience images and the control images differently than novices do. However, if the differences between novices and geoscience students reflect domain understanding, then differences should emerge only for the geoscience-related sketches. Furthermore, such expertise-related differences should be most pronounced for images on which the geoscience students rate themselves as having greater prior knowledge.

Method

Participants

Sixteen participants from Northwestern University took part: eight geoscience students (five graduate students and there advanced undergraduates) who were paid for participating, and eight undergraduates with no prior course experience in geosciences, who received course credit in an introductory psychology course. None of the participants took part in experiment 1. All participated individually.

Materials

There were three geoscience images and three nongeoscience images, all listed in the Appendix. The geoscience images were selected to be novel, unfamiliar examples of the kinds

of structures that geoscience students would be likely to encounter in a geoscience course. Sketches were again collected on a tablet PC running CogSketch software.

Procedures

The procedures for the pre-task tutorial on CogSketch and the sketching task were the same as in experiment 1. Each participant received the images in a different random order. At the end of the study, we asked participants to rate their level of prior knowledge of the object or structure depicted in each photo on a scale from 1 (poor prior knowledge) to 7 (excellent prior knowledge). Each session lasted less than 45 min.

Results

Two raters coded each sketch independently with high overall inter-rater reliability (Cohen's Kappa=0.88). Coders were blind to any participant identification information. Coding disagreements were settled through discussion.

Self-ratings of Prior Knowledge for the Objects and Structures in the Images

The mean prior knowledge ratings for the geoscience and control photographs are shown in Table 1. The knowledge level×image type interaction was significant $F(1, 14)=6.60$, $p<.05$, $\eta_p^2=.32$. Geoscience students rated their prior knowledge for the objects and structures in the geoscience photographs higher than the novices did, $t(14)=2.66$, $p<.05$, $d=1.33$. However, the two groups rated their prior knowledge at roughly the same level for the control photos, $t<1$, ns. There were no significant main effects.

Presence of Key Features

As in experiment 1, a geoscience expert analyzed the geoscience photos for their key features. He identified an average of about five features ($M=4.67$) in each photo. For the control images, we asked two judges who were unaffiliated with the research to identify the key features in each photograph—i.e., those they considered most central to the overall structure.

Table 1 Mean ratings of prior knowledge (out of seven) and mean proportion of key features in participants' sketches for experiment 2 photographs

Experience level	Image type	
	Geoscience	Control
Mean ratings of prior knowledge		
Geoscience student	5.13 (1.61)	4.79 (1.02)
Novice	2.88 (1.76)	4.86 (1.28)
Mean proportion of key features		
Geoscience student	.89 (.12)	.91 (.10)
Novice	.62 (.16)	.86 (.16)

Note. Ratings of prior knowledge were made on a 7-point scale, where 1=poor prior knowledge of the object or structure in the image, and 7=excellent prior knowledge. SDs in parentheses

The judges agreed entirely for two of the photos, and reached agreement on the third after a short discussion. They identified an average of about four features (M=3.67) in each control photo. Participants' sketches were coded for the presence of these key features. Table 1 shows the mean proportion of key features for novices and geoscience students for both kinds of images.

We found that geoscience students drew more key features than novices, but only for the geoscience images. For the control images, both groups drew a high proportion of key features. This finding was revealed in a 2 (experience level)×2 (image type) ANOVA; the interaction between experience level and image type, $F(1,14)$=10.24, p<.05, η_p^2=.42. As predicted, geoscience students sketched a higher proportion of key features in geoscience images than did novices, $t(14)$=4.07, p<.05, d=2.07. No such difference was found for the control images, t<1, ns. This interaction relates two main effects: (1) Geoscience students drew more key features overall than novices, $F(1, 14)$=7.96, p<.05, η_p^2=.36, and (2) Combining both groups, the sketches of the control images contained more key features than those of the geoscience images, $F(1, 14)$=3.37, p=.09, η_p^2=.19. Another way to describe this pattern is to note that both groups captured a fairly high level of structural detail for the control photos, but only the geoscience students maintained this level on the geoscience photos. Figure 6 shows examples of geoscience student and novice sketches of each type of image.

Figure 6 also reveals another interesting trend. Several geoscience students added relational symbols (which were *not* present in the original photos) to their sketches. The half arrows in the geoscience student's sketch are conventionally used to indicate movement on a fault plane, as determined using "offset markers" (evidence that the previously adjacent beds of rock have been displaced). These arrows were not present in the original image (and in our scoring, were not counted as key features). Nevertheless, several geoscience students drew them, usually after the key features had been sketched. In a technical sense, adding these arrows is an error, as subjects were instructed to copy the photographs. This result
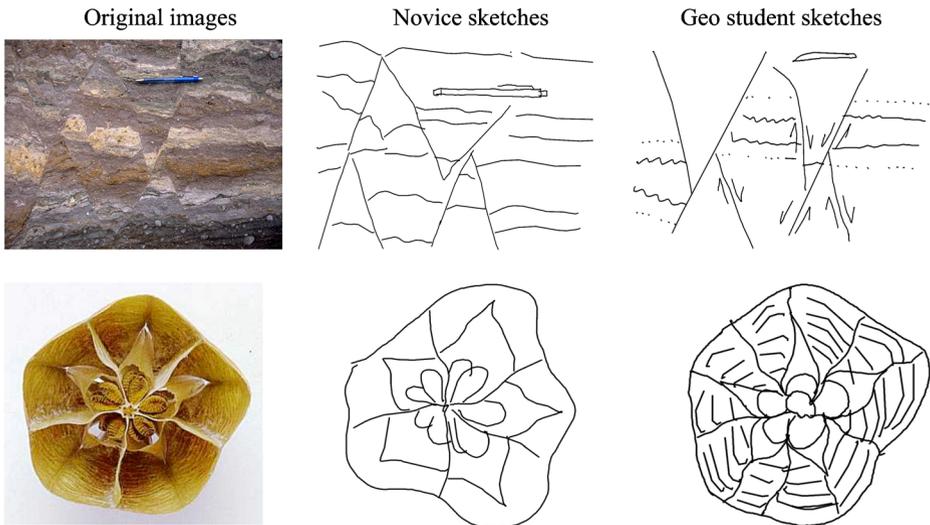

Fig. 6 Example sketches of a geoscience-related (cross-cutting faults; *top row*) and a nongeoscience photograph (fruit cross-section; *bottom row*) from experiment 2

strongly demonstrates that the geoscience students brought domain knowledge to bear on their construal of the materials.

Knowledge Ratings and Sketching Performance

If the differences between novices and geoscience students are driven by domain knowledge, then the differences should be greatest for those items on which geoscience students rated themselves as having high prior knowledge. We therefore conducted a regression analysis to predict the proportion of key features in a sketch from three factors: the participant's rating of knowledge for the content the image, the type of image (geoscience vs. control), and the interaction between these variables. Geoscience experience level (experienced vs. novice) was entered as a covariate and the variables were centered for the analysis. The model accounted for a significant amount of the variance in the proportion of key features, $F(4, 89)=7.68$, $p<.05$, $R^2=.26$; however, only the prior knowledge rating x image type interaction term was a significant predictor, $\beta=.40$, $t(93)=2.92$, $p<.05$, partial $r^2=.27$. This interaction reveals that prior knowledge ratings were a better predictor of the proportion of key features for the geoscience-related sketches than for the control sketches. Thus, a participant's self-rating of understanding of a particular geological structure was an important predictor of whether they included key geologic features in their sketch. By contrast, the key features of the control photos were sketched by nearly all of the participants, regardless of their self-assessed prior knowledge of the material in those images.

Discussion

The results of experiment 2 further bear out the predicted difference between novices and experienced students. As in experiment 1, geoscience students sketched a higher proportion of features that reflected geologic activity, such as fault lines and folds, for photos of geologic formations. When given photos of non-geologic structures, however, geoscience students drew sketches that were highly similar to those of novices (geoscience students included .91 key features vs. .86 for novices, a nonsignificant difference). Only when the images were related to geoscience was there a significant difference between geoscience students and novices. In fact, participants' self-assessed understanding predicted the presence of geologic features in their sketches of geological photos. These findings suggest that the difference between the groups stemmed chiefly from differences in domain knowledge, as opposed to some more general factor. Geoscience students appear to rely on domain-specific spatial and conceptual knowledge in order to capture the geologically relevant features of an image. Indeed, geoscience students occasionally added causal symbols to their drawings, a strong indication that causal knowledge was activated during the sketching process.

Experiment 3

This study focuses on sketching of causal diagrams. We sought to generalize the pattern found in experiment 1 (in which one causal diagram was drawn): that geoscience students differ from novices both in *what* they include in their sketches, and in the *order* in which they construct each sketch. As in experiment 2, we included control (nongeoscience) diagrams as well as geoscience diagrams to assess whether any differences between groups could be attributed to domain-general knowledge or skill vs. domain-specific knowledge of the causal processes

depicted. Our intention was to select control diagrams that were similar in causal complexity to the geoscience diagrams, and about equally familiar to the geoscience students and novices. Also as in experiment 2, we asked participants to rate their prior knowledge of the system or process displayed in each diagram at the end of the session. If geoscience students invoke knowledge of the causal processes depicted in the geoscience diagrams, then relative to novices (1) they should emphasize more relational information in their sketches of geoscience diagrams, and (2) the order in which they draw parts of the sketch should more closely reflect the sequence of causal events depicted in the diagrams. Further, if this predicted pattern stems from understanding of geoscience, then it should hold only for the geoscience diagrams, not for the control diagrams; and it should be most pronounced for diagrams on which geoscience students rate themselves as having most knowledge.

Method

Participants

Sixteen participants from Northwestern University took part: eight geoscience students (five graduate students and three advanced undergraduates) who were paid and eight undergraduates with no prior course experience in geosciences, who received course credit in introductory psychology. None of these participants took part in either experiment 1 or 2. All participated individually.

Materials

We used four geoscience-related diagrams and four nongeoscience (control) diagrams. The images are listed in the Appendix.

Procedures

The procedure for the sketching task was as in the other experiments, except that the participants were given 4 min per sketch. Each participant received the images in a different random order. At the end of the study we asked participants to rate their level of prior knowledge about the system or process displayed in each diagram on a scale from 1 (poor prior knowledge) to 7 (excellent prior knowledge).

Results

For the relational content and ordering analyses below, two raters coded each sketch independently with high overall inter-rater reliability (Cohen's Kappa=0.92). Sketches were scored blind to any participant identification information. Coding disagreements were resolved through discussion between the raters.

Relational Content Each part/glyph in each sketch was coded as either an *object* or a *relation* following the method used in Experiment 1. The mean proportion of relations for the geoscience and control causal diagrams is shown in Table 2. There was a statistically significant experience level x image type interaction, $F(1, 14)=6.17$, $p<.05$, $\eta_p^2=.31$. As predicted, for the geoscience-related images, geoscience students sketched a higher proportion

Table 2 Mean ratings of prior knowledge (out of seven) and mean proportion of relations in participants' sketches for experiment 3 diagrams

Experience level	Image type	
	Geoscience	Control
Mean ratings of prior knowledge		
Geoscience student	4.97 (0.94)	2.38 (0.63)
Novice	3.03 (1.14)	2.45 (0.86)
Mean proportion of relations		
Geoscience student	.48 (.15)	.35 (.15)
Novice	.27 (.16)	.33 (.15)

Note. Ratings were made on a 7-point scale, where 1=poor prior knowledge of the process or system in the diagram, and 7=excellent prior knowledge. SDs in parentheses

of relations than the novice group, $t(14)=3.45$, $p<.05$, $d=1.18$. For the control images, however, there was no significant difference between the geoscience students and novices[4], $t<1$, *ns*. There were no significant main effects. Figure 7 shows a representative sample of sketches from experienced and novice participants, with relational symbols indicated in red.

As our coding took into account the written labels as well as the nature of the glyphs (relational symbols, such as arrows), the difference in relational content for the geoscience-related images could be due to only one of these components or both. There was inadequate statistical power to test the experience level x image type interaction using either relational labels or relational symbols alone as a dependent variable, but trends in the data suggest that the geoscience students used more of each in their sketches of the geoscience-related images.

Geoscience students included a higher proportion of relational labels in their geoscience-related sketches ($M=.23$, $SD=.18$) than novices ($M=.09$, $SD=.13$), as well as a higher proportion of relational symbols ($M=.38$, $SD=.14$ vs. $M=.23$, $SD=.15$). (Note that the sum of the relational labels and symbols does not equal the number of relational parts in Table 2, because several parts included both a relational label and symbol, and these counted as a single relation in our coding.) For the control sketches, which did not show a difference in overall relational content between geoscience students and novices, the proportion of relational labels was similar for the geoscience students ($M=.08$, $SD=.14$) and novices ($M=.04$, $SD=.05$), as was the proportion of relational symbols ($M=.33$, $SD=.14$ and $M=.30$, $SD=.15$, respectively). In summary, relational labels were relatively uncommon, except for the geoscience students' use of them in their geoscience-related sketches. The effect of experience level on the overall number of relational parts in the geoscience-related sketches appears to be due to an increased use of both relational labels and symbols by the geoscience students.

Self-ratings of Prior Knowledge of the Systems and Processes in the Images

Table 2 shows each group's mean prior knowledge ratings for the system or process in the geoscience and control causal diagrams. A 2 (experience level)× 2 (image type) ANOVA

---

[4] Experiments 2 and 3 both showed a similar statistical interaction, i.e., an advantage for geoscience students on the geoscience images, but no difference between the groups on the control images. However, whereas in experiment 2 both groups produced a high proportion of relations on the control images, in experiment 3 both groups produced a low proportion of relations on the control images.
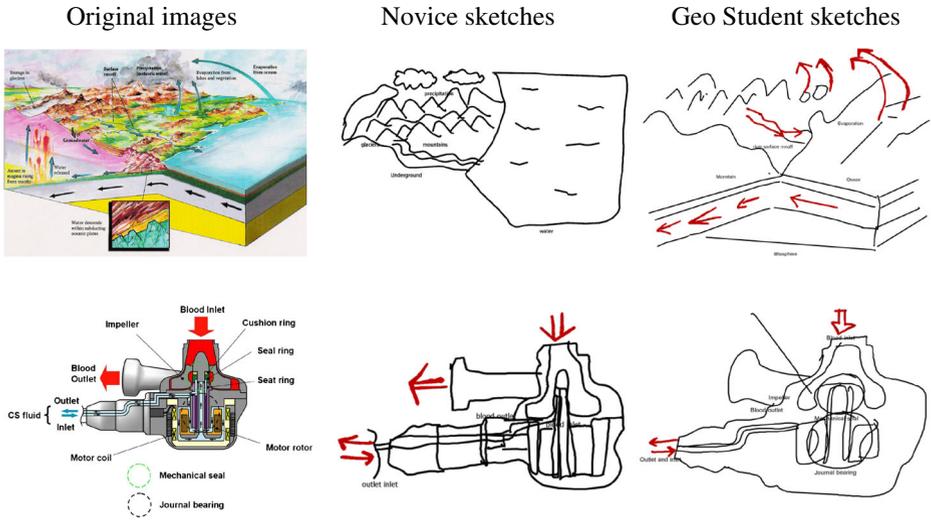
Original images	Novice sketches	Geo Student sketches

Fig. 7 Example sketches of a geoscience-related (water cycle; *top row*) and a nongeoscience causal diagram (blood pump for artificial heart; *bottom row*) from experiment 3. Relational symbols are indicated in *red*. In the materials presented to participants, all captions were legible. *From left to right across the top of the water cycle diagram*, the captions are storage in glaciers, surface runoff, precipitation (meteoric water), evaporation from lakes and vegetation, and evaporation from oceans. *Across the bottom, from left to right*, the captions are ascent in magma rising from mantle and water descends within subducting ocean plates

revealed a significant experience level x image type interaction, $F(1,14)=8.28$, $p<.05$, $\eta_p^2=.37$. As expected, the geoscience students rated their prior knowledge higher on the geoscience images than did the novices, $t(14)=6.14$, $p<.05$, $d=1.86$. By contrast, both groups rated their prior knowledge for the control images as about the same, $t<1$, *ns*. Two main effects are subsumed by this interaction: geoscience students rated their knowledge higher than novices, $F(1, 14)=10.19$, $p<.05$, $\eta_p^2=.42$, and the geoscience diagrams were given higher ratings than the control diagrams, $F(1,14)=20.68$, $p<.05$, $\eta_p^2=.60$. We also note that the prior knowledge ratings were lower for the control causal diagrams than the control photos from experiment 2. One reason is that participants rated their knowledge for the *systems* displayed in the diagrams, not the objects that were present. The diagram on water conservation in the home included many objects that were likely familiar, such as a shower, toilet, and garden hose. The system that connected these different objects was probably less familiar. Most important for our purposes is that the control images were about equally familiar to both groups.

Knowledge Ratings and Sketching Performance

As in experiment 2, we conducted a multiple regression analysis to predict the mean proportion of relations in a sketch from the participant's rating of prior knowledge, the type of image, and the interaction between these variables. Geoscience experience level (experienced vs. novice) was entered as a covariate and the variables were centered for the analysis. This model failed to account for a significant proportion of the variance in the proportion of relations, $F(4,117)=1.24$, *ns*, $R^2=.04$, and no variable accounted for more than 1 % of the unique variance ($\beta s<.16$, $ts\leq1.0$). Yet, as is evident from the results in Table 2, there was a strong relation between *mean* prior knowledge ratings for the geoscience images and the *mean* proportion of relations in sketches. Participants who gave higher mean prior knowledge ratings

for the geoscience images had a higher mean proportion of relations in their sketches, $r(16)=.58$, $p<.05$. However, there was no such relation for the control images, $r(16)=.08$, *ns*.

Sketch Ordering

Geoscience Diagrams For each diagram, a geoscience expert identified two to four key causal events (depicted by one or more directed arrows) and specified their causal order. These were a subset of the total number of arrows in the diagram. We scored the participants' sketches for the presence of these causal events, and found that geoscience students ($M=0.51$, $SD=0.17$) drew a higher proportion of key causal events than novices ($M=0.26$, $SD=0.22$), $t(14)=2.63$, $p<.05$, $d=1.27$. When key events were sketched, we used the timing data from CogSketch to determine their ordering. We compared the order in which geoscience students sketched the causal events to those of other participants and to the causal ordering identified by the expert.

Our main finding was that, unlike novices, geoscience students tended to sketch causal events in line with causal order. This finding emerged from an analysis of the relation between drawing order and the causal order of events in the diagrams. For geoscience sketches that contained at least two causal events (63 % of the geoscience student sketches and 34 % of the novice sketches), we scored the causal events according to the order in which they were drawn[5], assigning the first key event sketched a score of 1, the second a 2, and so on. These scores were organized into a single array that included the participant's scores for all four sketches. We then derived an average array for the geoscience students and novices, and compared each of these to the array representing an expert's ordering of causal events for the four diagrams, using Kendall's coefficient of concordance to compare the similarity between the arrays (see Table 3).

The geoscience students' ordering was highly similar to the expert's ordering, Kendall's $W=0.98$, $\chi^2(11)=21.64$, $p<.05$, implying that geoscience students sketched the events in causal order. In contrast, neither the similarity between the novices' ordering and the expert ordering, Kendall's $W=0.39$, $\chi^2(11)=7.72$, *ns*, nor the similarity between the geoscience students' and novices' orderings, Kendall's $W=0.43$, $\chi^2(11)=8.51$, *ns*, reached significance. Mean ordering was representative of ordering at the individual level. The median value of Kendall's W was 1.0 between the expert and the geoscience students and 0.38 between the expert and the novices.

Control Diagrams The control diagrams were from different domains, so two experienced individuals were recruited to identify the causal events. An engineer with over 20 years of experience identified four key causal events in each of the mechanical diagrams (carburetor, blood pump, and water conservation system). An advanced biology student identified five key causal events in the photosynthesis diagram. Each causal event was represented by one or more directed arrows in the diagram, and the experienced individuals specified the order of these events. We scored participants' sketches for the presence of the causal events. Although there was a suggestion of a higher proportion of key causal events for geoscience students, $M=0.55$ ($SD=0.12$), than for novices, $M=0.44$ ($SD=0.15$), this difference was not statistically significant, $t(14)=1.34$, $p=.20$.

We followed the procedure described above for analyzing the control diagrams, using the timing data from CogSketch to analyze the order in which participants sketched the causal events. We again included only the sketches that contained at least two causal events for a given diagram, leaving us with useable data from 100 % of the geoscience student sketches and 79 % of the novice sketches. As shown in Table 3, none of the similarity comparisons reached statistical significance, Kendall's $Ws<0.58$, $ps>.05$. Thus, in contrast to the finding for

---

[5] We used CogSketch's ordering of events but omitted events not considered to be key events.

Table 3 Agreement in mean event ordering (Kendall's coefficient of concordance) for sketches of experiment 3 diagrams

	Geoscience students	Expert coder
Geoscience diagrams		
Novices	0.43	0.39
Geoscience students		0.98*
Control diagrams		
Novices	0.61	0.51
Geoscience students		0.58

*$p<.05$, significance

the geoscience diagrams, we saw no evidence that either group followed causal order when sketching the control diagrams.

Discussion

These findings again bear out the prediction that geoscience expertise would be reflected in participants' sketches of geoscience materials. As in experiment 1, geoscience students included more relational content in their sketches of geoscience-related causal diagrams. Additional analyses suggest that this difference was due to increased use of both relational labels and relational symbols, such as arrows indicating movement or causality, by the geoscience students. The difference was found only for geoscience diagrams. For causal diagrams from other fields, there was no significant difference in the amount of relational information included by the two groups. This is consistent with the idea that the differences in geoscience sketching stem from greater domain knowledge on the part of the geoscience students. Further evidence for the role of domain knowledge comes from the finding that participants' self-ratings of knowledge of the systems and processes in the geoscience diagrams correlated with their sketching performance. However, unlike our finding for sketches of geological formations (experiment 2), knowledge ratings for individual geoscience diagrams were not a significant predictor.

Geoscience students also differed from novices in terms of *how* they sketched, and again, this difference was specific to geoscience materials. Geoscience students sketched causal events in the geoscience diagrams in an order consistent with the causal order identified by an expert. Novices, when they did sketch causal events from the geoscience diagrams, did not order them in terms of their causal sequence. Likewise, neither geoscience students nor novices tended to draw the causal events from the control diagrams in sequence. Again, this suggests an effect of domain-specific knowledge.

General Discussion

These studies show that people's sketching patterns can reveal differences in their understanding of geoscience. When sketching photographs of geological formations, geoscience students were more likely than novices to include spatial structures that reflected geologic activity, such as faults and folds. When sketching diagrams of geologic processes, geoscience students included more relational information in their sketches, and often drew causal events in sequential order. Strikingly, these effects of domain knowledge occurred despite the fact that the task was simply to copy the photograph or diagram. Our participants could view the photos

and diagrams for the entire time that they sketched; yet there were still sizable differences in *what* and *how* they sketched.

We also asked whether these group differences could have stemmed from some general skill or knowledge that the geoscience students possessed, rather than from greater domain understanding. The pattern of evidence suggests that the effects were due largely to domain-specific knowledge. For both photographs and causal diagrams, the differences between geoscience students and novices in what was sketched, and in the order in which it was sketched, held only for geoscience materials, not for control materials. Further, these patterns aligned with participants' self-ratings of knowledge for the content of the images. Of course, it is possible that effects of general knowledge or skill will emerge in studies involving different sketching tasks and materials. Further research is needed to address this issue, and to explicitly test the influence of general skills—such as spatial thinking and visualization—that have been associated with attainment in science (Black 2005; Kali, and Orion 1996; Kastens et al. 2009; Shea et al. 2001; Uttal and Cohen 2012; Wai et al. 2009).

Effects of Conceptual Knowledge on Sketching Our results suggest that the geoscience students relied on conceptual knowledge even when simply copying photographs of geological formations. Their construals were not solely based on the visual information present; they were influenced by participants' mental models, including knowledge of the forces that could produce a given structure (Forbus 2011; Gentner and Stevens 1983; Hay et al. 2013; Tversky 1991). This is underscored by the finding that geoscience students often added arrows depicting the direction of forces to their sketches of geological formations. Expertise-related exaggerations in spatial depiction have also been found in other arenas. For example, Landy and Goldstone (2007) found that when knowledgeable students copy math equations, they tend to group the elements of operations that must be done earlier in the equation more closely in space. Likewise, acquired categorical knowledge can lead to underestimating distances between same-category items (Holden et al. 2010; Tversky 1981; Uttal et al. 2010). Knowledge of forces and causal relations might lead to similar distortions in sketches of scientific phenomena. For example, understanding of the forces that caused displacement may lead the sketcher to exaggerate this displacement in the direction of the forces. Such findings would shed further light onto the ways that learning affects the organization of domain knowledge (Forbus 2011).

Effects of domain knowledge were also evident for causal diagrams. Geoscience students were more likely than novices to include causally relevant parts for diagrams depicting geoscience processes, but not for control diagrams. This fits with evidence that knowledge of the identity and causal role of the objects in a diagram affects diagram construction. For example, Heiser and Tversky (2006) found that novices who are provided with a functional description of the parts of a diagram were more likely to include relational symbols when sketching the diagram than those provided with a merely physical description.

A particularly intriguing finding was that when copying causal diagrams, geoscience students tended to sketch following the causal order of events. These results align with findings on expertise in engineering and design[6]. Moss et al. (2006) had freshmen and senior engineering students recreate diagrams of devices using a specialized computer interface that allowed them to drag and drop component parts into a diagram space. Moss et al. found that seniors tended to assemble the parts based on functionality, whereas freshman tended to chunk together the same types of components. More to the point, seniors were also more likely to

---

[6] This pattern has also been observed in biology. Louis Gomez (personal communication, May 2012) reports that students who understand the photosynthesis process tend to copy diagrams of photosynthesis in causal order.

follow the flow of energy as they recreated the diagrams, as with the geoscience students in the present research. If this pattern of causal ordering holds up in other domains, then students' order of sketching may provide an unobtrusive measure of their domain understanding.

In sum, this work advances our understanding of science learning by identifying markers of domain understanding that appear in students' sketches. The findings suggest that, whereas novices tended to focus on depicting the objects in the scene, experienced students activated causal knowledge as they copied images, affecting both what and how they sketched. It is intriguing that similar signs of domain knowledge have been found in other tasks, such as recreating a machine diagram, and with materials that are less intensely spatial, such as math equations. This raises the possibility that identifying certain properties of a sketch—such as its relational content, spatial layout, and temporal construction—may be generally indicative of a student's level of understanding.

Using Sketching Software to Assess and Enhance Science Learning We also aimed to make a methodological contribution, by testing whether data gathered by CogSketch can be used to detect markers of domain understanding. We found that it can. Two features were particularly useful. First, CogSketch automatically records the order in which objects are drawn. As our results show, such ordering information can be diagnostic of differences in knowledge. In the past, such timing data has been onerous to obtain. For example, in the Rey–Osterreith Complex Figure test, a widely used assessment of neurological disorder (e.g., Duley et al. 1993; Shin et al. 2006), timing data for participants' sketches are obtained by giving participants colored pencils. After each time segment, they are asked to switch to a new color. Clearly, this kind of task would be prohibitive in classroom settings. Timing data can also be obtained by analyzing audio and video transcripts, but this is time-consuming and unsuitable for most instructional purposes. Although more analysis are needed to discover when time-course data is most useful, our findings suggest that automatic tracking of the time course of drawing can be highly informative.

A second useful feature of CogSketch is that it requires participants to segment their drawing into objects and label them with words or phrases[7]. While people could be instructed to label the parts of their drawing in a paper-and-pencil drawing task, it is likely that without close supervision they would sometimes fail to do so, complicating the analysis. Furthermore, the fact that participants distinguished objects from relations using labels (in addition to symbols) was extremely useful in our analyses. In this study, some aspects of the coding— including some of the identification of key events—were done by human coders, but this work was facilitated by the fact that segmentation, labeling, and ordering of events was provided by CogSketch.

CogSketch's use of participant labeling distinguishes it from most other sketch understanding systems, which are recognition-based (e.g., de Silva et al. 2007; Pargas et al. 2007; Plimmer and Freeman 2007; Plimmer and Hammond 2008; Valentine et al. 2012). Recognition-based systems require that there is a one-to-one mapping between a shape and its intended meaning. This works well when a domain has a graphical symbol vocabulary (e.g., electronics, organic chemistry, and free body diagrams). However, this requirement does not fit many kinds of scientific diagrams. For example, in geoscience, a line can represent a fault, a bedding layer, a horizon line, and many other objects. Moreover, many entities in scientific diagrams are not graphical symbols. When annotating a photograph to show faults and marker beds, the shape of these entities is determined by the structure shown in the photograph, not by

---

[7] CogSketch normally requires the user to categorize their glyphs as objects or relations, but to allow for more spontaneity, this feature was not used in this research.

a set of domain conventions, as they are in electronics, for instance. Labeling removes any ambiguity about the meaning of a sketch object and dispenses with the need for perfect depiction of the objects in a scene. Although providing labels could be seen as burdensome compared with free sketching, it has the advantage of making the drawing task less forbidding for participants, since they do not need to rely on their drawing skills alone to communicate.

Sketch understanding software could eventually have a profound impact on educational assessment and instruction. Automated recording and analysis of sketches could facilitate the assessment of spatial scientific knowledge and the scoring of standardized tests that involve drawing, greatly reducing the burden on instructors and allowing more rapid feedback to the students. Sketch understanding software could also help improve student learning directly. CogSketch has been used to develop worksheets that are being used in classroom experiments (Yin et al. 2010) to attempt to provide such benefits. Furthermore, these benefits extend beyond geoscience. CogSketch has been used to support learning in fifth-grade biology classrooms (Miller et al. 2014) and in college engineering classes (Wetzel and Forbus 2012). Additional research is needed to explore the connections between sketching and learning, and new technologies are sure to play an important role in this endeavor (Ainsworth et al. 2011).

Conclusions

Our findings suggest that differences in domain-specific knowledge are reflected in both the products and the process of sketching. These differences were apparent even when participants were simply asked to copy the materials in front of them. Our results further suggest that the differences between novices and geoscience students were largely domain-specific. We suggest that the techniques used here could be applied in other spatially intense domains to explore students' knowledge and to support their learning. Ultimately, work on sketching will not only inform our understanding of STEM learning but will contribute to our understanding of the nature of scientific expertise and how it develops.

Appendix

Table 4 Image lists for the experiments

Experiment	Image type	
	Geoscience	Control
Experiment 1	Folded rock	
	Rock layers	
	Subduction 1	
Experiment 2 (geoformations)	Fold and unconformity	Deck of cards folded in a hand
	Cross-cutting faults	Fruit cross-section
	Cross-cutting dike with faults	Slice of layer cake
Experiment 3 (causal diagrams)	Subduction 2	Carburetor
	Water cycle	Blood pump for artificial heart
	Phosphorus cycle	Home water conservation system
	Sand budget	Photosynthesis

References

Ainsworth, S., Prain, V., & Tytler, R. (2011). Drawing to learn in science. *Science, 333*, 1096–1097.

Black, A. A. (2005). Spatial ability and earth science conceptual understanding. *Journal of Geoscience Education, 53*(4), 402–414.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*, 55–81.

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

Chi, M. T. H., Glaser, R., & Farr, M. (Eds.). (1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.

de Silva, R., Bischel, T. D., Lee, W., Peterson, E. J., Calfee, R. C. & Stahovich, T., 2007. Kirchhoff's pen: A pen-based circuit analysis tutor. In *Proceedings of the 4th Eurographics workshop on Sketch-based interfaces and modeling.*

Dihoff, R. E., Brosvic, G. M., & Epstein, M. (2004). Provision of feedback during preparation for academic testing: learning is enhanced by immediate but not delayed feedback. *Psychological Record, 54*, 207–231.

Do, E. Y. (2005). Design sketches and sketch design tools. *Knowledge Based Systems, 18*, 383–405.

Duley, J. F., Wilkins, J., Hamby, S., Hopkins, D., Burwell, R., & Barry, N. (1993). Explicit scoring criteria for the Rey–Osterrieth and Taylor complex figures. *The Clinical Neuropsychologist, 7*, 29–38.

Forbus, K. (2011). Qualitative modeling. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*(4), 374–391.

Forbus, K., Usher, J., Lovett, A., Lockwood, K., & Wetzel, J. (2008). CogSketch: open-domain sketch understanding for cognitive science research and for education. In the *Proceedings of the Fifth Eurographics Workshop on Sketch-Based Interfaces and Modeling*. Annecy, France

Forbus, K., Usher, J., Lovett, A., Lockwood, K., & Wetzel, J. (2011). CogSketch: sketch understanding for cognitive science research and for education. *Topics in Cognitive Science, 3*, 648–666.

Gentner, D., & Ratterman, M. J. (1991). Language and the career of similarity. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on thought and language: interrelations in development* (pp. 257–277). London: Cambridge University Press.

Gentner, D., & Stevens, A. L. (Eds.). (1983). *Mental models*. Hillsdale, NJ: Erlbaum.

Gobert, J. (2000). A typology of models for plate tectonics: inferential power and barriers to understanding. *International Journal of Science Education, 22*, 937–977.

Gobert, J., & Clement, J. J. (1999). Effects of student-generated diagrams versus student-generated summaries on conceptual understanding of causal and dynamic knowledge in plate tectonics. *Journal of Research in Science Teaching, 36*, 39–53.

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C. H., Jones, G., Oliver, I., et al. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Science, 5*, 236–243.

Goel, V. (1995). *Sketches of thought*. Cambridge: MA: MIT Press.

Hay, D. B., Williams, D., Stahl, D., & Wingate, R. J. (2013). Using drawings of the brain cell to exhibit expertise in neuroscience: exploring the boundaries of experimental culture. *Science Education, 97*(3), 468–491.

Heiser, J., & Tversky, B. (2006). Arrows in comprehending and producing mechanical diagrams. *Cognitive Science, 30*, 581–592.

Hmelo-Silver, C. E., & Pfeffer, M. G. (2004). Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. *Cognitive Science, 1*, 127–138.

Holden, M., Curby, K., Newcombe, N., & Shipley, T. F. (2010). Spatial memory: hierarchical encoding of location in natural scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(3), 590–604.

Jee, B. D., & Wiley, J. (2007). How goals affect the organization and use of domain knowledge. *Memory and Cognition, 35*, 837–851.

Jee, B. D., Uttal, D. H., Gentner, D., Manduca, C., Shipley, T., Sageman, B., et al. (2010). Analogical thinking in geoscience education. *Journal of Geoscience Education, 58*, 2–13.

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

Kali, Y., & Orion, N. (1996). Spatial abilities of high school students in the perception of geologic structures. *Journal of Research in Science Teaching, 33*, 369–391.

Kastens, K. A., Agrawal, S., & Liben, L. (2009). How students and field geologists reason in integrating spatial observations from outcrops to visualize a 3-D geological structure. *International Journal of Science Education, 31*, 365–394.

Kindfield, A. (1992). Expert diagrammatic reasoning in biology. Paper presented at the *AAAI Symposium on Reasoning with Diagrammatic Representations I*, Stanford University.

Kulik, J. A., & Kulik, C. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research, 58*, 79–97.

Landy, D., & Goldstone, R. L. (2007). Formal notations are diagrams: evidence from a production task. *Memory and Cognition, 35*, 2033–2040.

Larkin, J., & Simon, H. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, 65–99.

Libarkin, J. C., & Brick, C. (2002). Research methodologies in science education: visualization and the geosciences. *Journal of Geoscience Education, 50*, 449–455.

Liben, L. S., & Titus, S. J. (2012). The importance of spatial thinking for geoscience education: Insights from the crossroads of geoscience and cognitive science. *Geological Society of America Special Papers, 86*, 51–70.

Lowe, R. K. (1993). Constructing a mental representation from an abstract technical diagram. *Learning and Instruction, 3*, 157–179.

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

Miller, B. W., Cromley, J. G., Newcombe, N. S., Chang, M. D., & Forbus, K. D. (2014). *Supporting student science learning through sketching with on-demand feedback.* Manuscript in preparation.

Moss, J., Kotovsky, K., & Cagan, J. (2006). The role of functionality in the mental representations of engineering students: some differences in the early stages of expertise. *Cognitive Science, 30*(1), 65–93.

Pargas, R., Cooper, M., Williams, C., and Bryfczynski, S. (2007). Organicpad: A tablet pc based interactivity tool for organic chemistry. *First International International Workshop on Pen-based Learning Technologies. (pp. 1-6). IEEE.*

Piburn, M.D., van der Hoeven Kraft, K., and Pacheco, H. (2011). A new century for geoscience education research. Paper presented at the *Second Committee Meeting on the Status, Contributions, and Future Directions of Discipline-Based Education Research.* Available: http://www.nationalacademies.org/bose/DBER_Piburn_October_Paper.pdf.

Plimmer, B., & Freeman, I. (2007). A toolkit approach to sketched diagram recognition. *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI...but not as we know it*, (Vol. 1, 205–213). Swinton, UK: Computer Society.

Plimmer, B., & Hammond, T. (2008). Getting started with sketch tools: a tutorial on sketch recognition tools. *Lecture Notes in Computer Science, 5223*, 9–12.

Proffitt, J. B., Coley, J. D., & Medin, D. L. (2000). Expertise and category-based induction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 811–828.

Schunn, C. D., & Anderson, J. R. (1999). The generality/specificity of expertise in scientific reasoning. *Cognitive Science, 23*, 337–370.

Schwamborn, A., Mayer, R. E., Thillmann, H., Leopold, C., & Leutner, D. (2010). Drawing as a generative activity and drawing as a prognostic activity. *Journal of Educational Psychology, 102*(4), 872–879.

Shea, D. L., Lubinski, D., & Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: a 20-year longitudinal study. *Journal of Educational Psychology, 93*(3), 604–614.

Shin, M. S., Park, S. Y., Park, S. R., Seol, S. H., & Kwon, J. S. (2006). Clinical and empirical applications of the Rey–Osterrieth Complex Figure Test. *Nature Protocols, 1*, 892–899.

Suwa, M., & Tversky, B. (1997). What architects and students perceive in their sketches: a protocol analysis. *Design Studies, 18*, 385–403.

Taylor, H. A., & Tversky, B. (1992). Descriptions and depictions of environments. *Memory and Cognition, 20*, 483–496.

Tversky, B. (1981). Distortions in memory for maps. *Cognitive Psychology, 13*, 407–433.

Tversky, B. (1991). Spatial mental models. In G. H. Bower (Ed.), *The psychology of learning and motivation: advances in research and theory* (Vol. 27, pp. 109–145). N.Y.: Academic Press.

Tversky, B. (2002). What do sketches say about thinking? In the *Proceedings of the AAAI Spring Symposium on Sketch Understanding,* 148-151.

Uttal, D. H., & Cohen, C. A. (2012). Spatial thinking and STEM education: when, why and how. *Psychology of Learning and Motivation, 57*, 147–181.

Uttal, D. H., Friedman, A., Hand, L. L., & Warren, C. (2010). Learning fine-grained and category information in navigable real-world space. *Memory and Cognition, 38*, 1026–1040.

Valentine, S., Vides, F., Lucchese, G., Turner, D., Kim, H., Li, W., Linsey J., & Hammond, T. (2012). Mechanix: A sketch-based tutoring system for statics courses. In *Proceedings of AAAI.*

Van Meter, P., & Garner, J. (2005). The promise and practice of learner-generated drawing: literature review and synthesis. *Educational Psychology Review, 17*(2), 285–325.

Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: a study of conceptual change in childhood. *Cognitive Psychology, 24*, 535–585.

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101*(4), 817.

Wetzel, J., & Forbus, K. (2012). Teleological representations for multi-modal design explanations. *Proceedings of the 26th International Workshop on Qualitative Reasoning*. Los Angeles, California.

White, R. T., & Gunstone, R. F. (1992). *Probing understanding*. New York: Falmer Press.

Yin, P., Forbus, K., Usher, J., Sageman, B., & Jee, B. D. (2010). Sketch Worksheets: A sketch-based educational software system. *Proceedings of the 22nd Annual Conference on Innovative Applications of Artificial Intelligence*.